Learning WHO Saying WHAT to WHOM in Multi-Party Conversations

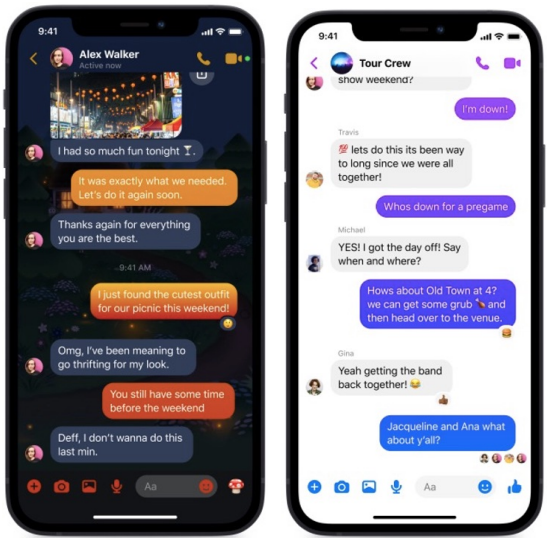
Jia-Chen Gu[1], **Zhuosheng Zhang**[2], **Zhen-Hua Ling**[1]
[1]University of Science and Technology of China
[2]Shanghai Jiao Tong University
{gujc,zhling}@ustc.edu.cn, zhangzs@sjtu.edu.cn

Abstract

Multi-party conversations (MPC) are a more practical and challenging scenario involving more than two interlocutors. This research topic has drawn significant attention from both academia and industry, and it is nowadays counted as one of the most promising research areas in the field of dialogue systems. In general, MPC algorithms aim at addressing the issues of *Who* saying *What* to *Whom*, specifically, who speaks, say what, and address whom. The complicated interactions between interlocutors, between utterances, and between interlocutors and utterances develop many variant tasks of MPC worth investigation. In this tutorial, we present a comprehensive survey of recent advances in MPC. In particular, we summarize recent advances on the research of MPC modeling which is categorized by *Who* saying *What* to *Whom*. Finally, we highlight the challenges which are not yet well addressed in MPC and present future research directions.

1 Introduction

The development of intelligent dialogue systems that are able to engage in conversations with humans, has been one of the longest running goals in artificial intelligence (Kepuska and Bohouta, 2018; Berdasco et al., 2019; Zhou et al., 2020). Thanks to breakthroughs in sequence modeling (Sutskever et al., 2014; Vaswani et al., 2017) and pre-trained language models (PLMs) (Radford et al., 2019; Devlin et al., 2019; Lewis et al., 2020), researchers have proposed various effective models for conversations between two participants (Serban et al., 2016; Wen et al., 2017; Zhang et al., 2020). Recently, researchers have paid more attention to a more practical and challenging scenario involving more than two participants, which is well known as multi-party conversations (MPC) (Ouchi and Tsuboi, 2016; Zhang et al., 2018; Le et al., 2019; Hu et al., 2019; Kummerfeld et al., 2019; Gu et al., 2021, 2022a,b, 2023; Li and Zhao, 2023). Figure 1 presents instances of two-party and multi-party conversations, corresponding to the common scenarios of one-to-one chat and group chat respectively in daily life. Utterances in a two-party conversation are posted between two interlocutors alternately, constituting a *sequential* information flow. On the other hand, each utterance in an MPC can be spoken by anyone and address anyone else in this conversation, which constitutes a *graphical* information flow.



(a) Two-party conversation (b) Multi-party conversation

Figure 1: Illustration of the two-party (one-on-one chat) and multi-party (group chat) conversation instances.

Modeling *Who* saying *What* to *Whom* in MPC poses unique research challenges, specifically, who speaks, say what, and address whom. All these have been revisited by researchers since the emergence of neural approaches as the dominant approach for solving MPC problems. The emergence of large language models (LLMs) (Brown et al., 2020; Ouyang et al., 2022), especially ChatGPT[1], has enabled more engaging conversations in the form of two-party conversations,

[1]https://chat.openai.com/

Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Tutorial Abstracts), pages 29–34
November 1–4, 2023.

Slot	Theme
Session 1:	
14:15 – 14:20	Tutorial presenters introduction
14:20 – 14:35	Introduce the task, history and application
14:35 – 14:45	Introduce the part of datasets
14:45 – 15:05	Introduce the part of "WHO Speaks"
15:05 – 15:35	Introduce the explicit part of "Address WHOM"
15:35 – 15:45	Q&A
15:45 – 16:15	Coffee break
Session 2:	
16:15 – 16:45	Introduce the implicit part of "Address WHOM"
16:45 – 17:15	Introduce the part of "Say WHAT"
17:15 – 17:30	Challenges and open questions
17:30 – 17:35	Conclusion
17:35 – 17:45	Q&A

Table 1: Tutorial schedule.

i.e., a conversation between an agent and a user. However, the open questions of whether ChatGPT still shows great performance in group chat, and can be adapted to the form of MPC effectively and efficiently are worth investigation. Furthermore, whether these chat-style LLMs are capable of simulating human interactions e.g., conversation and perception, among multiple agents in a virtual environment (Park et al., 2023) remains research-worthy. This tutorial will survey the cutting-edge methods for MPC, introducing key sub-areas whose combination is needed for a successful solution, which have not been covered in any previous AI & NLP conference tutorial.

2 Tutorial Outline

This will be a **three-hour** tutorial devoted to the **cutting-edge** topic of MPC. This tutorial will include two sessions. Each session will be 1.5h, followed by 10 minutes for Q&A and 30 minutes for a break. Each part includes an overview of the corresponding topic, widely used methods and a deep dive into representative research. The detailed tutorial schedule is shown in Table 1.

The complicated interactions between interlocutors and utterances develop many variant tasks of MPC worth investigation. For example, there are scenarios where multiple people may want to interact with a chatbot or one person interacts with several chatbots at the same time, as in a chat group, to coordinate among themselves and achieve a common goal. Thus, the ability to predict which agent in the conversation is the most likely to speak next, and conversely, when an agent must wait before interacting, is important for conducting engaging and social conversations (Pinhanez et al., 2018; de Bayser et al., 2019). Furthermore, detecting who is being addressed, i.e., who the current speaker is talking to, is also non-trivial in these conversation scenarios (Ouchi and Tsuboi, 2016; Zhang et al., 2018; Le et al., 2019; Gu et al., 2021, 2023). Last but not least, only after knowing a speaker and an addressee at the current dialogue state, can the system return an appropriate response following the conversation (Hu et al., 2019; Wang et al., 2020; Gu et al., 2022a; Li and Zhao, 2023).

2.1 WHO Speaks

A multi-party conversation may involve the interaction between human-human, human-machine, and machine-machine. Different speakers have different characteristics, with diverse styles, emotions, and opinions. The task of *speaker segmentation*, also known as *speaker diarisation* or *speaker change detection*, aims at finding speaker changing points in a conversation. Specifically, a speaker change occurs when the current and the previous utterances are not uttered by the same speaker. Meng et al. (2017) formulate this task as a binary utterance-pair classification to judge whether the speaker changes before and after a certain decision point. Meng et al. (2018) propose another surrogate task for general speaker modeling. This task is defined by segmenting an MPC into several parts according to speakers, each segment of which comprises one or a few consecutive sentences uttered by a particular speaker. A candidate set of speakers is also given and models are required to identify the speaker of each segment. Gu et al. (2021) propose an utterance semantics-based speaker searching task where models are asked to search for a speaker in history that shares the same speaker with the expected conversation turn.

2.2 Address WHOM

Existing methods on detecting who is being addressed can be generally categorized into *explicit* and *implicit* ones.

Explicit Addressee Recognition This task *explicitly* determines the intended addressee of an utterance. Previous studies mainly focus on predicting the addressee of only the last utterance of a conversation (Ouchi and Tsuboi, 2016; Zhang et al., 2018), while recent studies pay more attention to predicting the addressees of all utterances of a conversation (Le et al., 2019; Gu et al., 2021, 2023). Le et al. (2019) propose a who-to-whom (W2W) model to recognize and complete the addressees of all utterances in a conversation to

help understand the whole conversation, given an MPC where part of the addressees are unspecified. Gu et al. (2021) propose a pre-trained MPC-BERT language model for universal MPC understanding by designing self-supervised tasks, and test it on addressee recognition. Gu et al. (2023) present graph-induced fine-tuning (GIFT) which can adapt various Transformer-based LMs by designing four types of edges to integrate graph-induced signals into attention mechanisms.

Implicit Dialogue Disentanglement The messages from different interlocutors on different topics are heavily interwoven. Therefore, it is necessary to disentangle a whole conversation into several threads from a data stream so that each thread is about a specific topic. Basically, most of the existing methods are designed to find out which previous utterance in the history the current utterance is replying to. Thus, this task is in essence modeling addressees *implicitly*. Kummerfeld et al. (2019) create a large-scale `#Ubuntu` IRC corpus that is 16 times larger than all previously released datasets combined. Yu and Joty (2020) formalize the link prediction of disentanglement as a pointing problem following pointer networks. Each pointing operation is modeled as a multinomial distribution over the set of previous utterances. Ma et al. (2022) propose characteristic features of speaker property and reference dependency for dialogue structure.

2.3 Say WHAT

Generating appropriate responses in the context of MPC is challenging due to the diverse intent topic transition and information temporality. Existing methods enabling MPC systems to decide what to say can be generally categorized into *retrieval*-based and *generation*-based methods.

Retrieval-based The retrieval-based methods aims at selecting the best-matched response from a set of candidates, given the context of a multi-turn conversation. The key to this task is to rank the set of response candidates according to the semantic matching between the context of an MPC and a response candidate. Ouchi and Tsuboi (2016) propose jointly modeling the tasks of response selection and addressee selection to capture what is being said to whom at each time step in a context. Zhang et al. (2018) follow this framework and improve it by updating the interlocutor embeddings role-sensitively. Gu et al. (2021) propose jointly learning who says what to whom in a unified framework by considering addressee- and speaker-related tasks as complementary signals for response selection during the pre-training stage.

Generation-based The generation-based methods synthesize a response with generative models by maximizing its generation probability given the previous conversation history. Hu et al. (2019) propose a graph-structured neural network, the core of which is to encode utterances based on the graph topology rather than the sequence of their appearances in a conversation, to model the information flow as *graphical*. Gu et al. (2022a) propose to model complicated interactions between utterances and interlocutors with a heterogeneous graph, where two types of graph nodes and six types of edges are designed to model heterogeneity in MPC. Li and Zhao (2023) propose an EM approach that iteratively performs the expectation steps to generate missing addressees, and the maximization steps to optimize a generative model.

2.4 Open Challenges

Robustness to Addressee Scarcity Existing methods rely heavily on the necessary addressee labels and can only be applied to an ideal setting where each utterance must be tagged with an addressee label. In practice, statistics show that addressees of 55% of the utterances in the Ubuntu IRC dataset (Ouchi and Tsuboi, 2016) are not specified. Given an MPC with a few addressee labels missing, existing methods fail to build a consecutively connected conversation graph, but only a few separate conversation fragments instead. Despite Li and Zhao (2023) make preliminary exploration on the scarcity of addressee, the cost of missing addressee inference is too high, and currently only the silver label for the last utterance in a conversation is affordable to infer. A future research direction could be ensuring message passing between these conversation fragments and enhancing the robustness to addressee scarcity.

Universal MPC Understanding Most existing studies design models for each individual task in MPC separately. Intuitively, the complicated interactions between interlocutors and utterances might make these tasks complementary among each other. Making use of these tasks simultaneously may produce better contextualized representations of interlocutors and utterances, and would enhance the conversation understanding. Gu et al. (2021) focus on modeling only interlocutor structures and utterance semantics. A future research direction could be designing better supervised tasks for

augmenting LLMs with more abilities, and being tested on more tasks to evaluate model robustness and generalization. With the rapid growth of LLMs, machine-machine interaction also becomes an important direction, where the machines may collaborate, debate, and evolve via MPC.

Topic Transition Despite the remarkable ability of LLMs to comprehend and generate language, the performance would drop significantly when its comes to topic transition and long context. This is precisely where a good MPC system excels. An MPC system should be able to track topic transitions and update the conversation structure dynamically as a conversation proceeds, so that it can understand deep semantics and structures of an MPC. Modeling topic transitions in MPC which tracks the specific flows of multiple ongoing sub-conversations at a fine granularity could serve as a surrogate task for complex context understanding.

3 Tutorial Presenters

Jia-Chen Gu is currently a Postdoctoral Researcher at University of Science and Technology of China. His research interests lie within machine learning for dialogue systems. Homepage: http://home.ustc.edu.cn/~gujc/

Zhuosheng Zhang is currently an Assistant Professor at Shanghai Jiao Tong University. His research interests include natural language processing, dialogue systems, and large language models. He has given a tutorial on Machine Reading Comprehension: The Role of Contextualized Language Models and Beyond at IJCAI 2021. Homepage: https://bcmi.sjtu.edu.cn/~zhangzs/

Zhen-Hua Ling is a Professor with the University of Science and Technology of China. He was a Visiting Scholar at the University of Washington and a Marie Curie Fellow at the University of Edinburgh. His research interests include speech processing and natural language processing. He was the recipient of the IEEE Signal Processing Society Young Author Best Paper Award in 2010. He was an Associate Editor of IEEE/ACM Transcations on Audio, Speech, and Language Processing from 2014 to 2018. Homepage: http://staff.ustc.edu.cn/~zhling/

4 Diversity Considerations

The MPC techniques we introduce is language agnostic. Thus, they can be applied to data in various languages and localities with some extent of adaption to scale beyond English. Presenters will share this tutorial with a worldwide audience by promoting on social media. Presenters span over junior (J.-C. Gu and Z. Zhang) and senior researchers (Z.-H. Ling). Thus, we have diversified instructors which will also help encourage diverse audience. J.-C. Gu and Z. Zhang have experience co-organizing NLP Workshops, and actively work on inviting undergraduate students to research and promoting diversity. We will work with *ACL D&I teams, and consult resources to diversify our audience participation.

5 Prerequisite

The prerequisite includes familiarity with basic machine learning and deep learning models, especially those typically used in modern NLP for MPC, including representation learning, Transformers, etc. Furthermore, this tutorial assumes background in basic probability, linear algebra, and calculus. We will also provide introductory materials.

Reading List The presenters have survey papers for comprehensive references (Gu et al., 2022b; Zhang and Zhao, 2021). The following papers are also recommended: Ouchi and Tsuboi (2016); Hu et al. (2019); Kummerfeld et al. (2019); Gu et al. (2021); Ma et al. (2022); Gu et al. (2022a); Li and Zhao (2023); Gu et al. (2023).

Breadth While dozens of relevant papers over the tutorial are provided, we plan to cover around 10-15 research papers in detail. Only 3-5 of the "deep dive" papers come from the presenter team.

6 Tutorial Details

Audience Size Audience sizes for physical and virtual meetings are expected to be around 100 and 150 respectively.

Open Access The slides, code and other teaching materials will be released online for public access. The video recording of our tutorial will also be included in the ACL Anthology.

7 Ethics Statement

Certain conversation data might come from private dialogues between people. Thus, privacy considerations must be taken to ensure all data that is released conforms to regulations and are under consent. As conversations and pre-trained language models may have bias in various forms, MPC models may contain the same form of bias and should be reviewed and modified if necessary.

References

Ana Berdasco, Gustavo López, Ignacio Díaz-Oreiro, Luis Quesada, and Luis A. Guerrero. 2019. User experience comparison of intelligent personal assistants: Alexa, google assistant, siri and cortana. In *13th International Conference on Ubiquitous Computing and Ambient Intelligence, UCAmI 2019, Toledo, Spain, December 2-5, 2019*, volume 31 of *MDPI Proceedings*, page 51. MDPI.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, et al. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Maíra Gatti de Bayser, Paulo Rodrigo Cavalin, Claudio S. Pinhanez, and Bianca Zadrozny. 2019. Learning multi-party turn-taking models from dialogue logs. *CoRR*, abs/1907.02090.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Jia-Chen Gu, Zhenhua Ling, Quan Liu, Cong Liu, and Guoping Hu. 2023. GIFT: graph-induced fine-tuning for multi-party conversation understanding. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2023, Toronto, Canada, July 9-14, 2023*, pages 11645–11658. Association for Computational Linguistics.

Jia-Chen Gu, Chao-Hong Tan, Chongyang Tao, Zhen-Hua Ling, Huang Hu, Xiubo Geng, and Daxin Jiang. 2022a. HeterMPC: A heterogeneous graph neural network for response generation in multi-party conversations. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022*, pages 5086–5097. Association for Computational Linguistics.

Jia-Chen Gu, Chongyang Tao, and Zhen-Hua Ling. 2022b. Who says what to whom: A survey of multi-party conversations. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, pages 5486–5493. ijcai.org.

Jia-Chen Gu, Chongyang Tao, Zhen-Hua Ling, Can Xu, Xiubo Geng, and Daxin Jiang. 2021. MPC-BERT: A pre-trained language model for multi-party conversation understanding. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021*, pages 3682–3692. Association for Computational Linguistics.

Wenpeng Hu, Zhangming Chan, Bing Liu, Dongyan Zhao, Jinwen Ma, and Rui Yan. 2019. GSN: A graph-structured network for multi-party dialogues. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI 2019, Macao, China, August 10-16, 2019*, pages 5010–5016.

Veton Kepuska and Gamal Bohouta. 2018. Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home). In *IEEE 8th Annual Computing and Communication Workshop and Conference, CCWC 2018, Las Vegas, NV, USA, January 8-10, 2018*, pages 99–103. IEEE.

Jonathan K. Kummerfeld, Sai R. Gouravajhala, Joseph Peper, Vignesh Athreya, R. Chulaka Gunasekara, Jatin Ganhotra, Siva Sankalp Patel, Lazaros C. Polymenakos, and Walter S. Lasecki. 2019. A large-scale corpus for conversation disentanglement. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28- August 2, 2019, Volume 1: Long Papers*, pages 3846–3856.

Ran Le, Wenpeng Hu, Mingyue Shang, Zhenjun You, Lidong Bing, Dongyan Zhao, and Rui Yan. 2019. Who is speaking to whom? learning to identify utterance addressee in multi-party conversations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 1909–1919.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 7871–7880. Association for Computational Linguistics.

Yiyang Li and Hai Zhao. 2023. EM pre-training for multi-party dialogue response generation. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, ACL 2023, (Volume 1: Long Papers), Toronto, Canada, July 9-14, 2023*. Association for Computational Linguistics.

Xinbei Ma, Zhuosheng Zhang, and Hai Zhao. 2022. Structural characterization for dialogue disentanglement. In *Proceedings of the 60th Annual Meeting*

of the Association for Computational Linguistics (Volume 1: Long Papers), ACL 2022, Dublin, Ireland, May 22-27, 2022, pages 285–297. Association for Computational Linguistics.

Zhao Meng, Lili Mou, and Zhi Jin. 2017. Hierarchical RNN with static sentence-level attention for text-based speaker change detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management, CIKM 2017, Singapore, November 06 - 10, 2017*, pages 2203–2206. ACM.

Zhao Meng, Lili Mou, and Zhi Jin. 2018. Towards neural speaker modeling in multi-party conversation: The task, dataset, and models. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation, LREC 2018, Miyazaki, Japan, May 7-12, 2018*. European Language Resources Association (ELRA).

Hiroki Ouchi and Yuta Tsuboi. 2016. Addressee and response selection for multi-party conversation. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, EMNLP 2016, Austin, Texas, USA, November 1-4, 2016*, pages 2133–2143.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F. Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022*.

Joon Sung Park, Joseph C. O'Brien, Carrie J. Cai, Meredith Ringel Morris, Percy Liang, and Michael S. Bernstein. 2023. Generative agents: Interactive simulacra of human behavior. *CoRR*, abs/2304.03442.

Claudio S. Pinhanez, Heloisa Candello, Mauro Carlos Pichiliani, Marisa Vasconcelos, Melina Alberio Guerra, Maíra Gatti de Bayser, and Paulo Rodrigo Cavalin. 2018. Different but equal: Comparing user collaboration with digital personal assistants vs. teams of expert agents. *CoRR*, abs/1808.08157.

Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Iulian Vlad Serban, Alessandro Sordoni, Yoshua Bengio, Aaron C. Courville, and Joelle Pineau. 2016. Building end-to-end dialogue systems using generative hierarchical neural network models. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, pages 3776–3784.

Ilya Sutskever, Oriol Vinyals, and Quoc V. Le. 2014. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada*, pages 3104–3112.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 5998–6008.

Weishi Wang, Steven C. H. Hoi, and Shafiq R. Joty. 2020. Response selection for multi-party conversations with dynamic topic tracking. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 6581–6591.

Tsung-Hsien Wen, David Vandyke, Nikola Mrksic, Milica Gasic, Lina Maria Rojas-Barahona, Pei-Hao Su, Stefan Ultes, and Steve J. Young. 2017. A network-based end-to-end trainable task-oriented dialogue system. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 1: Long Papers*, pages 438–449.

Tao Yu and Shafiq R. Joty. 2020. Online conversation disentanglement with pointer networks. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 6321–6330. Association for Computational Linguistics.

Rui Zhang, Honglak Lee, Lazaros Polymenakos, and Dragomir R. Radev. 2018. Addressee and response selection in multi-party conversations with speaker interaction rnns. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pages 5690–5697.

Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2020. DIALOGPT : Large-scale generative pre-training for conversational response generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations, ACL 2020, Online, July 5-10, 2020*, pages 270–278. Association for Computational Linguistics.

Zhuosheng Zhang and Hai Zhao. 2021. Advances in multi-turn dialogue comprehension: A survey. *arXiv preprint arXiv:2103.03125*.

Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. 2020. The design and implementation of xiaoice, an empathetic social chatbot. *Comput. Linguistics*, 46(1):53–93.