Data Augmentation for Conversational AI

Heydar Soudani
heydar.soudani@ru.nl
Radboud University
Nijmegen, The Netherlands

Roxana Petcu
r.m.petcu@uva.nl
University of Amsterdam
Amsterdam, The Netherlands

Evangelos Kanoulas
e.kanoulas@uva.nl
University of Amsterdam
Amsterdam, The Netherlands

Faegheh Hasibi
faegheh.hasibi@ru.nl
Radboud University
Nijmegen, The Netherlands

ABSTRACT

Advancements in conversational systems have revolutionized information access, surpassing the limitations of single queries. However, developing dialogue systems requires a large amount of training data, which is a challenge in low-resource domains and languages. Traditional data collection methods like crowd-sourcing are labor-intensive and time-consuming, making them ineffective in this context. Data augmentation (DA) is an effective approach to alleviate the data scarcity problem in conversational systems. This tutorial provides a comprehensive and up-to-date overview of DA approaches in the context of conversational systems. It highlights recent advances in open domain and task-oriented conversation generation, and different paradigms of evaluating these models. We also discuss current challenges and future directions in order to help researchers and practitioners to further advance the field in this area.

CCS CONCEPTS

• **Information systems → Recommender systems**; • **Computing methodologies → Discourse, dialogue and pragmatics**.

KEYWORDS

Conversational AI, Data Augmentation, Dataset Creation, Conversation Generation

1 INTRODUCTION

Motivation. The development of dialogue systems has garnered significant attention and demand in both industry and everyday life due to their diverse range of functionalities that cater to user needs.

These systems can be broadly classified into two categories: task-oriented dialogue systems (TOD) and open-domain dialogue systems (ODD) [26]. TOD systems are specifically designed to address particular problems within a specific domain, with the objective of performing tasks like ticket or table reservations [27]. Thus, their primary focus is task completion. On the other hand, ODD systems engage in unrestricted conversations on a wide range of topics. The primary challenge faced by ODD systems is ensuring coherence and consistency when generating responses. This implies that the generated responses must be context-aware, taking into account the conversation history [26].

The progress of dialogue systems relies heavily on the use of large neural models, similar to other natural language processing (NLP) tasks, and as a result, their effectiveness is contingent upon the availability of substantial amounts of training data [45]. The dependence on large scale training data challenges development of dialogue agents, particularly for low resource settings with limited or no training data. While crowd-sourcing serves as the primary method for generating datasets in dialogue systems, it is labor intensive nature possesses limitations concerning time, cost, and scalability [18]. The scarcity of data for diverse and specific domains, compounded by the difficulty of adapting existing datasets or generating entirely new ones calls for alternative methods of training dialogue systems.

To tackle the issue of data shortage in dialogue systems, several methods have been proposed, including semi-supervised learning and data augmentation (DA) [42]. While semi-supervised learning is a promising approach, relying solely on existing dialogue datasets presents a classic chicken-and-egg problem. DA, on the other hand, involves generating conversation samples from external resources, such as unstructured text files and structured data like knowledge graphs. This approach serves multiple purposes: it diversifies datasets, introduces novel conversational scenarios, and enhances control over the flow of the generated conversation.

In this tutorial, we aim to offer a comprehensive overview of conversation augmentation and generation methods for TOD and ODD systems. We delve into the specific challenges that must be addressed when undertaking the task of creating new dialogue data. To provide a comprehensive package for dialogue data creation, we also present an overview of evaluation methods and available datasets that can be utilized to assess the quality and performance of the generated dialogue data. By covering these aspects, our tutorial offers a holistic understanding of the methods, challenges, and evaluation procedures associated with the dialogue data creation.

Previous tutorials. Our tutorial builds upon two key concepts: conversational systems and DA. Tutorials that discuss these concepts in recent years include:

- Conversational Recommender Systems [9] in RecSys 2020. This tutorial focuses on the foundations and algorithms for conversational recommendation and their application in real-world systems like search engines, e-commerce, and social networks.
- Conversational Information Seeking: Theory and Application [7] in SIGIR 2022. The tutorial aims to provide an introduction to Conversational Information Seeking (CIS), covering recent advanced topics and state-of-the-art approaches.
- Self-Supervised Learning for Recommendation [11] in CIKM 2022. This tutorial aims to present a systematic review of state-of-the-art self-supervised learning (SSL)-based recommender systems.
- Limited Data Learning [42] in ACL 2022. This tutorial offers an overview of methods alleviating the need for labeled data in NLP, including DA and semi-supervised learning.
- Proactive Conversational Agents [19] in WSDM 2023. This tutorial introduces methods to equip conversational agents with the ability to interact with end users proactively.

Unlike previous tutorials that focus on either conversational systems or the data scarcity problem, this tutorial provides an in-depth exploration of the challenges associated with augmenting and creating conversational data, highlighting the unique difficulties posed in conversational context. For the first time, we presented this tutorial at the CIKM 2023 conference held in Birmingham, UK. Approximately 25 individuals from various communities participated in our tutorial. The audience was researchers interested in the field of conversational systems as well as specific fields such as medicine, law, and education. Additionally, the subject captured the attention of researchers and practitioners in the industry seeking to develop dialogue systems tailored to specific domains.

Target audience and prerequisites. This tutorial is designed for professionals, students, and researchers in information retrieval and natural language processing, specifically interested in conversational AI. Familiarity with machine learning, deep learning, and transformers is required. No prior knowledge of dialogue system models or data augmentation methods is necessary.

Tutorial material. Tutorial slides, video teaser, a collection of essential references, and other support materials can be found at the tutorial website https://dataug-convai.github.io.

2 TUTORIAL OUTLINE

We plan to give a half-day lecture-style tutorial. The tutorial starts by an introduction, followed by three main sessions. We plan to have a short Q&A after each session and conclude the tutorial with a discussion on future direction and a final Q&A session.

2.1 Agenda

A tentative schedule of the tutorial is as follows.

(1) Introduction **(20 min)**
  (a) Conversational Systems
  (b) Problem of Data Scarcity
  (c) Data Augmentation Approach
(2) Evaluation **(20 min)**
  (a) Automatic Methods
  (b) Human-based Methods
(3) Conversation Generation: Task-oriented **(50 min)**
  (a) PreLLM:
    (i) Agenda-/Rule-/Schema-based
    (ii) Data-driven Generation
    (iii) Data-driven User Simulator
  (b) PostLLM:
    (i) With Fine-tuning
    (ii) With Prompting
    (iii) With Zero-shot
(4) Conversation Generation: Open Domain **(65 min)**
  (a) PreLLM:
    (i) Document-grounded
    (ii) Self-play Simulation
  (b) PostLLM:
    (i) With Fine-tuning
    (ii) With Prompting
(5) Conclusion and Future Direction **(10 min)**
(6) Q&A **(15 min)**

2.2 Content

Introduction. We start by introducing the audience to the basics of conversational systems, including TOD and ODD systems. We provide an overview of the components and concepts associated with TOD and ODD systems, ensuring that participants can grasp the necessary knowledge to follow the tutorial independently. We further discuss the data scarcity problem in creating dialogue systems, particularly in low-resource domains and languages and give an introduction to the proposed techniques to tackle this issue. Given that dialogue datasets may not be available for all languages and domains, we discuss dialogue data generation methods that leverage external resources such as unstructured text files, Knowledge Graphs (KG), and Large Language Models (LLMs).

Evaluation. Before delving into methods for creating dialogue data, it's essential to address how to evaluate the quality of this data. Synthetic conversation samples are primarily evaluated using two methods: intrinsic and extrinsic evaluation. The intrinsic evaluation approach directly assesses the quality of generated dialogue samples and comprises two categories: Automatic and Human evaluation. Conversely, in the extrinsic evaluation approach, the data augmentation method is evaluated on downstream tasks; i.e., when the synthetically generated dialogue data is used for training a dialogue model.

In this tutorial, our focus is mainly on the intrinsic approach. Under the automatic approach, we discuss both reference-based and reference-free methods. For reference-based methods, we explain techniques such as word overlap metrics, BERTScore [46], Coverage [40], and Coreference alignment [10]. Additionally, we explore simulation-based methods [34] and measures like Dist-n, Ent-n, Sent-BERT [31], and USR [24] for reference-free evaluation. For human evaluation, we discuss diverse methods such as Single-model per-turn, Single-model per-dialogue, Pairwise per-turn, and Pairwise per-dialogue methods [33]. We provide a comprehensive

discussion of the advantages and disadvantages of these evaluation methods, taking into account their suitability for different scenarios.

Conversation Generation: Task-oriented We next discuss the conversation generation methods for TOD systems. In such systems, the primary objective of the dialogue system is to understand the user's intent throughout a multi-turn conversation and subsequently provide relevant suggestions to assist the user in achieving their goal. However, accurately capturing the essential information from the user's utterances to ensure successful task completion requires meticulous attention and domain expertise [26].

At first, we give an overview of the problem by describing clear examples of TOD data, explaining its challenges, and showing commonly used datasets. Afterwards, we present current research by observing it from the perspective of leveraging LLMs. We begin by introducing methods that fit into the preLLM section. The first approach discussed embodies rule-guided generation methods [30, 32], which use slot-value schemes to automatically generate dialogues and annotations. A second approach focuses on data-driven generation [15, 44], where goal templates are filled in with values from a knowledge base that is further used in a pre-trained language models for dialogue generation. Another method revolves around user simulator [20, 37], where the dialogue is created by accessing two agents that each processes the dialogue history and generates a belief state for one side of the conversation. In the second part of TOD generation, we will focus on methods that leverage LLMs w.r.t. their generalization capabilities. Ordering them by needs of computational resources, we present dialogue generation using fine-tuning [25], prompting [18, 36], and zero-shot generalization [23].

Conversation Generation: Open Domain In this part, our focus is on the methods available for generating dialogue samples for an ODD system. We categorize current works into preLLM and postLLM methods. Under preLLM, we start by examining the document-grounded approach [4, 12, 13, 21, 40], which follows a pipeline method originating from synthetic QA pair generation [2, 17, 28]. This approach comprises four sequential stages: passage selection, span extraction, user & agent utterance generation, followed by a filtering process to maintain the quality of generated turns. Next, we explore self-play simulation methods involving a trained dialogue agent acting as both the user and the agent bot, generating conversation samples through interactions between two agents. Target-guided dialogue systems [8, 29, 34, 35, 39, 43, 47] are often leveraged for self-play simulation. Therefore, we delve into target-guided dialogue systems, focusing on concepts such as dialogue flow and dialogue planning strategies. These concepts enhance controllability in the generation process.

In the postLLM part, we explore methods that utilize either fine-tuning or prompting to generate conversation data. In the case of fine-tuning, we introduce the inpainting method [6], which defines the task as taking a partial dialog and generating a complete dialog. Additionally, for the prompting approach, we discuss various methods hinging on the three stages of (i) seed data generation, (ii) conversation generation, and (iii)filtering. Building on this pipeline, we elaborate current research in different domains, such as social dialogue [5, 14, 41], role-specified open-domain dialogue [3], math tutoring dialogue [22], persona-based dialogue [16], target-guided dialogue [38], and non-collaborative dialogue [1].

Conclusion and Future Direction We conclude the tutorial with an exploration of open research problems and future directions in the field.

3 PRESENTER BIOGRAPHY

Heydar Soudani is a first-year Ph.D. student at Radboud University's Institute of Computing and Information Sciences (iCIS), where he is being supervised jointly by Faegheh Hasibi and Evangelos Kanoulas. He holds a Bachelor's degree from Polytechnic of Tehran and a Master's degree from Sharif University of Technology. His research primarily focuses on conversational systems in low-resource domains and languages. Specifically, he is dedicated to the development of knowledge-grounded models that generate synthetic multi-turn conversation data.

Roxana Petcu is a starting Ph.D. student at the University of Amsterdam (UvA), supervised by prof Evangelos Kanoulas and dr Faegheh Hasibi. She completed her master's degree in Artificial Intelligence at UvA and obtained a BSc in Computer Science at Vrije University (Amsterdam). Her research focuses on data augmentation and generation for conversational agents. Her interests and previous experience also include speech recognition models, graph neural networks, and low-resource optimization.

Evangelos Kanoulas is a full professor of computer science at the University of Amsterdam, leading the Information Retrieval Lab at the Informatics Institute. His research lies in developing evaluation methods and algorithms for search, and recommendation, with a focus on learning robust models of language that can be used to understand noisy human language, retrieve textual data from large corpora, generate faithful and factual text, and converse with the user. Prior to joining the University of Amsterdam, he was a research scientist at Google and a Marie Curie fellow at the University of Sheffield. His research has been published at SIGIR, CIKM, WWW, WSDM, EMNLP, ACL, and other venues in the fields of IR and NLP. He has proposed and organized numerous search benchmarking competitions as part of the Text Retrieval Conference (TREC) and the Conference and Labs of the Evaluation Forum (CLEF). Furthermore, he is a member of the Ellis society (https://ellis.eu/).

Faegheh Hasibi is an assistant professor of information retrieval at the Institute of Computing and Information Sciences (iCIS) at Radboud University. Her research interests are at the intersection of Information Retrieval and Natural Language Processing, with a particular emphasis on conversational AI and semantic search systems. She explores various aspects, including knowledge-grounded conversational search, entity linking and retrieval, and the utilization of knowledge graphs for semantic search tasks. Her contributions to the field are published in renowned international conferences such as SIGIR, CIKM, COLING, and ICTIR and have been recognized by awards at the SIGIR and ICTIR conferences. She has given multiple invited talks and has extensive experience as a lecturer.

ACKNOWLEDGEMENTS

REFERENCES

[1] Zishan Ahmad, Suman Saurabh, Vaishakh Sreekanth Menon, Asif Ekbal, Roshni R. Ramnani, and Anutosh Maitra. 2023. INA: An Integrative Approach for Enhancing Negotiation Strategies with Reward-Based Dialogue System. *CoRR* abs/2310.18207 (2023).

[2] Chris Alberti, Daniel Andor, Emily Pitler, Jacob Devlin, and Michael Collins. 2019. Synthetic QA Corpora Generation with Roundtrip Consistency. In *ACL*. 6168–6173.

[3] Sanghwan Bae, Dong-Hyun Kwak, Sungdong Kim, Donghoon Ham, Soyoung Kang, Sang-Woo Lee, and Woo-Myoung Park. 2022. Building a Role Specified Open-Domain Dialogue System Leveraging Large-Scale Language Models. In *Proceedings of the 2022 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL.* Association for Computational Linguistics, 2128–2150.

[4] Jianzhu Bao, Rui Wang, Yasheng Wang, Aixin Sun, Yitong Li, Fei Mi, and Ruifeng Xu. 2023. A Synthetic Data Generation Framework for Grounded Dialogues. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics, ACL 2023, Toronto, Canada, July 9-14, 2023.* Association for Computational Linguistics, 10866–10882.

[5] Maximillian Chen, Alexandros Papangelis, Chenyang Tao, Seokhwan Kim, Andy Rosenbaum, Yang Liu, Zhou Yu, and Dilek Hakkani-Tur. 2023. PLACES: Prompting Language Models for Social Conversation Synthesis. In *Findings of the Association for Computational Linguistics: EACL 2023, Dubrovnik, Croatia, May 2-6, 2023.* Association for Computational Linguistics, 814–838.

[6] Zhuyun Dai, Arun Tejasvi Chaganty, Vincent Y. Zhao, Aida Amini, Qazi Mamunur Rashid, Mike Green, and Kelvin Guu. 2022. Dialog Inpainting: Turning Documents into Dialogs. In *Int'l Conf. on Machine Learning, ICML 2022, 17-23 July 2022, Baltimore, Maryland, USA (Proceedings of Machine Learning Research, Vol. 162).* PMLR, 4558–4586.

[7] Jeffrey Dalton, Sophie Fischer, Paul Owoicho, Filip Radlinski, Federico Rossetto, Johanne R. Trippas, and Hamed Zamani. 2022. Conversational Information Seeking: Theory and Application. In *SIGIR.* 3455–3458.

[8] Yang Deng, Wenqiang Lei, Lizi Liao, and Tat-Seng Chua. 2023. Prompting and Evaluating Large Language Models for Proactive Dialogues: Clarification, Target-guided, and Non-collaboration. *CoRR* abs/2305.13626 (2023).

[9] Zuohui Fu, Yikun Xian, Yongfeng Zhang, and Yi Zhang. 2020. Tutorial on Conversational Recommendation Systems. In *RecSys.* 751–753.

[10] Yifan Gao, Piji Li, Irwin King, and Michael R. Lyu. 2019. Interconnected Question Generation with Coreference Alignment and Conversation Flow Modeling. In *ACL.* 4853–4862.

[11] Chao Huang, Lianghao Xia, Xiang Wang, Xiangnan He, and Dawei Yin. 2022. Self-Supervised Learning for Recommendation. In *CIKM.* 5136–5139.

[12] Seonjeong Hwang, Yunsu Kim, and Gary Geunbae Lee. 2022. Multi-Type Conversational Question-Answer Generation with Closed-ended and Unanswerable Questions. In *AACL.* 169–177.

[13] Gangwoo Kim, Sungdong Kim, Kang Min Yoo, and Jaewoo Kang. 2022. Generating Information-Seeking Conversations from Unlabeled Documents. In *EMNLP.*

[14] Hyunwoo Kim, Jack Hessel, Liwei Jiang, Ximing Lu, Youngjae Yu, Pei Zhou, Ronan Le Bras, Malihe Alikhani, Gunhee Kim, Maarten Sap, and Yejin Choi. 2022. SODA: Million-scale Dialogue Distillation with Social Commonsense Contextualization. *CoRR* abs/2212.10465 (2022).

[15] Sungdong Kim, Minsuk Chang, and Sang-Woo Lee. 2021. NeuralWOZ: Learning to Collect Task-Oriented Dialogue via Model-Based Simulation. In *ACL.*

[16] Young-Jun Lee, Chae-Gyun Lim, Yunsu Choi, Ji-Hui Lm, and Ho-Jin Choi. 2022. PERSONACHATGEN: Generating Personalized Dialogues using GPT-3. In *Proceedings of the 1st Workshop on Customized Chat Grounding Persona and Knowledge.*

[17] Patrick Lewis, Yuxiang Wu, Linqing Liu, Pasquale Minervini, Heinrich Küttler, Aleksandra Piktus, Pontus Stenetorp, and Sebastian Riedel. 2021. PAQ: 65 Million Probably-Asked Questions and What You Can Do With Them. *TACL*, 1098–1115.

[18] Zekun Li, Wenhu Chen, Shiyang Li, Hong Wang, Jing Qian, and Xifeng Yan. 2022. Controllable Dialogue Simulation with In-context Learning. In *EMNLP.*

[19] Lizi Liao, Grace Hui Yang, and Chirag Shah. 2023. Proactive Conversational Agents. In *WSDM.* 1244–1247.

[20] Hsien-chin Lin, Nurul Lubis, Songbo Hu, Carel van Niekerk, Christian Geishauser, Michael Heck, Shutong Feng, and Milica Gasic. 2021. Domain-independent User Simulation with Transformers for Task-oriented Dialogue Systems. In *SIGDIAL.*

[21] Yen Ting Lin, Alexandros Papangelis, Seokhwan Kim, and Dilek Hakkani-Tur. 2022. Knowledge-Grounded Conversational Data Augmentation with Generative Conversational Networks. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue, SIGDIAL 2022, Edinburgh, UK, 07-09 September 2022.* Association for Computational Linguistics, 26–38.

[22] Jakub Macina, Nico Daheim, Sankalan Pal Chowdhury, Tanmay Sinha, Manu Kapur, Iryna Gurevych, and Mrinmaya Sachan. 2023. MathDial: A Dialogue Tutoring Dataset with Rich Pedagogical Properties Grounded in Math Reasoning Problems. *CoRR* abs/2305.14536 (2023).

[23] Shikib Mehri, Yasemin Altun, and Maxine Eskenazi. 2022. LAD: Language Models as Data for Zero-Shot Dialog. In *SIGDIAL.* 595–604.

[24] Shikib Mehri and Maxine Eskénazi. 2020. USR: An Unsupervised and Reference Free Evaluation Metric for Dialog Generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020,* Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault (Eds.). Association for Computational Linguistics, 681–707.

[25] Biswesh Mohapatra, Gaurav Pandey, Danish Contractor, and Sachindra Joshi. 2021. Simulated Chats for Building Dialog Systems: Learning to Generate Conversations from Instructions. In *EMNLP.* 1190–1203.

[26] Jinjie Ni, Tom Young, Vlad Pandelea, Fuzhao Xue, and Erik Cambria. 2023. Recent advances in deep learning based dialogue systems: a systematic survey. *Artif. Intell. Rev.,* 3055–3155.

[27] Jinjie Ni, Tom Young, Vlad Pandelea, Fuzhao Xue, and Erik Cambria. 2023. Recent advances in deep learning based dialogue systems: a systematic survey. *Artif. Intell. Rev.* 4 (2023), 3055–3155.

[28] Raul Puri, Ryan Spring, Mohammad Shoeybi, Mostofa Patwary, and Bryan Catanzaro. 2020. Training Question Answering Models From Synthetic Data. In *EMNLP.* 5811–5826.

[29] Jinghui Qin, Zheng Ye, Jianheng Tang, and Xiaodan Liang. 2020. Dynamic Knowledge Routing Network for Target-Guided Open-Domain Conversation. In *AAAI.* 8657–8664.

[30] Abhinav Rastogi, Xiaoxue Zang, Srinivas Sunkara, Raghav Gupta, and Pranav Khaitan. 2020. Towards Scalable Multi-Domain Conversational Agents: The Schema-Guided Dialogue Dataset. In *AAAI.* 8689–8696.

[31] Nils Reimers and Iryna Gurevych. 2019. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. In *Proceedings of the 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int'l Joint Conf. on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China,* Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (Eds.). Association for Computational Linguistics, 3980–3990.

[32] Pararth Shah, Dilek Hakkani-Tür, Gökhan Tür, Abhinav Rastogi, Ankur Bapna, Neha Nayak, and Larry P. Heck. 2018. Building a Conversational Agent Overnight with Dialogue Self-Play. *Arxiv.*

[33] Eric Michael Smith, Orion Hsu, Rebecca Qian, Stephen Roller, Y-Lan Boureau, and Jason Weston. 2022. Human Evaluation of Conversations is an Open Problem: comparing the sensitivity of various methods for evaluating dialogue agents. In *Proceedings of the 4th Workshop on NLP for Conversational AI, ConvAI@ACL 2022, Dublin, Ireland, May 27, 2022.* Association for Computational Linguistics, 77–97.

[34] Jianheng Tang, Tiancheng Zhao, Chenyan Xiong, Xiaodan Liang, Eric Xing, and Zhiting Hu. 2019. Target-Guided Open-Domain Conversation. In *ACL.*

[35] Zee Hen Tang and Mi-Yen Yeh. 2023. EAGLE: Enhance Target-Oriented Dialogs by Global Planning and Topic Flow Integration. In *Proceedings of the 32nd ACM Int'l Conf. on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023.* ACM, 2402–2411.

[36] Silvia Terragni, Modestas Filipavicius, Nghia Khau, Bruna Guedes, André Manso, and Roland Mathis. 2023. In-Context Learning User Simulators for Task-Oriented Dialog Systems. *arXiv.*

[37] Bo-Hsiang Tseng, Yinpei Dai, Florian Kreyssig, and Bill Byrne. 2021. Transferable Dialogue Systems and User Simulators. 152–166.

[38] Jian Wang, Yi Cheng, Dongding Lin, Chak Tou Leong, and Wenjie Li. 2023. Target-oriented Proactive Dialogue Systems with Personalization: Problem Formulation and Dataset Curation. *CoRR* abs/2310.07397 (2023).

[39] Jian Wang, Dongding Lin, and Wenjie Li. 2023. Dialogue Planning via Brownian Bridge Stochastic Process for Goal-directed Proactive Dialogue. In *Findings of the Association for Computational Linguistics: ACL 2023, Toronto, Canada, July 9-14, 2023.* Association for Computational Linguistics, 370–387.

[40] Qingyang Wu, Song Feng, Derek Chen, Sachindra Joshi, Luis Lastras, and Zhou Yu. 2022. DG2: Data Augmentation Through Document Grounded Dialogue Generation. In *SIGDIAL.* 204–216.

[41] Canwen Xu, Daya Guo, Nan Duan, and Julian J. McAuley. 2023. Baize: An Open-Source Chat Model with Parameter-Efficient Tuning on Self-Chat Data. *CoRR* abs/2304.01196 (2023).

[42] Diyi Yang, Ankur Parikh, and Colin Raffel. 2022. Learning with Limited Text Data. In *ACL.* 28–31.

[43] Zhitong Yang, Bo Wang, Jinfeng Zhou, Yue Tan, Dongming Zhao, Kun Huang, Ruifang He, and Yuexian Hou. 2022. TopKG: Target-oriented Dialog via Global Planning on Knowledge Graph. In *COLING.* 745–755.

[44] Kang Min Yoo, Hanbit Lee, Franck Dernoncourt, Trung Bui, Walter Chang, and Sang-goo Lee. 2020. Variational Hierarchical Dialog Autoencoder for Dialog State Tracking Data Augmentation. In *EMNLP.* 3406–3425.

[45] Hamed Zamani, Johanne R. Trippas, Jeff Dalton, and Filip Radlinski. 2022. Conversational Information Seeking. *Arxiv.*

[46] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating Text Generation with BERT. In *8th Int'l Conf. on Learning Representations, ICLR, Addis Ababa, Ethiopia.*

[47] Peixiang Zhong, Yong Liu, Hao Wang, and Chunyan Miao. 2021. Keyword-Guided Neural Conversational Model. In *AAAI.* 14568–14576.