Goal Awareness for Conversational AI: Proactivity, Non-collaborativity, and Beyond

Yang Deng[1], Wenqiang Lei[2], Minlie Huang[3], Tat-Seng Chua[4]
[1] The Chinese University of Hong Kong   [2]Sichuan University
[3] Tsinghua University   [4] National University of Singapore
ydeng@se.cuhk.edu.hk wenqianglei@gmail.com
aihuang@tsinghua.edu.cn chuats@comp.nus.edu.sg

1 Introduction

Tutorial Description   Conversational systems are envisioned to provide social support or functional service to human users via natural language interactions. Conventional conversation researches mainly focus on the response-ability of the system, such as dialogue context understanding and response generation, but overlooks the design of an essential property in intelligent conversations, *i.e.*, goal awareness. The awareness of goals means the state of not only being responsive to the users but also aware of the target conversational goal and capable of leading the conversation towards the goal, which is a significant step towards higher-level intelligence and artificial consciousness. It can not only largely improve user engagement and service efficiency in the conversation, but also empower the system to handle more complicated conversation tasks that involve strategical and motivational interactions. In this tutorial, we will introduce the recent advances on the design of agent's awareness of goals in a wide range of conversational systems.

Type of Tutorial   Cutting-edge

Targeted Audience   Target audiences are researchers and practitioners who interested in natural language processing and human-computer interaction. The audience will learn about the state-of-the-art research in conversational AI and the cutting-edge designs of agent's awareness in various conversational systems.

Suggested Duration   Half day (3 hours)

2 Tutorial Outline

Part I: Preliminary (20 minutes)

Conversational agents are generally envisioned to achieve the conversational goal by providing social support or functional service to human users via natural language interactions. In terms of the goal, Part I will present a brief overview of the widely-studied problems and correponding main-stream approaches in several typical conversational systems, including open-domain dialogue (ODD) systems (Zhang et al., 2018a; Li et al., 2017; Roller et al., 2021), task-oriented dialogue (TOD) systems (Budzianowski et al., 2018; Lei et al., 2018; Su et al., 2022), conversational question answering (CQA) systems (Choi et al., 2018; Reddy et al., 2019; Anantha et al., 2021; Qiu et al., 2021), and conversational recommender systems (CRS) (Li et al., 2018; Deng et al., 2021; Wang et al., 2022).

Part II: Proactive Conversational Systems (50 minutes)

As opposed to responding to users, proactivity is the most prominent feature of goal awareness in conversational systems, which can improve the collaboration between the users and system towards the ultimate conversation goal. Derived from the definition of proactivity in organizational behaviors (Grant and Ashford, 2008) and its dictionary definitions, conversational agents' proactivity can be defined as the capability to create or control the conversation by taking the initiative and anticipating impacts on themselves or human users (Deng et al., 2023a). In this part, we will provide a comprehensive introduction about such efforts on the design of agent's proactivity that span various task formulations and application scenarios. In specific, we categorize them in three directions according to the application scenario, and plan to discuss their research problems and methods as follows:

- **Topic Shifting and Planning in Open-domain Dialogues**   The goal of OOD systems is to maintain engaging social conversations with users. Proactive OOD systems can consciously change topics (Rachna et al., 2021; Xie et al., 2021) and lead directions (Tang et al., 2019; Wu et al., 2019; Yang et al., 2022) for improving user engagement in the conversation. We will present the existing methods for topic shifting and planning in open-domain dialogues, including graph-based topic

planning (Qin et al., 2020; Zhong et al., 2021; Xu et al., 2020; Ni et al., 2022), responding plan generation (Kishinami et al., 2022), and learning from interactions with users (Lei et al., 2022).

- **Additional Information Delivery in Task-oriented Dialogues** The goal of TOD systems is to provide functional service for users, such as making reservations or managing schedule. The proactivity in TOD systems is firstly defined as the capability of consciously providing additional information that is not requested by but useful to the users (Balaraman and Magnini, 2020a,b), which can improve the quality and effectiveness of conveying functional service in the conversation. We will introduce the recent studies of proactive TOD systems with various designs. For instance, Sun et al. (2021) add topical chit-chats into the responses for TODs. Chen et al. (2022c) enrich task-oriented dialogues with relevant entity knowledge.

- **Uncertainty Elimination in Information-seeking Dialogues** The goal of CIS systems (Zamani et al., 2022) is to fulfill the user's information needs and its typical applications include conversational search, conversational recommendation, and conversational question answering. Conventional CIS systems assume that users always convey clear information requests, while the user queries, in reality, are often brief and succinct. Recent years have witnessed several advances on developing proactive CIS systems that can consciously eliminate the uncertainty for more efficient and precise information seeks by initiating a subdialogue. Such a subdialogue can either clarify the ambiguity of the query or question in conversational search (Aliannejadi et al., 2019, 2021; Zamani et al., 2020) and conversation question answering (Guo et al., 2021; Deng et al., 2022a), or elicit the user preference in conversational recommendation (Zhang et al., 2018b; Lei et al., 2020a,b; Deng et al., 2022b).

Part III: Non-collaborative Conversational Systems (40 minutes)

Most of existing conversational systems are built upon the assumption that the users willingly collaborate with the conversational agent to reach the mutual goal. However, this assumption may not always hold in some real-world scenarios, where the users and the system do not share the same

goal (He et al., 2018; Wang et al., 2019) or the users are not willing to coordinate with the agent (Yang et al., 2019; Kim et al., 2022). In these cases, the conversational agent requires another feature of goal awareness, *i.e.*, non-collaborativity (Li et al., 2020; Zhou et al., 2020), which means the capability of handling both in-goal and off-goal dialogues appropriately for ultimately leading back to the system's goal. In this part, we will categorize the non-collaborative settings into two groups as follows and cover their to-date work respectively.

- **The users and the system do not share the same goal.** Typical applications include persuasion dialogues (Wang et al., 2019), negotiation dialogues (He et al., 2018; Chawla et al., 2021), and anti-scam dialogues (Li et al., 2020). We will present the approaches for the system to consciously mitigate and resolve the conflict goals with users, including dialogue strategy learning (Dutt et al., 2021; Yamaguchi et al., 2021; Joshi et al., 2021), user personality modeling (Shi et al., 2021; Yang et al., 2021), and response style transfer (Mishra et al., 2022; Wu et al., 2021).

- **The users are not willing to coordinate with the agent.** Example scenarios include calming down the emotional users before solving their problems (Liu et al., 2021b), managing the users' complaints before providing service (Yang et al., 2019), and handling problematic content during the conversations (Kim et al., 2022). We will introduce the pioneering studies for the system to consciously deal with non-collaborative users during the conversation, including emotion cause analysis (Cheng et al., 2022; Deng et al., 2023c), user satisfaction estimation (Liu et al., 2021a; Deng et al., 2022c), and safe response generation (Baheti et al., 2021; Ung et al., 2022).

Part IV: Multi-goal Conversational Systems (30 minutes)

All the aforementioned conversational systems assume that users always know what they want and the system solely targets at reaching a certain goal, such as chit-chat, question answering, recommendation, etc. The system with a higher level of agent's awareness of goals should also be capable of handling conversations with multiple and various goals. As for multi-goal conversational systems (Liu et al., 2022; Deng et al., 2022d), the agent is expected to consciously discover users' intentions and naturally lead user-engaged dialogues

with multiple conversation goals. We will cover the newly proposed problems in multi-goal conversational systems with their corresponding data resources (Sun et al., 2021; Zhao et al., 2022; Young et al., 2022; Chiu et al., 2022). Then we will discuss two problem settings of multi-goal conversational systems with corresponding state-of-the-art approaches: (i) The goal sequence is predefined (Bai et al., 2021; Zhang et al., 2021b), and (ii) The next goal needs to be predicted (Liu et al., 2020; Chen et al., 2022b; Deng et al., 2022d).

Part V: Open Challenges for Conversational Agents' Awareness and Beyond (40 minutes)

In the last part, we will discuss the main open challenges in developing agent's awareness in conversational systems and several potential research directions for future studies.

- **Evaluation for Conversational Agent's Awareness** The development of robust evaluation protocols has already been a long-standing problem for different kinds of conversational systems (Zhang et al., 2021a; Peng et al., 2021; Li et al., 2022b). The evaluation for conversational agent's awareness is a more challenging problem, since it is involved the evaluation not only from the perspective of natural language, but also from the perspectives of human-computer interaction, sociology, psychology, etc. We will cover the latest studies for shedding some lights on this topic, inclusive of popular metrics such as goal completion and user satisfaction (Liu et al., 2020; Lei et al., 2022; Gupta et al., 2022), and model-based methods such as user simulator (Zhang and Balog, 2020; Sekulic et al., 2022).

- **Ethics for Conversational Agent's Awareness** Although existing designs of agent's awareness of goals in conversational systems generally aim at social goodness (Wang et al., 2019; Liu et al., 2021b; Kim et al., 2022), it is inevitably a double-edged sword that can be used for good or evil. For responsible NLP researches, we will discuss several important aspects of ethical issues in conscious conversational systems: (i) Factuality: Factual incorrectness and hallucination of knowledge are common in conversational systems (Dziri et al., 2022; Honovich et al., 2021). When enabling the conversational agent with awareness, it becomes more crucial to guarantee the factuality of the system-provided information (Chen et al., 2022a). (ii) Safety: Besides

general dialogue safety problems, such as toxic language and social bias (Saveski et al., 2021; Barikeri et al., 2021), conscious conversational systems need to pay more attentions to the aggressiveness issue during the non-collaborative conversations (Kim et al., 2022; Hu et al., 2022). (iii) Privacy: The privacy issue is overlooked in current studies on conversational systems (Li et al., 2022a; Shi et al., 2022), but the agent's awareness raises concerns about how these conversational systems handle personal information obtained from the users. Furthermore, we will introduce some recent released resources that can be adopted for studying this topic (Ziems et al., 2022; Sun et al., 2022; Kim et al., 2022).

- **Agent's Awareness in LLM-based Conversational AI** Large Language Models (LLMs) have been demonstrated to be powerful of handling various NLP tasks in the form of conversations, such as ChatGPT (Schulman et al., 2022), LaMDA (Thoppilan et al., 2022), BlenderBot (Shuster et al., 2022), etc. However, these applications are typically designed to follow the user's instructions and intents. There are still several limitations that attribute to the lack of agent's awareness, such as passively providing randomly-guessed answers to ambiguous user queries, failing to refuse or handle problematic user requests that may exhibit harmful or biased conversations, etc. In addition, they also fall short of interacting under non-collaborative or system-oriented settings. Therefore, we will discuss the role of LLMs in goal awareness for conversational AI with some latest studies (Huang et al., 2022; Deng et al., 2023b; Yao et al., 2022).

3 Presenters

Yang Deng is a Research Fellow in National University of Singapore. His research lies in natural language processing and information retrieval, especially for dialogue and QA systems. He has published over 20 papers at top venues such as ACL, EMNLP, SIGIR, WWW, TKDE, and TOIS. Additional information is available at `https://dengyang17.github.io`.

Wenqiang Lei is a Professor in Sichuan University. His research interests focus on conversational AI, including conversational recommendation, dialogue and QA systems. He has published relevant papers at top venues such as ACL, EMNLP, KDD, SIGIR, TOIS, and received the ACM MM 2020

best paper award. He has given tutorials on the topic of conversational recommendation at RecSys 2021, SIGIR 2020, and co-organized special issues about conversational information seeking on ACM Trans. on Web. Specifically, his tutorial on SIGIR 2020 accepts over 1600 audiences, being one of the most popular tutorials in SIGIR 2020. Additional information is available at `https://sites.google.com/view/wenqianghome/home`.

Minlie Huang is an Associate Professor with the Department of Computer Science and Technology, Tsinghua University. He has authored or coauthored more than 100 papers in premier conferences and journals (ACL, EMNLP, TACL, etc). His research interests include natural language processing, particularly in dialog systems, reading comprehension, and sentiment analysis. He is an editor of TACL, CL, TNNLS, the Area Chair or SAC of ACL/EMNLP for more than 10 times. He is the recipient of IJCAI 2018 distinguished paper award, a nominee of ACL 2019 best demo papers, and SIGDIAL 2020 best paper award. Additional information is available at `http://coai.cs.tsinghua.edu.cn/hml`.

Tat-Seng Chua is the KITHCT Chair Professor with the School of Computing, National University of Singapore. His main research interest include multimedia information retrieval and social media analytics. He is the 2015 winner of the prestigious ACM SIGMM Technical Achievement Award and receives the best papers (or candidates) over 10 times in top conferences (SIGIR, WWW, MM, etc). He serves as the general co-chair of top conferences multiple times (MM 2005, SIGIR 2008, WSDM 2023, etc), and the editors of multiple journals (TOIS, TMM, etc). He has given invited keynote talks at multiple top conferences, including the recent one on the topic of multimodal conversational search and recommendation. Additional information is available at `https://www.chuatatseng.com/`.

4 Reading Lists

Previous Tutorials:
(Chen et al., 2017b) ACL 2017 - Deep Learning for Dialogue Systems;
(Su et al., 2018) NAACL 2018 - Deep Learning for Conversational AI;
(Gao et al., 2018) ACL 2018/SIGIR 2018 - Neural Approaches to Conversational AI;
(Gao et al., 2020) SIGIR 2020 - Recent Advances in Conversational Information Retrieval;
(Dalton et al., 2022) SIGIR 2022 - Conversational Information Seeking: Theory and Application.
(Liao et al., 2023) WSDM 2023 - Proactive Conversational Agents.

Related Surveys or Book Chapters:
(Chen et al., 2017a) A Survey on Dialogue Systems: Recent Advances and New Frontiers;
(Gao et al., 2019) Neural Approaches to Conversational AI;
(Huang et al., 2020) Challenges in Building Intelligent Open-domain Dialog Systems;
(Zamani et al., 2022) Conversational Information Seeking;
(Gao et al., 2022) Neural Approaches to Conversational Information Retrieval;
(Yan et al., 2022) Deep Learning for Dialogue Systems: Chit-Chat and Beyond.

5 Other Tutorial Information

Breadth and Diversity Considerations According to the representative set of papers listed in the selected bibliography, the concerned work in this tutorial will contain only 10%-15% of work that involves at least one of the four presenters. The rest of the tutorial will present a comprehensive overview of the tutorial topic by discussing the related work as much as possible from other researchers. The discussed approaches are problem-driven and language-agnostic, which means that the introduced content are generally applicable to all languages. The techniques are also not limited to a certain type of dialogues and can be generalized to diverse conversational systems.

Ethical Considerations Artificial consciousness is a broad and essential topic towards "Strong AI" in the whole AI community (Searle, 1992), which can and should be used for social goodness, but inevitably comes with potential risks. In fact, the awareness of goals is just one of the cognitive aspects of consciousness (Baars, 1993). As part of this tutorial, we will provide a specific section for discussing the ethical considerations and designs for agent's awareness in conversational systems. This tutorial also provides the opportunity to arouse discussions on how far we can and should go for agent's consciousness in conversational AI from the view of ethical and responsible NLP researches.

Open Access of Materials All tutorial materials will be made publicly available.

References

Mohammad Aliannejadi, Julia Kiseleva, Aleksandr Chuklin, Jeff Dalton, and Mikhail S. Burtsev. 2021. Building and evaluating open-domain dialogue corpora with clarifying questions. In *EMNLP 2021*, pages 4473–4484.

Mohammad Aliannejadi, Hamed Zamani, Fabio Crestani, and W. Bruce Croft. 2019. Asking clarifying questions in open-domain information-seeking conversations. In *SIGIR 2019*, pages 475–484.

Raviteja Anantha, Svitlana Vakulenko, Zhucheng Tu, Shayne Longpre, Stephen Pulman, and Srinivas Chappidi. 2021. Open-domain question answering goes conversational via question rewriting. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021*, pages 520–534.

Bernard J Baars. 1993. *A cognitive theory of consciousness*. Cambridge University Press.

Ashutosh Baheti, Maarten Sap, Alan Ritter, and Mark O. Riedl. 2021. Just say no: Analyzing the stance of neural dialogue generation in offensive contexts. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021*, pages 4846–4862.

Jiaqi Bai, Ze Yang, Xinnian Liang, Wei Wang, and Zhoujun Li. 2021. Learning to copy coherent knowledge for response generation. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021*, pages 12535–12543.

Vevake Balaraman and Bernardo Magnini. 2020a. Investigating proactivity in task-oriented dialogues. In *Proceedings of the Seventh Italian Conference on Computational Linguistics, CLiC-it 2020*, volume 2769.

Vevake Balaraman and Bernardo Magnini. 2020b. Proactive systems and influenceable users: Simulating proactivity in task-oriented dialogues. In *Proc. 24th Workshop Semantics Pragmatics Dialogue (SEMDIAL)*, pages 1–12.

Soumya Barikeri, Anne Lauscher, Ivan Vulic, and Goran Glavas. 2021. Redditbias: A real-world resource for bias evaluation and debiasing of conversational language models. 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*, pages 1941–1955.

Pawel Budzianowski, Tsung-Hsien Wen, Bo-Hsiang Tseng, Iñigo Casanueva, Stefan Ultes, Osman Ramadan, and Milica Gasic. 2018. Multiwoz - A large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 5016–5026.

Kushal Chawla, Jaysa Ramirez, Rene Clever, Gale M. Lucas, Jonathan May, and Jonathan Gratch. 2021. Casino: A corpus of campsite negotiation dialogues for automatic negotiation systems. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021*, pages 3167–3185.

Hongshen Chen, Xiaorui Liu, Dawei Yin, and Jiliang Tang. 2017a. A survey on dialogue systems: Recent advances and new frontiers. *SIGKDD Explor.*, 19(2):25–35.

Maximillian Chen, Weiyan Shi, Feifan Yan, Ryan Hou, Jingwen Zhang, Saurav Sahay, and Zhou Yu. 2022a. Seamlessly integrating factual information and social content with persuasive dialogue. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing, AACL/IJCNLP 2022*, pages 399–413.

Yun-Nung Chen, Asli Celikyilmaz, and Dilek Hakkani-Tür. 2017b. Deep learning for dialogue systems. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts*, pages 8–14.

Zhi Chen, Lu Chen, Bei Chen, Libo Qin, Yuncong Liu, Su Zhu, Jian-Guang Lou, and Kai Yu. 2022b. Unidu: Towards A unified generative dialogue understanding framework. In *Proceedings of the 23rd Annual Meeting of the Special Interest Group on Discourse and Dialogue, SIGDIAL 2022*, pages 442–455.

Zhiyu Chen, Bing Liu, Seungwhan Moon, Chinnadhurai Sankar, Paul A. Crook, and William Yang Wang. 2022c. KETOD: knowledge-enriched task-oriented dialogue. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 2581–2593.

Yi Cheng, Wenge Liu, Wenjie Li, Jiashuo Wang, Ruihui Zhao, Bang Liu, Xiaodan Liang, and Yefeng Zheng. 2022. Improving multi-turn emotional support dialogue generation with lookahead strategy planning. *CoRR*, abs/2210.04242.

Ssu Chiu, Maolin Li, Yen-Ting Lin, and Yun-Nung Chen. 2022. Salesbot: Transitioning from chit-chat to task-oriented dialogues. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, ACL 2022*, pages 6143–6158.

Eunsol Choi, He He, Mohit Iyyer, Mark Yatskar, Wentau Yih, Yejin Choi, Percy Liang, and Luke Zettlemoyer. 2018. Quac: Question answering in context. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2174–2184.

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 '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3455–3458.

Yang Deng, Wenqiang Lei, Wai Lam, and Tat-Seng Chua. 2023a. A survey on proactive dialogue systems: Problems, methods, and prospects. *CoRR*, abs/2305.02750.

Yang Deng, Wenqiang Lei, Lizi Liao, and Tat-Seng Chua. 2023b. Prompting and evaluating large language models for proactive dialogues: Clarification, target-guided, and non-collaboration. *CoRR*, abs/2305.13626.

Yang Deng, Wenqiang Lei, Wenxuan Zhang, Wai Lam, and Tat-Seng Chua. 2022a. PACIFIC: towards proactive conversational question answering over tabular and textual data in finance. *CoRR*, abs/2210.08817.

Yang Deng, Yaliang Li, Bolin Ding, and Wai Lam. 2022b. Leveraging long short-term user preference in conversational recommendation via multi-agent reinforcement learning. *IEEE Transactions on Knowledge and Data Engineering*.

Yang Deng, Yaliang Li, Fei Sun, Bolin Ding, and Wai Lam. 2021. Unified conversational recommendation policy learning via graph-based reinforcement learning. In *SIGIR '21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1431–1441.

Yang Deng, Wenxuan Zhang, Wai Lam, Hong Cheng, and Helen Meng. 2022c. User satisfaction estimation with sequential dialogue act modeling in goal-oriented conversational systems. In *WWW '22: The ACM Web Conference 2022*, pages 2998–3008.

Yang Deng, Wenxuan Zhang, Weiwen Xu, Wenqiang Lei, Tat-Seng Chua, and Wai Lam. 2022d. A unified multi-task learning framework for multi-goal conversational recommender systems. *CoRR*, abs/2204.06923.

Yang Deng, Wenxuan Zhang, Yifei Yuan, and Wai Lam. 2023c. Knowledge-enhanced mixed-initiative dialogue system for emotional support conversations. *CoRR*, abs/2305.10172.

Ritam Dutt, Sayan Sinha, Rishabh Joshi, Surya Shekhar Chakraborty, Meredith Riggs, Xinru Yan, Haogang Bao, and Carolyn P. Rosé. 2021. Resper: Computationally modelling resisting strategies in persuasive conversations. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021*, pages 78–90.

Nouha Dziri, Sivan Milton, Mo Yu, Osmar R. Zaïane, and Siva Reddy. 2022. On the origin of hallucinations in conversational models: Is it the datasets or the models? In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022*, pages 5271–5285.

Jianfeng Gao, Michel Galley, and Lihong Li. 2018. Neural approaches to conversational AI. In *Proceedings of ACL 2018, Tutorial Abstracts*, pages 2–7.

Jianfeng Gao, Michel Galley, and Lihong Li. 2019. Neural approaches to conversational AI. *Found. Trends Inf. Retr.*, 13(2-3):127–298.

Jianfeng Gao, Chenyan Xiong, and Paul Bennett. 2020. Recent advances in conversational information retrieval. In *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020*, pages 2421–2424.

Jianfeng Gao, Chenyan Xiong, Paul Bennett, and Nick Craswell. 2022. Neural approaches to conversational information retrieval. *CoRR*, abs/2201.05176.

Adam M Grant and Susan J Ashford. 2008. The dynamics of proactivity at work. *Research in organizational behavior*, 28:3–34.

Meiqi Guo, Mingda Zhang, Siva Reddy, and Malihe Alikhani. 2021. Abg-coqa: Clarifying ambiguity in conversational question answering. In *AKBC 2021*.

Prakhar Gupta, Harsh Jhamtani, and Jeffrey P. Bigham. 2022. Target-guided dialogue response generation using commonsense and data augmentation. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 1301–1317.

He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2333–2343.

Or Honovich, Leshem Choshen, Roee Aharoni, Ella Neeman, Idan Szpektor, and Omri Abend. 2021. q^2: Evaluating factual consistency in knowledge-grounded dialogues via question generation and question answering. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021*, pages 7856–7870.

Zhiqiang Hu, Roy Ka-Wei Lee, and Nancy F. Chen. 2022. Are current task-oriented dialogue systems able to satisfy impolite users? *CoRR*, abs/2210.12942.

Minlie Huang, Xiaoyan Zhu, and Jianfeng Gao. 2020. Challenges in building intelligent open-domain dialog systems. *ACM Trans. Inf. Syst.*, 38(3):21:1–21:32.

Wenlong Huang, Pieter Abbeel, Deepak Pathak, and Igor Mordatch. 2022. Language models as zero-shot planners: Extracting actionable knowledge for embodied agents. In *International Conference on Machine Learning, ICML 2022*, volume 162, pages 9118–9147.

Rishabh Joshi, Vidhisha Balachandran, Shikhar Vashishth, Alan W. Black, and Yulia Tsvetkov. 2021. Dialograph: Incorporating interpretable strategy-graph networks into negotiation dialogues. In *9th International Conference on Learning Representations, ICLR 2021*.

Hyunwoo Kim, Youngjae Yu, Liwei Jiang, Ximing Lu, Daniel Khashabi, Gunhee Kim, Yejin Choi, and Maarten Sap. 2022. Prosocialdialog: A prosocial backbone for conversational agents. *CoRR*, abs/2205.12688.

Yosuke Kishinami, Reina Akama, Shiki Sato, Ryoko Tokuhisa, Jun Suzuki, and Kentaro Inui. 2022. Target-guided open-domain conversation planning. In *COLING 2022*, pages 660–668.

Wenqiang Lei, Xiangnan He, Yisong Miao, Qingyun Wu, Richang Hong, Min-Yen Kan, and Tat-Seng Chua. 2020a. Estimation-action-reflection: Towards deep interaction between conversational and recommender systems. In *WSDM 2020*, pages 304–312.

Wenqiang Lei, Xisen Jin, Min-Yen Kan, Zhaochun Ren, Xiangnan He, and Dawei Yin. 2018. Sequicity: Simplifying task-oriented dialogue systems with single sequence-to-sequence architectures. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018*, pages 1437–1447.

Wenqiang Lei, Gangyi Zhang, Xiangnan He, Yisong Miao, Xiang Wang, Liang Chen, and Tat-Seng Chua. 2020b. Interactive path reasoning on graph for conversational recommendation. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2073–2083.

Wenqiang Lei, Yao Zhang, Feifan Song, Hongru Liang, Jiaxin Mao, Jiancheng Lv, Zhenglu Yang, and Tat-Seng Chua. 2022. Interacting with non-cooperative user: A new paradigm for proactive dialogue policy. In *SIGIR '22: The 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 212–222.

Haoran Li, Yangqiu Song, and Lixin Fan. 2022a. You don't know my favorite color: Preventing dialogue representations from revealing speakers' private personas. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022*, pages 5858–5870.

Huihan Li, Tianyu Gao, Manan Goenka, and Danqi Chen. 2022b. Ditch the gold standard: Re-evaluating conversational question answering. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, ACL 2022*, pages 8074–8085.

Raymond Li, Samira Ebrahimi Kahou, Hannes Schulz, Vincent Michalski, Laurent Charlin, and Chris Pal. 2018. Towards deep conversational recommendations. In *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018*, pages 9748–9758.

Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. Dailydialog: A manually labelled multi-turn dialogue dataset. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing, IJCNLP 2017*, pages 986–995.

Yu Li, Kun Qian, Weiyan Shi, and Zhou Yu. 2020. End-to-end trainable non-collaborative dialog system. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020*, pages 8293–8302.

Lizi Liao, Grace Hui Yang, and Chirag Shah. 2023. Proactive conversational agents. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining, WSDM 2023*, pages 1244–1247.

Jiawei Liu, Kaisong Song, Yangyang Kang, Guoxiu He, Zhuoren Jiang, Changlong Sun, Wei Lu, and Xiaozhong Liu. 2021a. A role-selected sharing network for joint machine-human chatting handoff and service satisfaction analysis. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, EMNLP 2021*, pages 9731–9741.

Siyang Liu, Chujie Zheng, Orianna Demasi, Sahand Sabour, Yu Li, Zhou Yu, Yong Jiang, and Minlie Huang. 2021b. Towards emotional support dialog systems. 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*, pages 3469–3483.

Zeming Liu, Haifeng Wang, Zheng-Yu Niu, Hua Wu, Wanxiang Che, and Ting Liu. 2020. Towards conversational recommendation over multi-type dialogs. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020*, pages 1036–1049.

Zeming Liu, Jun Xu, Zeyang Lei, Haifeng Wang, Zheng-Yu Niu, and Hua Wu. 2022. Where to go for the holidays: Towards mixed-type dialogs for clarification of user goals. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, ACL 2022*, pages 1024–1034.

Kshitij Mishra, Azlaan Mustafa Samad, Palak Totala, and Asif Ekbal. 2022. PEPDS: A polite and empathetic persuasive dialogue system for charity donation. In *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022*, pages 424–440.

Jinjie Ni, Vlad Pandelea, Tom Young, Haicang Zhou, and Erik Cambria. 2022. Hitkg: Towards goal-oriented conversations via multi-hierarchy learning. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022*, pages 11112–11120.

Baolin Peng, Chunyuan Li, Zhu Zhang, Chenguang Zhu, Jinchao Li, and Jianfeng Gao. 2021. RADDLE: an evaluation benchmark and analysis platform for robust task-oriented dialog systems. 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*, pages 4418–4429.

Jinghui Qin, Zheng Ye, Jianheng Tang, and Xiaodan Liang. 2020. Dynamic knowledge routing network for target-guided open-domain conversation. In *AAAI 2020*, pages 8657–8664.

Minghui Qiu, Xinjing Huang, Cen Chen, Feng Ji, Chen Qu, Wei Wei, Jun Huang, and Yin Zhang. 2021. Reinforced history backtracking for conversational question answering. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021*, pages 13718–13726.

Konigari Rachna, Saurabh Ramola, Vijay Vardhan Alluri, and Manish Shrivastava. 2021. Topic shift detection for mixed initiative response. In *SIGdial 20211*, pages 161–166.

Siva Reddy, Danqi Chen, and Christopher D. Manning. 2019. Coqa: A conversational question answering challenge. *Trans. Assoc. Comput. Linguistics*, 7:249–266.

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. 2021. Recipes for building an open-domain chatbot. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021,*, pages 300–325.

Martin Saveski, Brandon Roy, and Deb Roy. 2021. The structure of toxic conversations on twitter. In *WWW '21: The Web Conference 2021*, pages 1086–1097.

J Schulman, B Zoph, C Kim, J Hilton, J Menick, J Weng, JFC Uribe, L Fedus, L Metz, M Pokorny, et al. 2022. Chatgpt: Optimizing language models for dialogue.

John R Searle. 1992. *The rediscovery of the mind*. MIT press.

Ivan Sekulic, Mohammad Aliannejadi, and Fabio Crestani. 2022. Evaluating mixed-initiative conversational search systems via user simulation. In *WSDM '22: The Fifteenth ACM International Conference on Web Search and Data Mining*, pages 888–896.

Weiyan Shi, Aiqi Cui, Evan Li, Ruoxi Jia, and Zhou Yu. 2022. Selective differential privacy for language modeling. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL 2022*, pages 2848–2859.

Weiyan Shi, Yu Li, Saurav Sahay, and Zhou Yu. 2021. Refine and imitate: Reducing repetition and inconsistency in persuasion dialogues via reinforcement learning and human demonstration. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3478–3492.

Kurt Shuster, Jing Xu, Mojtaba Komeili, Da Ju, Eric Michael Smith, Stephen Roller, Megan Ung, Moya Chen, Kushal Arora, Joshua Lane, Morteza Behrooz, William Ngan, Spencer Poff, Naman Goyal, Arthur Szlam, Y-Lan Boureau, Melanie Kambadur, and Jason Weston. 2022. Blenderbot 3: a deployed conversational agent that continually learns to responsibly engage. *CoRR*, abs/2208.03188.

Pei-Hao Su, Nikola Mrkšić, Iñigo Casanueva, and Ivan Vulić. 2018. Deep learning for conversational AI. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorial Abstracts*, pages 27–32.

Yixuan Su, Lei Shu, Elman Mansimov, Arshit Gupta, Deng Cai, Yi-An Lai, and Yi Zhang. 2022. Multi-task pre-training for plug-and-play task-oriented dialogue system. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, ACL 2022*, pages 4661–4676.

Hao Sun, Guangxuan Xu, Jiawen Deng, Jiale Cheng, Chujie Zheng, Hao Zhou, Nanyun Peng, Xiaoyan Zhu, and Minlie Huang. 2022. On the safety of conversational models: Taxonomy, dataset, and benchmark. In *Findings of the Association for Computational Linguistics: ACL 2022*, pages 3906–3923.

Kai Sun, Seungwhan Moon, Paul A. Crook, Stephen Roller, Becka Silvert, Bing Liu, Zhiguang Wang, Honglei Liu, Eunjoon Cho, and Claire Cardie. 2021. Adding chit-chat to enhance task-oriented dialogues. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2021*, pages 1570–1583.

Jianheng Tang, Tiancheng Zhao, Chenyan Xiong, Xiaodan Liang, Eric P. Xing, and Zhiting Hu. 2019. Target-guided open-domain conversation. In *ACL 2019*, pages 5624–5634.

Romal Thoppilan, Daniel De Freitas, Jamie Hall, Noam Shazeer, Apoorv Kulshreshtha, Heng-Tze Cheng, Alicia Jin, Taylor Bos, Leslie Baker, Yu Du, YaGuang Li, Hongrae Lee, Huaixiu Steven Zheng, Amin Ghafouri, Marcelo Menegali, Yanping Huang, Maxim Krikun, Dmitry Lepikhin, James Qin, Dehao Chen, Yuanzhong Xu, Zhifeng Chen, Adam Roberts, Maarten Bosma, Yanqi Zhou, Chung-Ching Chang, Igor Krivokon, Will Rusch, Marc Pickett, Kathleen S. Meier-Hellstern, Meredith Ringel Morris, Tulsee Doshi, Renelito Delos Santos, Toju Duke, Johnny Soraker, Ben Zevenbergen, Vinodkumar Prabhakaran, Mark Diaz, Ben Hutchinson, Kristen Olson, Alejandra Molina, Erin Hoffman-John, Josh Lee, Lora Aroyo, Ravi Rajakumar, Alena Butryna, Matthew

Lamm, Viktoriya Kuzmina, Joe Fenton, Aaron Cohen, Rachel Bernstein, Ray Kurzweil, Blaise Aguera-Arcas, Claire Cui, Marian Croak, Ed H. Chi, and Quoc Le. 2022. Lamda: Language models for dialog applications. *CoRR*, abs/2201.08239.

Megan Ung, Jing Xu, and Y-Lan Boureau. 2022. Safer-dialogues: Taking feedback gracefully after conversational safety failures. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, ACL 2022*, pages 6462–6481.

Lingzhi Wang, Shafiq R. Joty, Wei Gao, Xingshan Zeng, and Kam-Fai Wong. 2022. Improving conversational recommender system via contextual and time-aware modeling with less domain-specific knowledge. *CoRR*, abs/2209.11386.

Xuewei Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019*, pages 5635–5649.

Qingyang Wu, Yichi Zhang, Yu Li, and Zhou Yu. 2021. Alternating recurrent dialog model with large-scale pre-trained language models. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021*, pages 1292–1301.

Wenquan Wu, Zhen Guo, Xiangyang Zhou, Hua Wu, Xiyuan Zhang, Rongzhong Lian, and Haifeng Wang. 2019. Proactive human-machine conversation with explicit conversation goal. In *ACL 2019*, pages 3794–3804.

Huiyuan Xie, Zhenghao Liu, Chenyan Xiong, Zhiyuan Liu, and Ann A. Copestake. 2021. TIAGE: A benchmark for topic-shift aware dialog modeling. In *Findings of ACL: EMNLP 2021*, pages 1684–1690.

Jun Xu, Haifeng Wang, Zhengyu Niu, Hua Wu, and Wanxiang Che. 2020. Knowledge graph grounded goal planning for open-domain conversation generation. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020*, pages 9338–9345.

Atsuki Yamaguchi, Kosui Iwasa, and Katsuhide Fujita. 2021. Dialogue act-based breakdown detection in negotiation dialogues. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, EACL 2021*, pages 745–757.

Rui Yan, Juntao Li, and Zhou Yu. 2022. Deep learning for dialogue systems: Chit-chat and beyond. *Foundations and Trends in Information Retrieval*, 15(5):417–589.

Runzhe Yang, Jingxiao Chen, and Karthik Narasimhan. 2021. Improving dialog systems for negotiation with personality modeling. 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, pages 681–693.

Wei Yang, Luchen Tan, Chunwei Lu, Anqi Cui, Han Li, Xi Chen, Kun Xiong, Muzi Wang, Ming Li, Jian Pei, and Jimmy Lin. 2019. Detecting customer complaint escalation with recurrent neural networks and manually-engineered features. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019*, pages 56–63.

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 *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022*, pages 745–755.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. 2022. React: Synergizing reasoning and acting in language models. *CoRR*, abs/2210.03629.

Tom Young, Frank Xing, Vlad Pandelea, Jinjie Ni, and Erik Cambria. 2022. Fusing task-oriented and open-domain dialogues in conversational agents. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022*, pages 11622–11629.

Hamed Zamani, Susan T. Dumais, Nick Craswell, Paul N. Bennett, and Gord Lueck. 2020. Generating clarifying questions for information retrieval. In *WWW 2020*, pages 418–428.

Hamed Zamani, Johanne R. Trippas, Jeff Dalton, and Filip Radlinski. 2022. Conversational information seeking. *CoRR*, abs/2201.08808.

Chen Zhang, Yiming Chen, Luis Fernando D'Haro, Yan Zhang, Thomas Friedrichs, Grandee Lee, and Haizhou Li. 2021a. Dynaeval: Unifying turn and dialogue level evaluation. 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*, pages 5676–5689.

Jun Zhang, Yan Yang, Chencai Chen, Liang He, and Zhou Yu. 2021b. KERS: A knowledge-enhanced framework for recommendation dialog systems with multiple subgoals. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1092–1101.

Saizheng Zhang, Emily Dinan, Jack Urbanek, Arthur Szlam, Douwe Kiela, and Jason Weston. 2018a. Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018*, pages 2204–2213.

Shuo Zhang and Krisztian Balog. 2020. Evaluating conversational recommender systems via user simulation. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1512–1520.

Yongfeng Zhang, Xu Chen, Qingyao Ai, Liu Yang, and W. Bruce Croft. 2018b. Towards conversational search and recommendation: System ask, user respond. In *CIKM 2018*, pages 177–186.

Xinyan Zhao, Bin He, Yasheng Wang, Yitong Li, Fei Mi, Yajiao Liu, Xin Jiang, Qun Liu, and Huanhuan Chen. 2022. Unids: A unified dialogue system for chit-chat and task-oriented dialogues. In *Proceedings of the Second DialDoc Workshop on Document-grounded Dialogue and Conversational Question Answering, DialDoc@ACL 2022*, pages 13–22.

Peixiang Zhong, Yong Liu, Hao Wang, and Chunyan Miao. 2021. Keyword-guided neural conversational model. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021*, pages 14568–14576.

Yiheng Zhou, Yulia Tsvetkov, Alan W. Black, and Zhou Yu. 2020. Augmenting non-collaborative dialog systems with explicit semantic and strategic dialog history. In *8th International Conference on Learning Representations, ICLR 2020*.

Caleb Ziems, Jane A. Yu, Yi-Chia Wang, Alon Y. Halevy, and Diyi Yang. 2022. The moral integrity corpus: A benchmark for ethical dialogue systems. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, ACL 2022*, pages 3755–3773.