Deep Learning for Dialogue Systems: Chit-Chat and Beyond

Other titles in Foundations and Trends® in Information Retrieval

Psychology-informed Recommender Systems
Elisabeth Lex, Dominik Kowald, Paul Seitlinger, Thi Ngoc Trang Tran, Alexander Felfernig and Markus Schedl
ISBN: 978-1-68083-844-2

Search and Discovery in Personal Email Collections
Michael Bendersky, Xuanhui Wang, Marc Najork and Donald Metzler
ISBN: 978-1-68083-838-1

Extracting, Mining and Predicting Users' Interests from Social Media
Fattane Zarrinkalam, Stefano Faralli, Guangyuan Piao and Ebrahim Bagheri
ISBN: 978-1-68083-738-4

Knowledge Graphs: An Information Retrieval Perspective
Ridho Reinanda, Edgar Meij and Maarten de Rijke
ISBN: 978-1-68083-728-5

Deep Learning for Matching in Search and Recommendation
Jun Xu, Xiangnan He and Hang Li
ISBN: 978-1-68083-706-3

Deep Learning for Dialogue Systems: Chit-Chat and Beyond

Rui Yan
Gaoling School of Artificial Intelligence
Renmin University of China
ruiyan@ruc.edu.cn

Juntao Li
School of Computer Science and Technology
Soochow University
ljt@suda.edu.cn

Zhou Yu
Computer Science Department
Columbia University
zy2461@columbia.edu

now
the essence of knowledge
Boston — Delft

Foundations and Trends® in Information Retrieval

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

R. Yan *et al.*. *Deep Learning for Dialogue Systems: Chit-Chat and Beyond*. Foundations and Trends® in Information Retrieval, vol. 15, no. 5, pp. 417–589, 2022.

ISBN: 978-1-63828-023-1

Full text available at: http://dx.doi.org/10.1561/1500000083

Editorial Scope

Topics

Foundations and Trends® in Information Retrieval publishes survey and tutorial articles in the following topics:

- Applications of IR
- Architectures for IR
- Collaborative filtering and recommender systems
- Cross-lingual and multilingual IR
- Distributed IR and federated search
- Evaluation issues and test collections for IR
- Formal models and language models for IR
- IR on mobile platforms
- Indexing and retrieval of structured documents
- Information categorization and clustering
- Information extraction
- Information filtering and routing
- Metasearch, rank aggregation and data fusion
- Natural language processing for IR
- Performance issues for IR systems, including algorithms, data structures, optimization techniques, and scalability
- Question answering
- Summarization of single documents, multiple documents, and corpora
- Text mining
- Topic detection and tracking
- Usability, interactivity, and visualization issues in IR
- User modelling and user studies for IR
- Web search

Information for Librarians

Foundations and Trends® in Information Retrieval, 2022, Volume 15, 5 issues. ISSN paper version 1554-0669. ISSN online version 1554-0677. Also available as a combined paper and online subscription.

Contents

Deep Learning for Dialogue Systems: Chit-Chat and Beyond

Rui Yan[1], Juntao Li[2] and Zhou Yu[3]

[1] *Gaoling School of Artificial Intelligence, Renmin University of China, China; ruiyan@ruc.edu.cn*
[2] *School of Computer Science and Technology, Soochow University, China; ljt@suda.edu.cn*
[3] *Computer Science Department, Columbia University, USA; zy2461@columbia.edu*

ABSTRACT

With the rapid progress of deep neural models and the explosion of available data resources, dialogue systems that supports extensive topics and chit-chat conversations are emerging as a research hot-spot for many communities, e.g., information retrieval (IR), natural language processing (NLP), and machine learning (ML). Building a chit-chat system with retrieval techniques is an essential task and has achieved great success in the past few years. The advance of chit-chat systems, in turn, can support extensive IR tasks, e.g., conversational search and conversational recommendation. To facilitate the development of both retrieval-based chit-chat systems and IR tasks supported by these systems, we survey chit-chat systems from two perspectives: (1) techniques to build chit-chat systems, i.e., deep retrieval-based models, generative methods, and their ensembles, and (2) chit-chat components in completing IR tasks. In each aspect,

---

Rui Yan, Juntao Li and Zhou Yu (2022), "Deep Learning for Dialogue Systems: Chit-Chat and Beyond", Foundations and Trends® in Information Retrieval: Vol. 15, No. 5, pp 417–589. DOI: 10.1561/1500000083.

we present cutting-edge neural methods and summarize the core challenges encountered and possible research directions.

1

Introduction

Starting from the 1960s, conversational artificial intelligence has become a crucial research field and has grabbed much more attention in recent years. Empowered by deep neural models, dialogue systems have demonstrated very impressive and appealing performance in virtual assistants and social bots. In viewing its potential and values, mainstream NLP, IR, and even ML communities have started contributing to dialogue systems. Dialogue systems can be roughly grouped into two classes, i.e., task-oriented and chit-chat systems. The former group focuses on completing predefined tasks with task-specific constraints and goals, e.g., restaurant booking and making calls. The later systems are mainly designed for modeling the 'chats' characteristic of human-human conversations (Daniel and James, 2020) without specific goals and constraints, i.e., the topics of the conversation could be any. Given predefined constraints and goals, task-oriented systems can achieve impressive performance with limited data and computational resources. In contrast, chit-chat systems require massive training conversations to mimic human chatting with extensive topics. Unlike task-oriented systems that have achieved great success for decades, learning-based chit-chat systems have not made great strides until recent years with the explosion of both data resources, model capacity (data modeling capability of deep neural networks), and computational power. To facilitate the development of chit-chat systems and their supported IR tasks and bridge the gap between different research communities, especially for the NLP and IR fields, we propose to systematically review state-of-the-art chit-chat systems and draw the connections between chit-chat and tasks, from being supporting tasks and the unified modeling framework in the paradigm of pre-trained language models.

Specifically, our work has a deep concentration on deep neural chit-chat systems using IR techniques and NLP methods, i.e., this monograph presents lessons and experiences of how to establish relevant, coherent, diverse, knowledgeable, and human-like chit-chat systems. Besides, we also discuss the connections between chit-chat systems and tasks, ranging from the perspectives of treating chit-chat components as supporting tasks to make task completion more natural (e.g., recommendation) to the trend of leveraging a unified framework for various downstream tasks in the era of pre-trained language models. To the best of our knowledge, it is the first survey to cover these topics and features.

The main contributions of this survey are as follows:

- We thoroughly survey the deep neural models in recent years for chit-chat systems, ranging from retrieval-based methods to generation-based approaches and the ensemble of these two types of models.

- We provide the connections between the recently resurgent chit-chat systems and task-oriented systems, e.g., conversational recommendation and conversational search, which enables us to explore more possibilities of building either better chit-chat systems or improving user experience in constructing IR systems.

- We introduce various solutions for addressing or mitigating the confronted challenges (e.g., context modeling, one-to-diversity, human factors learning) from different perspectives, including data-side and model-side solutions and utilization of extra resources.

- We present necessary data resources and evaluation methods for building retrieval-based and generation-based chit-chat systems.

- We also analyze the main challenges that we are facing and give the possible exploration directions and the rising trends, which will shed light on building human-like systems.

1.1 Intended Audience and Scope

This survey is intended to bridge the researchers of IR and the NLP community to move chit-chat systems forward and support more IR tasks. Our target audience includes, but is not limited to, IR or NLP researchers who want to study chit-chat from different perspectives, e.g., compensating retrieval-based models with the generation or vice versa, IR researchers who need to complete their tasks with the assistance of chit-chat systems, engineers with hands-on experience in building chit-chat systems to leverage advanced chit-chat modeling techniques, anyone who intends to quickly keep up with the frontier of chit-chat systems, anyone who wants to learn how to build chit-chat systems with deep neural architectures.

The main scope of this survey is based on the tutorial of SIGIR 2019 and WWW 2019 (Wu and Yan, 2019a, 2019b). We expand the tutorial contents with up-to-date techniques for building chit-chat systems, covering retrieval-based methods, generation components, and their ensembles. Besides the above contents, we also discuss the role of chit-chat systems in completing tasks, especially for some emerging IR tasks, e.g., conversational search and conversational recommendation. Considering the new trend of utilizing a unified self-supervised pre-training framework for both chit-chat and IR tasks, we further review a few recent works in this line and point out the possible future direction.

The rest of this survey is structured as follows:

- The remainder of this section summarizes the importance of chit-chat systems and presents the core problems of chit-chat systems. Besides, the landscape of chit-chat systems is also introduced. At the end of this section, we clarify the relationship and discrepancy between this survey and recent papers.

- Section 2 briefly reviews classic chit-chat systems before the neural age, including rule-based, template-based, and learning-based methods, and summarizes the characteristics of these methods.

- Section 3 sorts out and elaborates retrieval-based dialogue systems in recent years. This section starts with the pre-processing of conversation data and then discusses the core problems of retrieval-based chit-chat systems in detail (e.g., context modeling, knowledge utilization, human factors learning), which ends with necessary data resources and evaluation metrics for building retrieval-based chit-chat systems.

- Section 4 provides an alternative option for building chit-chat systems, i.e., generation-based methods, focusing on the pros and cons of generation-based methods in building chit-chat systems and their relationships with retrieval-based solutions. The last part of this section gives essential data resources, evaluation methods, and current challenges.

- In Section 5, we describe the ensemble of the aforementioned two types of frameworks, focusing on the scenarios of integration and re-ranking, template and prototype, and adversarial learning. Section 6 connects chit-chat systems with tasks, including vanilla tasks and newly appeared IR tasks like conversational search, and reveals the trend of unifying chit-chat dialogues and tasks with large-scale pre-trained language models.

- Section 7 first concludes this survey with the progress of chit-chat systems and the chit-chat component in IR tasks, and then points out the ongoing struggles and the possible future trends.

1.2 The Importance of Chit-Chat Systems

Chit-chat systems have become more and more popular and important in both academia and industry. Studying chit-chat systems have various benefits, including providing helpful services to human users, promoting the development of artificial intelligence technologies, holding tremendous potential and commercial values in the future.

To human users, chit-chat systems can satisfy a myriad of human needs, such as communication, social belongings, emotional engagement,

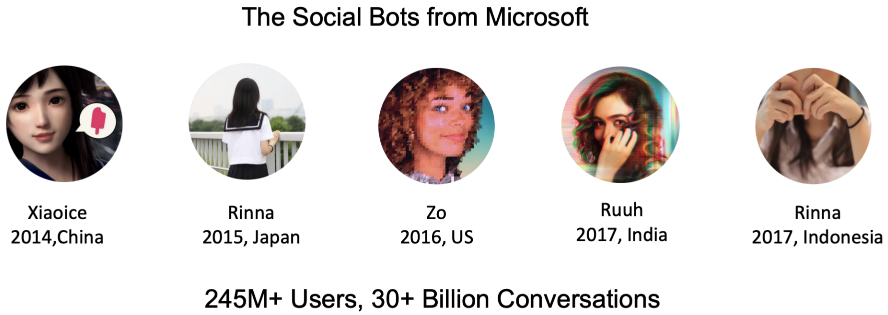

Figure 1.1: User size of social bots from Microsoft (Wu and Yan, 2019b).

etc. (Huang *et al.*, 2020b). On account of these merits, various applications, including but not limited to virtual assistants, smart speakers, social bots, and virtual customer services, are developed. As shown in Figure 1.1, chit-chat systems from the Microsoft corporation alone attracted over 245 million users and achieved over 30 billion conversations by 2019.

As for the connections between chit-chat systems and technology development, it is an indicator to calibrate the progress of artificial intelligence by launching the Turing test which is designed to test whether a machine can exhibit intelligent behaviours equivalent or indistinguishable from a human[1]. Building chit-chat systems also poses various unique challenges to state-of-the-art deep neural models, e.g., one-to-diversity, long-range context modeling, topic shift, long-term engagement computation, human factors learning, and the settlement of these problems, in turn, facilitates the progress of deep learning methods and encourages technical development.

Except for contributing to technology development and human needs, chit-chat systems also connect to various online commercial services. As demonstrated in Figure 1.2, chit-chat conversation might mix with goal-oriented demands, such as question answering, image search, and recommendation. Exploring chit-chat conversations could seamlessly find the demands of users and complete different tasks in a more efficient manner accordingly, i.e., without introducing multiple

[1] https://en.wikipedia.org/wiki/Turing_test (date accessed: 11 April 2022)

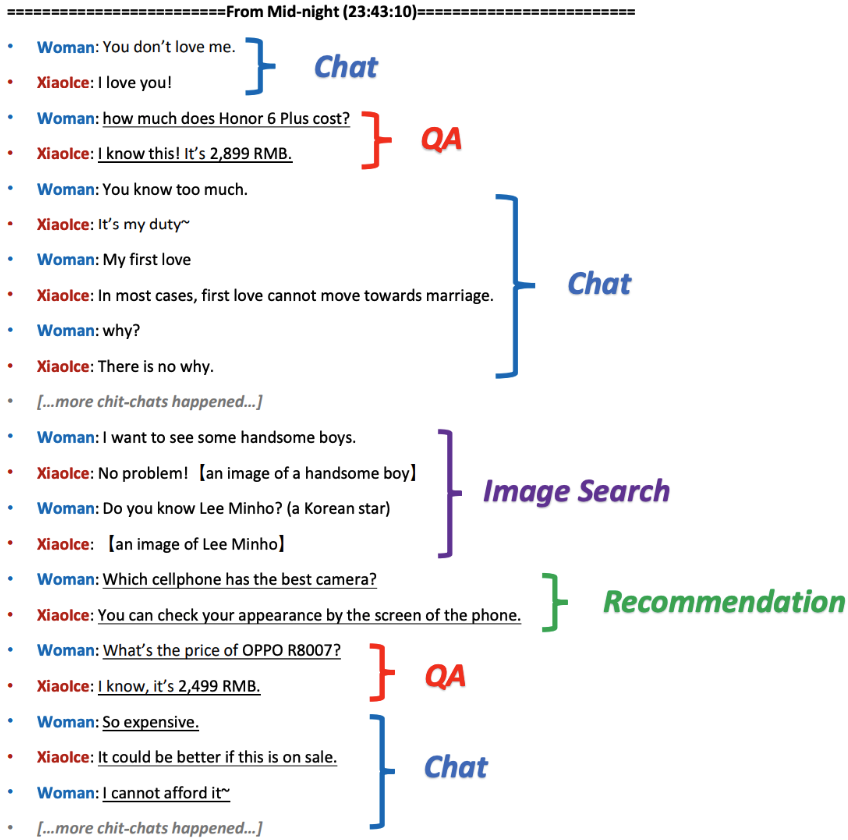

Figure 1.2: A case study that demonstrates the connections between chatbots and goal-oriented applications (Wu and Yan, 2019b).

task-specific systems. It also serves an essential role in intelligent entities and devices by providing the human-machine interface. The progress of chatbots could assist the development of robotics.

With the rapid progress of conversational AI techniques that support human-like interactions between computers and humans, it can be imaged that chit-chat systems are likely to have more industrial applications and broad market prospects. We believe that the potential of these systems is far more than we have seen in recent few years on social bots, virtual assistants, information seeking systems. In the far future, conversational AI systems might change almost everything in our daily life, e.g., the games will be more immersive, robotics are more intelligent that is able to sing, talk and even make friends with humans.

1.3 The Core Problems of Chit-Chat Systems

One of the main goals of chit-chat systems is to pass the Turing test so as to prove that an artificial program can chat like humans. Thus, the properties of human conversations should be considered and modeled in chit-chat systems. Seeing that human conversations are intricate and difficult to formulate, we utilize the qualitative analysis results of human conversation properties in Daniel and James (2020) to divide and shape the core problems of chit-chat systems. There are mainly six basic properties for human conversations: (1) **turns**, (2) **speech acts**, (3) **grounding**, (4) **sub-dialogues and dialogue structure**, (5) **initiative**, and (6) **inference and implicature**. For deep neural models trained on massive conversation data, **speech acts** and **dialogue structure** are implicitly modelled by neural networks. As for **initiative**, **user-initiative** and **system-initiative** frameworks are more common for task-specific systems while **mixed initiative** are very difficult to achieve. Thus, researchers mainly focus on the following problems in deep neural chit-chat systems.

Context Modeling. One of the main challenges we encountered is the long-range context modeling. Unlike task-oriented conversations that mainly consist of task-specific contents and usually complete a user demand in no more than a few dozens of conversation turns, chit-chat conversation is tied up with over hundreds of turns in usual, owing to the non-goal-oriented nature of chit-chat. In view of this, long-range context modeling has become a crucial issue for chit-chat dialogues to make conversations more consistent and coherent.

One-to-Diversity. In addition to multi-turn context modeling, one-to-diversity has also hindered the development of chit-chat systems. Unlike task-oriented conversations that take task completion as the evaluation metric, chit-chat further needs to mimic human-like conversations. Among various characteristics of human conversations, modeling expression diversity and one-to-many correlations bear the brunt.

Knowledge and Grounding. Beyond learning statistical patterns from existing human conversations, advanced chit-chat systems are expected to master and leverage knowledge like human beings. Besides commonsense knowledge, chit-chat conversations often correlate with non-contextual information, i.e., information and content that are not in context. Hereafter, we denote these extra information as grounding.

Human Factors. For chit-chat systems, user experience and engagement are always the core. To build a better chit-chat system, we have to consider the influence of various human factors, such as personalized expression preference, emotional changes, and beyond.

1.4 Landscape of Chit-Chat Systems and Beyond

A View from Chit-Chat. Advanced chit-chat systems mainly utilize cutting-edge deep neural techniques to automatically obtain responses for any newly given query or dialogue contexts. We group existing chit-chat models into three categories, i.e., frameworks based on retrieval techniques, generation-based models, and the ensemble of these two kinds of solutions. Retrieval-based frameworks mainly study how to automatically select feasible response candidates, covering the multi-turn context matching, extra resource utilization, human factors constraining, and pre-trained context-aware representation usages. Generation-based research focuses on the limitations of sequence to sequence networks, exploring from the perspective of data manipulation, generation pipelines, training objectives, large-scale pre-trained language models, and aforementioned context modeling, as well as human factors. Ensemble solutions investigate how to compensate retrieval-based dialogue systems with the merits of generation models and vice versa.

Linking Chit-Chat with Tasks. The connections between chit-chat and tasks can be categorized into three different directions. One is to discover and complete specific goals from chit-chat human-machine conversations to achieve better user engagement. The second is to enhance downstream tasks with chit-chat components, e.g., it can make it easier for users to accept recommended items from commercial recommendation systems.

Another possible direction is to utilize a unified large-scale pre-training framework to complete chit-chat conversations and tasks.

1.5 Comparisons with Existing Surveys

Recently, several tutorials and survey papers on dialogue systems have been presented (Yih *et al.*, 2015; Yih *et al.*, 2016), focusing on deep learning techniques and various IR related tasks (Gao, 2017), e.g., question answering (QA). Li and Yan (2018) briefly reviewed the multi-turn conversation methods involved in the NLPCC 2018 shared task, including both retrieval models and generation solutions. Chen *et al.* (2018b) provided a tutorial on spoken dialogue systems, which is mainly about traditional task-oriented dialogue systems. Serban *et al.* (2018) offered a thorough investigation on the public data available for building dialogue systems. Gao *et al.* (2019a) covered a myriad of topics in dialogue systems, including question answering, reading comprehension, task-oriented systems, social bots and industrial applications. Huang *et al.* (2020b) comprehensively studied three challenges that researchers are facing at present in building intelligent dialogue systems. Yan and Wu (2021) briefly summarized the progress and future of chit-chat dialogues with limited coverage and insufficient in-depth study. Zamani *et al.* (2022) recently provided an overview of existing research related to conversational information seeking. Gao *et al.* (2022) also wrote a book about conversational information seeking but focused on recent advances and technical details for building the main modules of conversational information retrieval systems. Considering that deep neural-based systems are the mainstream and are still in the process of development, we mainly compare this paper with recent surveys in closely related fields. More concretely, we conduct comparisons with two recent papers presented by Gao *et al.* (2019a) and Huang *et al.* (2020b), respectively.

This survey differs with Gao *et al.* (2019a) from the following aspects:

- We mainly focus on chit-chat systems rather than focus on task-oriented systems, question answering, and machine reading comprehension.

- We group recent research from the view of chit-chat, a specific type of conversation system that has attracted millions of users, instead of connecting goal-oriented dialogues and fully data-driven social bots from a unified perspective of optimal decision making.
- We mainly survey end-to-end methods built upon deep learning methods, instead of presenting task-oriented pipeline models or connecting traditional machine learning methods with modern neural models.
- We expose the recently explosive progress of completing tasks with the assistance of chit-chat systems, e.g., conversational recommendation (Lei *et al.*, 2020). We review the new paradigm of building chit-chat systems and completing tasks in recent large-scale pre-trained language models.

Compared with the short survey written by Huang *et al.* (2020b), we further present the following contents.

- Instead of focusing on surveying research that relates to specific challenges of chit-chat systems, we present a comprehensive study of modern chit-chat systems based on deep neural models.
- Except for discussing the main challenges that we are facing, this survey presents various solutions for addressing a myriad of challenges in the chit-chat conversations, which can provide guidance for anyone who wants to build chit-chat systems.
- This survey also has a border coverage, which draws the connection between chit-chat and goal-oriented systems, and emerging tasks of the IR community.

References

Adiwardana, D., M.-T. Luong, D. R. So, J. Hall, N. Fiedel, R. Thoppilan, Z. Yang, A. Kulshreshtha, G. Nemade, Y. Lu, *et al.* (2020). "Towards a human-like open-domain chatbot". *arXiv preprint arXiv:2001.09977*.

Akama, R., K. Inada, N. Inoue, S. Kobayashi, and K. Inui. (2017). "Generating stylistically consistent dialog responses with transfer learning". In: *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 408–412.

Alamri, H., V. Cartillier, A. Das, J. Wang, A. Cherian, I. Essa, D. Batra, T. K. Marks, C. Hori, P. Anderson, *et al.* (2019). "Audio visual scene-aware dialog". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7558–7567.

Antol, S., A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. L. Zitnick, and D. Parikh. (2015). "Vqa: Visual question answering". In: *Proceedings of the IEEE international conference on computer vision*. 2425–2433.

Bahdanau, D., K. Cho, and Y. Bengio. (2015). "Neural Machine Translation by Jointly Learning to Align and Translate". In: *3rd International Conference on Learning Representations, ICLR 2015*.

Baheti, A., A. Ritter, J. Li, and W. B. Dolan. (2018). "Generating More Interesting Responses in Neural Conversation Models with Distributional Constraints". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 3970–3980.

Baheti, A., A. Ritter, and K. Small. (2020). "Fluent Response Generation for Conversational Question Answering". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 191–207.

Bak, J. and A. Oh. (2019). "Variational hierarchical user-based conversation model". 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)*. 1941–1950.

Bak, J. and A. Oh. (2020). "Speaker Sensitive Response Evaluation Model". *arXiv preprint arXiv:2006.07015*.

Banerjee, S. and A. Lavie. (2005). "METEOR: An automatic metric for MT evaluation with improved correlation with human judgments". In: *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*. 65–72.

Bao, S., H. He, F. Wang, H. Wu, and H. Wang. (2020a). "PLATO: Pre-trained Dialogue Generation Model with Discrete Latent Variable". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 85–96.

Bao, S., H. He, F. Wang, H. Wu, H. Wang, W. Wu, Z. Guo, Z. Liu, and X. Xu. (2020b). "Plato-2: Towards building an open-domain chatbot via curriculum learning". *arXiv preprint arXiv:2006.16779*.

Bao, S., H. He, F. Wang, H. Wu, H. Wang, W. Wu, Z. Wu, Z. Guo, H. Lu, X. Huang, *et al.* (2021). "Plato-xl: Exploring the large-scale pre-training of dialogue generation". *arXiv preprint arXiv:2109.09519*.

Bradeško, L. and D. Mladenić. (2012). "A survey of chatbot systems through a loebner prize competition". In: *Proceedings of Slovenian language technologies society eighth conference of language technologies*. Institut Jožef Stefan Ljubljana, Slovenia. 34–37.

Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, *et al.* (2020). "Language models are few-shot learners". *arXiv preprint arXiv:2005.14165*.

Cai, D., Y. Wang, W. Bi, Z. Tu, X. Liu, W. Lam, and S. Shi. (2019a). "Skeleton-to-Response: Dialogue Generation Guided by Retrieval Memory". In: *NAACL*.

Cai, D., Y. Wang, W. Bi, Z. Tu, X. Liu, and S. Shi. (2019b). "Retrieval-guided Dialogue Response Generation via a Matching-to-Generation Framework". In: *EMNLP*.

Cai, H., H. Chen, Y. Song, Z. Ding, Y. Bao, W. Yan, and X. Zhao. (2020a). "Group-wise Contrastive Learning for Neural Dialogue Generation". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 793–802.

Cai, H., H. Chen, Y. Song, C. Zhang, X. Zhao, and D. Yin. (2020b). "Data Manipulation: Towards Effective Instance Learning for Neural Dialogue Generation via Learning to Augment and Reweight". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 6334–6343.

Cai, Y., M. Zuo, Q. Zhang, H. Xiong, and K. Li. (2020c). "A Bichannel Transformer with Context Encoding for Document-Driven Conversation Generation in Social Media". *Complexity*. 2020.

Campos, J. A., K. Cho, A. Otegi, A. Soroa, E. Agirre, and G. Azkune. (2020). "Improving Conversational Question Answering Systems after Deployment using Feedback-Weighted Learning". In: *Proceedings of the 28th International Conference on Computational Linguistics*. 2561–2571.

Cao, K. and S. Clark. (2017). "Latent Variable Dialogue Models and their Diversity". In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*. 182–187.

Cao, Y., W. Bi, M. Fang, and D. Tao. (2020). "Pretrained Language Models for Dialogue Generation with Multiple Input Sources". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 909–917.

Chan, Z., J. Li, X. Yang, X. Chen, W. Hu, D. Zhao, and R. Yan. (2019). "Modeling personalization in continuous space for response generation via augmented wasserstein autoencoders". 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)*. 1931–1940.

Chan, Z., L. Liu, J. Li, H. Zhang, D. Zhao, S. Shi, and R. Yan. (2021). "Enhancing the open-domain dialogue evaluation in latent space". In: *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*. 4889–4900.

Chang, J., R. He, L. Wang, X. Zhao, T. Yang, and R. Wang. (2019). "A Semi-Supervised Stable Variational Network for Promoting Replier-Consistency in Dialogue Generation". 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)*. 1920–1930.

Charniak, E. *et al.* (2016). "Parsing as language modeling". In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 2331–2336.

Chen, H., X. Liu, D. Yin, and J. Tang. (2017). "A survey on dialogue systems: Recent advances and new frontiers". *Acm Sigkdd Explorations Newsletter*. 19(2): 25–35.

Chen, H., Z. Ren, J. Tang, Y. E. Zhao, and D. Yin. (2018a). "Hierarchical variational memory network for dialogue generation". In: *Proceedings of the 2018 World Wide Web Conference*. 1653–1662.

Chen, Q., J. Lin, Y. Zhang, M. Ding, Y. Cen, H. Yang, and J. Tang. (2019). "Towards knowledge-based recommender dialog system". *arXiv preprint arXiv:1908.05391*.

Chen, X., F. Meng, P. Li, F. Chen, S. Xu, B. Xu, and J. Zhou. (2020a). "Bridging the Gap between Prior and Posterior Knowledge Selection for Knowledge-Grounded Dialogue Generation". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 3426–3437.

Chen, Y., L. Wu, and M. J. Zaki. (2020b). "GraphFlow: Exploiting Conversation Flow with Graph Neural Networks for Conversational Machine Comprehension".

Chen, Y.-N., A. Celikyilmaz, and D. Hakkani-Tur. (2018b). "Deep learning for dialogue systems". In: *Proceedings of the 27th International Conference on Computational Linguistics: Tutorial Abstracts*. 25–31.

Chu, E., P. Vijayaraghavan, and D. Roy. (2018). "Learning Personas from Dialogue with Attentive Memory Networks". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2638–2646.

Clark, P., N. Balasubramanian, S. Bhakthavatsalam, K. Humphreys, J. Kinkead, A. Sabharwal, and O. Tafjord. (2014). "Automatic construction of inference-supporting knowledge bases". In: *4th Workshop on Automated Knowledge Base Construction (AKBC)*. Citeseer.

Colby, K. M., S. Weber, and F. D. Hilf. (1971). "Artificial paranoia". *Artificial Intelligence*. 2(1): 1–25.

Colombo, P., W. Witon, A. Modi, J. Kennedy, and M. Kapadia. (2019). "Affect-Driven Dialog Generation". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 3734–3743.

Copple, K. L. (2008). *Bringing AI to life: putting today's tools and resources to work*. na.

Csáky, R., P. Purgai, and G. Recski. (2019). "Improving Neural Conversational Models with Entropy-Based Data Filtering". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 5650–5669.

Cui, Z., Y. Li, J. Zhang, J. Cui, C. Wei, and B. Wang. (2020). "Focus-Constrained Attention Mechanism for CVAE-based Response Generation". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 2021–2030.

Daniel, J. and H. M. James. (2020). "Speech and Language Processing". *Online Draft*.

Das, A., S. Kottur, K. Gupta, A. Singh, D. Yadav, J. M. Moura, D. Parikh, and D. Batra. (2017a). "Visual dialog". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 326–335.

Das, A., S. Kottur, J. M. Moura, S. Lee, and D. Batra. (2017b). "Learning cooperative visual dialog agents with deep reinforcement learning". *arXiv preprint arXiv:1703.06585*.

Devlin, J., M.-W. Chang, K. Lee, and K. Toutanova. (2018). “Bert: Pre-training of deep bidirectional transformers for language understanding”. *arXiv preprint arXiv:1810.04805*.

Devlin, J., M.-W. Chang, K. Lee, and K. 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, Volume 1 (Long and Short Papers)*. 4171–4186.

Dinan, E., S. Roller, K. Shuster, A. Fan, M. Auli, and J. Weston. (2018). “Wizard of wikipedia: Knowledge-powered conversational agents”. *arXiv preprint arXiv:1811.01241*.

Du, J., W. Li, Y. He, R. Xu, L. Bing, and X. Wang. (2018). “Variational autoregressive decoder for neural response generation”. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 3154–3163.

Du, W. and A. W. Black. (2019). “Boosting dialog response generation”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 38–43.

Dziri, N., E. Kamalloo, K. Mathewson, and O. R. Zaiane. (2019). “Augmenting Neural Response Generation with Context-Aware Topical Attention”. In: *Proceedings of the First Workshop on NLP for Conversational AI*. 18–31.

Engonopoulos, N., C. Teichmann, and A. Koller. (2018). “Discovering User Groups for Natural Language Generation”. In: *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. 171–179.

Feng, S., X. Ren, H. Chen, B. Sun, K. Li, and X. Sun. (2020). “Regularizing Dialogue Generation by Imitating Implicit Scenarios”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 6592–6604.

Forgues, G., J. Pineau, J.-M. Larchevêque, and R. Tremblay. (2014). “Bootstrapping dialog systems with word embeddings”. In: *Nips, workshop*. Vol. 2.

Galley, M., C. Brockett, A. Sordoni, Y. Ji, M. Auli, C. Quirk, M. Mitchell, J. Gao, and B. Dolan. (2015). "deltableu: A discriminative metric for generation tasks with intrinsically diverse targets". *arXiv preprint arXiv:1506.06863*.

Gan, Z., Y. Cheng, A. Kholy, L. Li, J. Liu, and J. Gao. (2019). "Multi-step Reasoning via Recurrent Dual Attention for Visual Dialog". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 6463–6474.

Gao, J. (2017). "An introduction to deep learning for natural language processing". *International Summer School on Deep Learning, Bilbao*.

Gao, J., M. Galley, L. Li, *et al.* (2019a). "Neural Approaches to Conversational AI". *Foundations and Trends® in Information Retrieval*. 13(2-3): 127–298.

Gao, J., C. Xiong, P. Bennett, and N. Craswell. (2022). "Neural Approaches to Conversational Information Retrieval". *arXiv preprint arXiv:2201.05176*.

Gao, J., W. Bi, X. Liu, J. Li, and S. Shi. (2019b). "Generating multiple diverse responses for short-text conversation". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. No. 01. 6383–6390.

Gao, J., W. Bi, X. Liu, J. Li, G. Zhou, and S. Shi. (2019c). "A Discrete CVAE for Response Generation on Short-Text Conversation". 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)*. 1898–1908.

Gao, X., S. Lee, Y. Zhang, C. Brockett, M. Galley, J. Gao, and W. B. Dolan. (2019d). "Jointly Optimizing Diversity and Relevance in Neural Response Generation". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 1229–1238.

Gao, X., Y. Zhang, S. Lee, M. Galley, C. Brockett, J. Gao, and B. Dolan. (2019e). "Structuring Latent Spaces for Stylized Response Generation". 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)*. 1814–1823.

Gao, Y., C.-S. Wu, S. Joty, C. Xiong, R. Socher, I. King, M. Lyu, and S. C. Hoi. (2020). "Explicit memory tracker with coarse-to-fine reasoning for conversational machine reading". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 935–945.

Gehring, J., M. Auli, D. Grangier, D. Yarats, and Y. N. Dauphin. (2017). "Convolutional Sequence to Sequence Learning". In: *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org. 1243–1252.

Ghazarian, S., J. T.-Z. Wei, A. Galstyan, and N. Peng. (2019). "Better automatic evaluation of open-domain dialogue systems with contextualized embeddings". *arXiv preprint arXiv:1904.10635*.

Ghazvininejad, M., C. Brockett, M.-W. Chang, B. Dolan, J. Gao, W.-t. Yih, and M. Galley. (2018). "A knowledge-grounded neural conversation model". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. No. 1.

Goodfellow, I. J., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio. (2014). "Generative Adversarial Nets". In: *NIPS*.

Gopalakrishnan, K., B. Hedayatnia, Q. Chen, A. Gottardi, S. Kwatra, A. Venkatesh, R. Gabriel, D. Hakkani-Tür, and A. A. AI. (2019). "Topical-Chat: Towards Knowledge-Grounded Open-Domain Conversations." In: *INTERSPEECH*. 1891–1895.

Gu, J.-C., T. Li, Q. Liu, Z.-H. Ling, Z. Su, S. Wei, and X. Zhu. (2020a). "Speaker-aware bert for multi-turn response selection in retrieval-based chatbots". In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2041–2044.

Gu, J.-C., Z.-H. Ling, X. Zhu, and Q. Liu. (2019a). "Dually Interactive Matching Network for Personalized Response Selection in Retrieval-Based Chatbots". 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)*. 1845–1854.

Gu, J.-C., Z. Ling, Q. Liu, Z. Chen, and X. Zhu. (2020b). "Filtering before Iteratively Referring for Knowledge-Grounded Response Selection in Retrieval-Based Chatbots". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 1412–1422.

Gu, X., K. Cho, J.-W. Ha, and S. Kim. (2019b). "DialogWAE: Multimodal Response Generation with Conditional Wasserstein Auto-Encoder". In: *International Conference on Learning Representations*. URL: https://openreview.net/forum?id=BkgBvsC9FQ.

Guo, D., C. Xu, and D. Tao. (2019a). "Image-question-answer synergistic network for visual dialog". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10434–10443.

Guo, D., H. Wang, and M. Wang. (2019b). "Dual Visual Attention Network for Visual Dialog." In: *IJCAI*. 4989–4995.

Guo, D., H. Wang, H. Zhang, Z.-J. Zha, and M. Wang. (2020). "Iterative context-aware graph inference for visual dialog". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10055–10064.

Guo, X., H. Wu, Y. Cheng, S. Rennie, G. Tesauro, and R. S. Feris. (2018). "Dialog-based interactive image retrieval". In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. 676–686.

Gupta, P., S. Mehri, T. Zhao, A. Pavel, M. Eskenazi, and J. P. Bigham. (2019). "Investigating evaluation of open-domain dialogue systems with human generated multiple references". *arXiv preprint arXiv: 1907.10568*.

Gupta, S., B. P. S. Rawat, and H. Yu. (2020). "Conversational Machine Comprehension: a Literature Review". In: *Proceedings of the 28th International Conference on Computational Linguistics*. 2739–2753.

Guu, K., T. Hashimoto, Y. Oren, and P. Liang. (2018). "Generating Sentences by Editing Prototypes". *Transactions of the Association for Computational Linguistics*. 6: 437–450.

Ham, D., J.-G. Lee, Y. Jang, and K.-E. Kim. (2020). "End-to-end neural pipeline for goal-oriented dialogue systems using GPT-2". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 583–592.

Han, X., Z. Zhang, N. Ding, Y. Gu, X. Liu, Y. Huo, J. Qiu, L. Zhang, W. Han, M. Huang, *et al.* (2021). "Pre-trained models: Past, present and future". *AI Open*.

He, K., X. Zhang, S. Ren, and J. Sun. (2016). "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

He, T. and J. Glass. (2020). "Negative Training for Neural Dialogue Response Generation". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2044–2058.

He, W., Y. Dai, Y. Zheng, Y. Wu, Z. Cao, D. Liu, P. Jiang, M. Yang, F. Huang, L. Si, *et al.* (2021). "GALAXY: A Generative Pre-trained Model for Task-Oriented Dialog with Semi-Supervised Learning and Explicit Policy Injection". *arXiv preprint arXiv:2111.14592*.

Hendricks, L. A., S. Venugopalan, M. Rohrbach, R. Mooney, K. Saenko, and T. Darrell. (2016). "Deep compositional captioning: Describing novel object categories without paired training data". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1–10.

Hewitt, J. and C. D. Manning. (2019). "A structural probe for finding syntax in word representations". In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 4129–4138.

Hixon, B., P. Clark, and H. Hajishirzi. (2015). "Learning knowledge graphs for question answering through conversational dialog". In: *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 851–861.

Hosseini-Asl, E., B. McCann, C.-S. Wu, S. Yavuz, and R. Socher. (2020). "A simple language model for task-oriented dialogue". *arXiv preprint arXiv:2005.00796*.

Hu, B., Z. Lu, H. Li, and Q. Chen. (2014). "Convolutional neural network architectures for matching natural language sentences". In: *NIPS*.

Hu, W., Z. Chan, B. Liu, D. Zhao, J. Ma, and R. Yan. (2019). "GSN: A Graph-Structured Network for Multi-Party Dialogues".

Hu, Z., J. E. F. Tree, and M. Walker. (2018). "Modeling linguistic and personality adaptation for natural language generation". In: *Proceedings of the 19th annual SIGdial meeting on discourse and dialogue*. 20–31.

Hua, K., Z. Feng, C. Tao, R. Yan, and L. Zhang. (2020). "Learning to Detect Relevant Contexts and Knowledge for Response Selection in Retrieval-based Dialogue Systems". In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 525–534.

Huang, C., O. R. Zaiane, A. Trabelsi, and N. Dziri. (2018). "Automatic dialogue generation with expressed emotions". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*. 49–54.

Huang, L., Z. Ye, J. Qin, L. Lin, and X. Liang. (2020a). "GRADE: Automatic Graph-Enhanced Coherence Metric for Evaluating Open-Domain Dialogue Systems". *arXiv preprint arXiv:2010.03994*.

Huang, M., X. Zhu, and J. Gao. (2020b). "Challenges in building intelligent open-domain dialog systems". *ACM Transactions on Information Systems (TOIS)*. 38(3): 1–32.

Hutchens, J. L. (1996). "How to pass the Turing test by cheating". *School of Electrical, Electronic and Computer Engineering research report TR97-05. Perth: University of Western Australia*.

Jawahar, G., B. Sagot, and D. Seddah. (2019). "What Does BERT Learn about the Structure of Language?" In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 3651–3657.

Ji, Z., Z. Lu, and H. Li. (2014). "An information retrieval approach to short text conversation". *arXiv preprint arXiv:1408.6988*.

Jiang, B., W. Zhou, J. Yang, C. Yang, S. Wang, and L. Pang. (2020). "PEDNet: A Persona Enhanced Dual Alternating Learning Network for Conversational Response Generation". In: *Proceedings of the 28th International Conference on Computational Linguistics*. 4089–4099.

Jiang, S., P. Ren, C. Monz, and M. de Rijke. (2019). "Improving neural response diversity with frequency-aware cross-entropy loss". In: *The World Wide Web Conference*. 2879–2885.

Jiang, S. and M. de Rijke. (2018). "Why are Sequence-to-Sequence Models So Dull? Understanding the Low-Diversity Problem of Chatbots". In: *Proceedings of the 2018 EMNLP Workshop SCAI: The 2nd International Workshop on Search-Oriented Conversational AI*. 81–86.

Jung, J., B. Son, and S. Lyu. (2020). "AttnIO: Knowledge Graph Exploration with In-and-Out Attention Flow for Knowledge-Grounded Dialogue". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 3484–3497.

Kaiser, M., R. Saha Roy, and G. Weikum. (2020). "Conversational Question Answering over Passages by Leveraging Word Proximity Networks". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2129–2132.

Kang, G.-C., J. Lim, and B.-T. Zhang. (2019a). "Dual Attention Networks for Visual Reference Resolution in Visual Dialog". 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)*. 2024–2033.

Kang, D., A. Balakrishnan, P. Shah, P. Crook, Y.-L. Boureau, and J. Weston. (2019b). "Recommendation as a communication game: Self-supervised bot-play for goal-oriented dialogue". *arXiv preprint arXiv:1909.03922*.

Khan, K., G. Sahu, V. Balasubramanian, L. Mou, and O. Vechtomova. (2020). "Adversarial Learning on the Latent Space for Diverse Dialog Generation". In: *Proceedings of the 28th International Conference on Computational Linguistics*. 5026–5034.

Khayrallah, H. and J. Sedoc. (2020). “SMRT Chatbots: Improving Non-Task-Oriented Dialog with Simulated Multiple Reference Training”. *arXiv preprint arXiv:2011.00547*.

Kim, H., H. Tan, and M. Bansal. (2020). “Modality-balanced models for visual dialogue”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 05. 8091–8098.

Ko, W.-J., G. Durrett, and J. J. Li. (2019). “Linguistically-informed specificity and semantic plausibility for dialogue generation”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 3456–3466.

Ko, W.-J., A. Ray, Y. Shen, and H. Jin. (2020). “Generating Dialogue Responses from a Semantic Latent Space”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 4339–4349.

Kong, J., Z. Zhong, Y. Cai, X. Wu, and D. Ren. (2020). “TSDG: Content-aware Neural Response Generation with Two-stage Decoding Process”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 2121–2126.

Kongyoung, S., C. Macdonald, and I. Ounis. (2020). “Multi-Task Learning using Dynamic Task Weighting for Conversational Question Answering”. In: *Proceedings of the 5th International Workshop on Search-Oriented Conversational AI (SCAI)*. 17–26.

Kottur, S., J. M. Moura, D. Parikh, D. Batra, and M. Rohrbach. (2018). “Visual coreference resolution in visual dialog using neural module networks”. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 153–169.

Kottur, S., X. Wang, and V. Carvalho. (2017). “Exploring Personalized Neural Conversational Models.” In: *IJCAI*. 3728–3734.

Kundu, S., Q. Lin, and H. T. Ng. (2020). “Learning to Identify Follow-Up Questions in Conversational Question Answering”. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 959–968.

Laban, P., J. Canny, and M. A. Hearst. (2020). "What's The Latest? A Question-driven News Chatbot". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. 380–387.

Lan, T., X.-L. Mao, W. Wei, X. Gao, and H. Huang. (2020). "Pone: A novel automatic evaluation metric for open-domain generative dialogue systems". *ACM Transactions on Information Systems (TOIS)*. 39(1): 1–37.

Lan, Z., M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut. (2019). "ALBERT: A Lite BERT for Self-supervised Learning of Language Representations". In: *International Conference on Learning Representations*.

Lei, W., X. He, M. de Rijke, and T.-S. Chua. (2020). "Conversational Recommendation: Formulation, Methods, and Evaluation". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2425–2428.

Lewis, M., Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. 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*. 7871–7880.

Lewis, P., P. Stenetorp, and S. Riedel. (2021). "Question and Answer Test-Train Overlap in Open-Domain Question Answering Datasets". In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. 1000–1008.

Li, J. and X. Sun. (2018). "A Syntactically Constrained Bidirectional-Asynchronous Approach for Emotional Conversation Generation". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 678–683.

Li, J., M. Galley, C. Brockett, J. Gao, and W. B. Dolan. (2016a). "A Diversity-Promoting Objective Function for Neural Conversation Models". In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 110–119.

Li, J., M. Galley, C. Brockett, G. Spithourakis, J. Gao, and W. B. Dolan. (2016b). "A Persona-Based Neural Conversation Model". In: *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 994–1003.

Li, J., W. Monroe, A. Ritter, D. Jurafsky, M. Galley, and J. Gao. (2016c). "Deep Reinforcement Learning for Dialogue Generation". In: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 1192–1202.

Li, J., W. Monroe, T. Shi, S. Jean, A. Ritter, and D. Jurafsky. (2017a). "Adversarial Learning for Neural Dialogue Generation". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2157–2169.

Li, J., Z. Zhang, H. Zhao, X. Zhou, and X. Zhou. (2020a). "Task-specific Objectives of Pre-trained Language Models for Dialogue Adaptation". *arXiv preprint arXiv:2009.04984*.

Li, J., C. Liu, C. Tao, Z. Chan, D. Zhao, M. Zhang, and R. Yan. (2021a). "Dialogue history matters! Personalized response selection in multi-turn retrieval-based chatbots". *ACM Transactions on Information Systems (TOIS)*. 39(4): 1–25.

Li, J., L. Qiu, B. Tang, D. Chen, D. Zhao, and R. Yan. (2019). "Insufficient data can also rock! learning to converse using smaller data with augmentation". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. No. 01. 6698–6705.

Li, J. and R. Yan. (2018). "Overview of the NLPCC 2018 shared task: Multi-turn human-computer conversations". In: *CCF International Conference on Natural Language Processing and Chinese Computing*. Springer. 446–451.

Li, L., C. Xu, W. Wu, Y. Zhao, X. Zhao, and C. Tao. (2020b). "Zero-Resource Knowledge-Grounded Dialogue Generation". *arXiv preprint arXiv:2008.12918*.

Li, M., S. Roller, I. Kulikov, S. Welleck, Y.-L. Boureau, K. Cho, and J. Weston. (2020c). "Don't Say That! Making Inconsistent Dialogue Unlikely with Unlikelihood Training". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 4715–4728.

Li, Q., H. Chen, Z. Ren, P. Ren, Z. Tu, and Z. Chen. (2020d). "EmpDG: Multi-resolution Interactive Empathetic Dialogue Generation". In: *Proceedings of the 28th International Conference on Computational Linguistics*. 4454–4466.

Li, R., S. Kahou, H. Schulz, V. Michalski, L. Charlin, and C. Pal. (2018). "Towards deep conversational recommendations". *arXiv preprint arXiv:1812.07617*.

Li, S., S. Feng, D. Wang, K. Song, Y. Zhang, and W. Wang. (2020e). "EmoElicitor: An Open Domain Response Generation Model with User Emotional Reaction Awareness". In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI*. 3637–3643.

Li, Y., H. Su, X. Shen, W. Li, Z. Cao, and S. Niu. (2017b). "Dailydialog: A manually labelled multi-turn dialogue dataset". *arXiv preprint arXiv:1710.03957*.

Li, Y., S. A. Hayati, W. Shi, and Z. Yu. (2021b). "DEUX: An Attribute-Guided Framework for Sociable Recommendation Dialog Systems". *arXiv preprint arXiv:2105.00825*.

Lian, R., M. Xie, F. Wang, J. Peng, and H. Wu. (2019). "Learning to Select Knowledge for Response Generation in Dialog Systems".

Liao, L., Y. Ma, X. He, R. Hong, and T.-s. Chua. (2018). "Knowledge-aware multimodal dialogue systems". In: *Proceedings of the 26th ACM international conference on Multimedia*. 801–809.

Liao, L., R. Takanobu, Y. Ma, X. Yang, M. Huang, and T.-S. Chua. (2019). "Deep conversational recommender in travel". *arXiv preprint arXiv:1907.00710*.

Lin, C.-Y. (2004). "Rouge: A package for automatic evaluation of summaries". In: *Text summarization branches out*. 74–81.

Lin, X., W. Jian, J. He, T. Wang, and W. Chu. (2020a). "Generating Informative Conversational Response using Recurrent Knowledge-Interaction and Knowledge-Copy". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 41–52.

Lin, Z., A. Madotto, G. I. Winata, and P. Fung. (2020b). "MinTL: Minimalist Transfer Learning for Task-Oriented Dialogue Systems". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 3391–3405.

Lintean, M. C. and V. Rus. (2012). "Measuring Semantic Similarity in Short Texts through Greedy Pairing and Word Semantics." In: *Flairs conference*. 244–249.

Litman, D., S. Singh, M. S. Kearns, and M. Walker. (2000). "NJFun-a reinforcement learning spoken dialogue system". In: *ANLP-NAACL 2000 Workshop: Conversational Systems*.

Liu, C., K. Liu, S. He, Z. Nie, and J. Zhao. (2019a). "Incorporating Interlocutor-Aware Context into Response Generation on Multi-Party Chatbots". In: *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*. 718–727.

Liu, P., W. Yuan, J. Fu, Z. Jiang, H. Hayashi, and G. Neubig. (2021). "Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing". *arXiv preprint arXiv:2107.13586*.

Liu, Q., Y. Chen, B. Chen, L. Jian-Guang, Z. Chen, B. Zhou, and D. Zhang. (2020). "You Impress Me: Dialogue Generation via Mutual Persona Perception". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 1417–1427.

Liu, S., H. Chen, Z. Ren, Y. Feng, Q. Liu, and D. Yin. (2018a). "Knowledge diffusion for neural dialogue generation". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1489–1498.

Liu, Y., W. Bi, J. Gao, X. Liu, J. Yao, and S. Shi. (2018b). "Towards less generic responses in neural conversation models: A statistical re-weighting method". In: *Proceedings of the 2018 conference on empirical methods in natural language processing*. 2769–2774.

Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. (2019b). "Roberta: A robustly optimized bert pretraining approach". *arXiv preprint arXiv:1907.11692*.

Liu, Z., Z.-Y. Niu, H. Wu, H. Wang, *et al.* (2019c). "Knowledge Aware Conversation Generation with Explainable Reasoning over Augmented Graphs". 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)*. 1782–1792.

Lowe, R., M. Noseworthy, I. V. Serban, N. Angelard-Gontier, Y. Bengio, and J. Pineau. (2017). "Towards an automatic turing test: Learning to evaluate dialogue responses". *arXiv preprint arXiv:1708.07149*.

Lowe, R., N. Pow, I. Serban, and J. Pineau. (2015). "The Ubuntu Dialogue Corpus: A Large Dataset for Research in Unstructured Multi-Turn Dialogue Systems". In: *SIGDIAL*. 285–294.

Lowe, R., I. V. Serban, M. Noseworthy, L. Charlin, and J. Pineau. (2016). "On the evaluation of dialogue systems with next utterance classification". *arXiv preprint arXiv:1605.05414*.

Lu, J., A. Kannan, J. Yang, D. Parikh, and D. Batra. (2017). "Best of both worlds: transferring knowledge from discriminative learning to a generative visual dialog model". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 313–323.

Lu, J., X. Ren, Y. Ren, A. Liu, and Z. Xu. (2020). "Improving contextual language models for response retrieval in multi-turn conversation". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1805–1808.

Lu, Z. and H. Li. (2013). "A deep architecture for matching short texts". In: *NIPS*.

Luan, Y., C. Brockett, W. B. Dolan, J. Gao, and M. Galley. (2017). "Multi-Task Learning for Speaker-Role Adaptation in Neural Conversation Models". In: *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. 605–614.

Lubis, N., S. Sakti, K. Yoshino, and S. Nakamura. (2019). "Positive emotion elicitation in chat-based dialogue systems". *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 27(4): 866–877.

Ma, Z., R. Yang, B. Du, and Y. Chen. (2020). “A Control Unit for Emotional Conversation Generation”. *IEEE Access*. 8: 43168–43176.

Madotto, A., Z. Lin, G. I. Winata, and P. Fung. (2021). “Few-Shot Bot: Prompt-Based Learning for Dialogue Systems”. *arXiv preprint arXiv:2110.08118*.

Madotto, A., Z. Lin, C.-S. Wu, and P. Fung. (2019). “Personalizing dialogue agents via meta-learning”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 5454–5459.

Majumder, B. P., H. Jhamtani, T. Berg-Kirkpatrick, and J. McAuley. (2020a). “Like Hiking? You Probably Enjoy Nature: Persona-grounded Dialog with Commonsense Expansions”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 9194–9206.

Majumder, N., P. Hong, S. Peng, J. Lu, D. Ghosal, A. Gelbukh, R. Mihalcea, and S. Poria. (2020b). “MIME: MIMicking Emotions for Empathetic Response Generation”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 8968–8979.

Mangrulkar, S., S. Shrivastava, V. Thenkanidiyoor, and D. A. Dinesh. (2018). “A context-aware convolutional natural language generation model for dialogue systems”. In: *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. 191–200.

Mazare, P.-E., S. Humeau, M. Raison, and A. Bordes. (2018). “Training Millions of Personalized Dialogue Agents”. In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2775–2779.

McCann, B., N. S. Keskar, C. Xiong, and R. Socher. (2018). “The natural language decathlon: Multitask learning as question answering”. *arXiv preprint arXiv:1806.08730*.

Mehri, S. and M. Eskenazi. (2020a). “Unsupervised evaluation of interactive dialog with dialogpt”. *arXiv preprint arXiv:2006.12719*.

Mehri, S. and M. Eskenazi. (2020b). “Usr: An unsupervised and reference free evaluation metric for dialog generation”. *arXiv preprint arXiv:2005.00456*.

Mi, F., W. Zhou, L. Kong, F. Cai, M. Huang, and B. Faltings. (2021). "Self-training Improves Pre-training for Few-shot Learning in Task-oriented Dialog Systems". In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 1887–1898.

Mikolov, T., K. Chen, G. Corrado, and J. Dean. (2013). "Efficient estimation of word representations in vector space". *arXiv preprint arXiv:1301.3781*.

Misu, T., K. Georgila, A. Leuski, and D. Traum. (2012). "Reinforcement learning of question-answering dialogue policies for virtual museum guides". In: *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. 84–93.

Mitchell, J. and M. Lapata. (2008). "Vector-based models of semantic composition". *NAACL-HLT*: 236–244.

Moghe, N., S. Arora, S. Banerjee, and M. M. Khapra. (2018). "Towards Exploiting Background Knowledge for Building Conversation Systems". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2322–2332.

Moon, S., P. Shah, A. Kumar, and R. Subba. (2019). "Opendialkg: Explainable conversational reasoning with attention-based walks over knowledge graphs". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 845–854.

Mrkšić, N., D. Ó. Séaghdha, T.-H. Wen, B. Thomson, and S. Young. (2017). "Neural Belief Tracker: Data-Driven Dialogue State Tracking". In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1777–1788.

Murahari, V., D. Batra, D. Parikh, and A. Das. (2020). "Large-scale pretraining for visual dialog: A simple state-of-the-art baseline". In: *European Conference on Computer Vision*. Springer. 336–352.

Nayak, N., D. Hakkani-Tür, M. A. Walker, and L. P. Heck. (2017). "To Plan or not to Plan? Discourse Planning in Slot-Value Informed Sequence to Sequence Models for Language Generation." In: *INTERSPEECH*. 3339–3343.

Niu, Y., H. Zhang, M. Zhang, J. Zhang, Z. Lu, and J.-R. Wen. (2019). "Recursive visual attention in visual dialog". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6679–6688.

Noh, H., P. H. Seo, and B. Han. (2016). "Image question answering using convolutional neural network with dynamic parameter prediction". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 30–38.

Ohsugi, Y., I. Saito, K. Nishida, H. Asano, and J. Tomita. (2019). "A Simple but Effective Method to Incorporate Multi-turn Context with BERT for Conversational Machine Comprehension". In: *Proceedings of the First Workshop on NLP for Conversational AI*. 11–17.

Olabiyi, O., A. Khazane, A. Salimov, and E. Mueller. (2019). "An Adversarial Learning Framework For A Persona-Based Multi-Turn Dialogue Model". In: *Proceedings of the Workshop on Methods for Optimizing and Evaluating Neural Language Generation*. 1–10.

Pan, B., H. Li, Z. Yao, D. Cai, and H. Sun. (2019). "Reinforced Dynamic Reasoning for Conversational Question Generation". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2114–2124.

Papineni, K., S. Roukos, T. Ward, and W.-J. Zhu. (2002). "BLEU: a method for automatic evaluation of machine translation". In: *ACL*. ACL. 311–318.

Park, S., T. Whang, Y. Yoon, and H. Lim. (2020). "Multi-View Attention Networks for Visual Dialog". *arXiv preprint arXiv:2004.14025*.

Park, Y., J. Cho, and G. Kim. (2018). "A Hierarchical Latent Structure for Variational Conversation Modeling". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. 1792–1801.

Parthasarathi, P. and J. Pineau. (2018). "Extending Neural Generative Conversational Model using External Knowledge Sources". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 690–695.

Peng, B., C. Li, J. Li, S. Shayandeh, L. Liden, and J. Gao. (2020). "Soloist: Few-shot task-oriented dialog with a single pretrained autoregressive model". *arXiv preprint arXiv:2005.05298*.

Penha, G. and C. Hauff. (2020). “What does BERT know about books, movies and music? Probing BERT for Conversational Recommendation”. In: *Fourteenth ACM Conference on Recommender Systems*. 388–397.

Pennington, J., R. Socher, and C. D. Manning. (2014). “Glove: Global vectors for word representation”. In: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

Peters, M. E., M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer. (2018). “Deep contextualized word representations”. *arXiv preprint arXiv:1802.05365*.

Petroni, F., T. Rocktäschel, S. Riedel, P. Lewis, A. Bakhtin, Y. Wu, and A. Miller. (2019). “Language Models as Knowledge Bases?” 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)*. 2463–2473.

Qi, P., Y. Zhang, and C. D. Manning. (2020). “Stay Hungry, Stay Focused: Generating Informative and Specific Questions in Information-Seeking Conversations”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 25–40.

Qi, W., Y. Gong, Y. Yan, C. Xu, B. Yao, B. Zhou, B. Cheng, D. Jiang, J. Chen, R. Zhang, *et al.* (2021). “ProphetNet-X: Large-Scale Pretraining Models for English, Chinese, Multi-lingual, Dialog, and Code Generation”. *arXiv preprint arXiv:2104.08006*.

Qian, Q., M. Huang, H. Zhao, J. Xu, and X. Zhu. (2018). “Assigning Personality/Profile to a Chatting Machine for Coherent Conversation Generation.” In: *IJCAI*. 4279–4285.

Qin, L., M. Galley, C. Brockett, X. Liu, X. Gao, W. B. Dolan, Y. Choi, and J. Gao. (2019). “Conversing by Reading: Contentful Neural Conversation with On-demand Machine Reading”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 5427–5436.

Qiu, L., J. Li, W. Bi, D. Zhao, and R. Yan. (2019). “Are training samples correlated? learning to generate dialogue responses with multiple references”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 3826–3835.

Qiu, L., Y. Shiu, P. Lin, R. Song, Y. Liu, D. Zhao, and R. Yan. (2020). "What If Bots Feel Moods?" In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1161–1170.

Qu, C., L. Yang, C. Chen, M. Qiu, W. B. Croft, and M. Iyyer. (2020). "Open-retrieval conversational question answering". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 539–548.

Radford, A., K. Narasimhan, T. Salimans, and I. Sutskever. (2018a). "Improving Language Understanding by Generative Pre-Training". *Open AI, blog*.

Radford, A., K. Narasimhan, T. Salimans, and I. Sutskever. (2018b). "Improving language understanding by generative pre-training".

Radford, A., J. Wu, R. Child, D. Luan, D. Amodei, I. Sutskever, *et al.* (2019). "Language models are unsupervised multitask learners". *OpenAI blog*. 1(8): 9.

Raffel, C., N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu. (2020). "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer". *Journal of Machine Learning Research*. 21(140): 1–67.

Rashkin, H., E. M. Smith, M. Li, and Y.-L. Boureau. (2018). "Towards empathetic open-domain conversation models: A new benchmark and dataset". *arXiv preprint arXiv:1811.00207*.

Rashkin, H., E. M. Smith, M. Li, and Y.-L. Boureau. (2019). "Towards Empathetic Open-domain Conversation Models: A New Benchmark and Dataset". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 5370–5381.

Reddy, S., D. Chen, and C. D. Manning. (2019). "Coqa: A conversational question answering challenge". *Transactions of the Association for Computational Linguistics*. 7: 249–266.

Ren, S., K. He, R. B. Girshick, and J. Sun. (2015). "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: *NIPS*.

Ritter, A., C. Cherry, and W. B. Dolan. (2011). "Data-driven response generation in social media". In: *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*. 583–593.

Roberts, A., C. Raffel, and N. Shazeer. (2020). "How Much Knowledge Can You Pack into the Parameters of a Language Model?" In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 5418–5426.

Roller, S., E. Dinan, N. Goyal, D. Ju, M. Williamson, Y. Liu, J. Xu, M. Ott, E. M. Smith, Y.-L. Boureau, *et al.* (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*. 300–325.

Saeidi, M., M. Bartolo, P. Lewis, S. Singh, T. Rocktäschel, M. Sheldon, G. Bouchard, and S. Riedel. (2018). "Interpretation of Natural Language Rules in Conversational Machine Reading". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2087–2097.

Sai, A. B., A. K. Mohankumar, S. Arora, and M. M. Khapra. (2020). "Improving Dialog Evaluation with a Multi-reference Adversarial Dataset and Large Scale Pretraining". *Transactions of the Association for Computational Linguistics*. 8: 810–827.

Sanh, V., A. Webson, C. Raffel, S. H. Bach, L. Sutawika, Z. Alyafeai, A. Chaffin, A. Stiegler, T. L. Scao, A. Raja, *et al.* (2021). "Multitask prompted training enables zero-shot task generalization". *arXiv preprint arXiv:2110.08207*.

Sankar, C., S. Subramanian, C. Pal, S. Chandar, and Y. Bengio. (2019). "Do Neural Dialog Systems Use the Conversation History Effectively? An Empirical Study". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 32–37.

Sarkar, R., K. Goswami, M. Arcan, and J. P. McCrae. (2020). "Suggest me a movie for tonight: Leveraging Knowledge Graphs for Conversational Recommendation". In: *Proceedings of the 28th International Conference on Computational Linguistics*. 4179–4189.

Schatzmann, J., K. Weilhammer, M. Stuttle, and S. Young. (2006). "A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies". *The knowledge engineering review*. 21(2): 97–126.

Schick, T. and H. Schütze. (2021a). “Exploiting Cloze-Questions for Few-Shot Text Classification and Natural Language Inference”. In: *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*. 255–269.

Schick, T. and H. Schütze. (2021b). “It’s Not Just Size That Matters: Small Language Models Are Also Few-Shot Learners”. In: *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 2339–2352.

Schwartz, I., S. Yu, T. Hazan, and A. G. Schwing. (2019). “Factor graph attention”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2039–2048.

Scialom, T., S. S. Tekiroğlu, J. Staiano, and M. Guerini. (2020). “Toward Stance-based Personas for Opinionated Dialogues”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 2625–2635.

See, A., S. Roller, D. Kiela, and J. Weston. (2019). “What makes a good conversation? How controllable attributes affect human judgments”. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 1702–1723.

Sellam, T., D. Das, and A. P. Parikh. (2020). “BLEURT: Learning robust metrics for text generation”. *ACL*.

Seo, P. H., A. M. Lehrmann, B. Han, and L. Sigal. (2017). “Visual Reference Resolution using Attention Memory for Visual Dialog”. In: *NIPS*.

Serban, I. V., R. Lowe, P. Henderson, L. Charlin, and J. Pineau. (2018). “A Survey of Available Corpora For Building Data-Driven Dialogue Systems: The Journal Version”. *Dialogue & Discourse*. 9(1): 1–49.

Shah, M., S. Mehri, and T. Srinivasan. (2020). “Reasoning Over History: Context Aware Visual Dialog”. In: *Proceedings of the First International Workshop on Natural Language Processing Beyond Text*. 75–83.

Shah, P., D. Hakkani-Tur, B. Liu, and G. Tur. (2018). "Bootstrapping a neural conversational agent with dialogue self-play, crowdsourcing and on-line reinforcement learning". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 3 (Industry Papers)*. 41–51.

Shang, L., Z. Lu, and H. Li. (2015). "Neural responding machine for short-text conversation". *arXiv preprint arXiv:1503.02364*.

Shawar, B. A. and E. S. Atwell. (2005). "Using corpora in machine-learning chatbot systems". *International journal of corpus linguistics*. 10(4): 489–516.

Shen, L. and Y. Feng. (2020). "CDL: Curriculum Dual Learning for Emotion-Controllable Response Generation". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 556–566.

Shen, L., Y. Feng, and H. Zhan. (2019a). "Modeling Semantic Relationship in Multi-turn Conversations with Hierarchical Latent Variables". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 5497–5502.

Shen, T., X. Geng, Q. Tao, D. Guo, D. Tang, N. Duan, G. Long, and D. Jiang. (2019b). "Multi-Task Learning for Conversational Question Answering over a Large-Scale Knowledge Base". 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)*. 2442–2451.

Shen, X., H. Su, W. Li, and D. Klakow. (2018). "Nexus network: Connecting the preceding and the following in dialogue generation". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 4316–4327.

Shen, X., H. Su, Y. Li, W. Li, S. Niu, Y. Zhao, A. Aizawa, and G. Long. (2017). "A Conditional Variational Framework for Dialog Generation". In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 504–509.

Shi, W. and Z. Yu. (2018). "Sentiment Adaptive End-to-End Dialog Systems". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1509–1519.

Shum, M., S. Zheng, W. Kryscinski, C. Xiong, and R. Socher. (2020). "Sketch-Fill-AR: A Persona-Grounded Chit-Chat Generation Framework". In: *Proceedings of the 2nd Workshop on Natural Language Processing for Conversational AI*. 118–131.

Simonyan, K. and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition". *arXiv preprint arXiv:1409.1556*.

Song, H., Y. Wang, W. Zhang, X. Liu, and T. Liu. (2020a). "Generate, Delete and Rewrite: A Three-Stage Framework for Improving Persona Consistency of Dialogue Generation". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 5821–5831.

Song, H., W.-N. Zhang, Y. Cui, D. Wang, and T. Liu. (2019a). "Exploiting Persona Information for Diverse Generation of Conversational Responses".

Song, H., W.-N. Zhang, J. Hu, and T. Liu. (2020b). "Generating persona consistent dialogues by exploiting natural language inference". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 05. 8878–8885.

Song, Y., C.-T. Li, J.-Y. Nie, M. Zhang, D. Zhao, and R. Yan. (2018). "An Ensemble of Retrieval-Based and Generation-Based Human-Computer Conversation Systems". In: *IJCAI*.

Song, Y., Z. Tian, D. Zhao, M. Zhang, and R. Yan. (2017). "Diversifying neural conversation model with maximal marginal relevance". In: *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*. 169–174.

Song, Z., X. Zheng, L. Liu, M. Xu, and X.-J. Huang. (2019b). "Generating responses with a specific emotion in dialog". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 3685–3695.

Sordoni, A., M. Galley, M. Auli, C. Brockett, Y. Ji, M. Mitchell, J.-Y. Nie, J. Gao, and W. B. Dolan. (2015). "A Neural Network Approach to Context-Sensitive Generation of Conversational Responses". In: *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 196–205.

Su, H., X. Shen, S. Zhao, Z. Xiao, P. Hu, C. Niu, J. Zhou, *et al.* (2020a). "Diversifying Dialogue Generation with Non-Conversational Text". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 7087–7097.

Su, Y., L. Shu, E. Mansimov, A. Gupta, D. Cai, Y.-A. Lai, and Y. Zhang. (2021). "Multi-Task Pre-Training for Plug-and-Play Task-Oriented Dialogue System". *arXiv preprint arXiv:2109.14739*.

Su, Y., Y. Wang, S. Baker, D. Cai, X. Liu, A. Korhonen, and N. Collier. (2020b). "Prototype-to-Style: Dialogue Generation with Style-Aware Editing on Retrieval Memory". *ArXiv*. abs/2004.02214.

Suganuma, M., T. Okatani, *et al.* (2020). "Efficient Attention Mechanism for Visual Dialog that Can Handle All the Interactions Between Multiple Inputs". In: *16th European Conference on Computer Vision, ECCV 2020*. Springer Science and Business Media Deutschland GmbH. 223–240.

Sun, K., D. Yu, J. Chen, D. Yu, Y. Choi, and C. Cardie. (2019). "Dream: A challenge data set and models for dialogue-based reading comprehension". *Transactions of the Association for Computational Linguistics*. 7: 217–231.

Sun, T., X. Liu, X. Qiu, and X. Huang. (2021). "Paradigm Shift in Natural Language Processing". *arXiv preprint arXiv:2109.12575*.

Sutskever, I., O. Vinyals, and Q. V. Le. (2014). "Sequence to Sequence Learning with Neural Networks". In: *Advances in Neural Information Processing systems*. 3104–3112.

Takayama, J. and Y. Arase. (2020). "Consistent Response Generation with Controlled Specificity". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 4418–4427.

Tang, J., T. Zhao, C. Xiong, X. Liang, E. Xing, and Z. Hu. (2019). "Target-Guided Open-Domain Conversation". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 5624–5634.

Tao, C., J. Feng, C. Liu, J. Li, X. Geng, and D. Jiang. (2021a). "Building an Efficient and Effective Retrieval-based Dialogue System via Mutual Learning". *arXiv preprint arXiv:2110.00159*.

Tao, C., J. Feng, R. Yan, W. Wu, and D. Jiang. (2021b). "A survey on response selection for retrieval-based dialogues". In: *IJCAI*.

Tao, C., L. Mou, D. Zhao, and R. Yan. (2018). "Ruber: An unsupervised method for automatic evaluation of open-domain dialog systems". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. No. 1.

Tao, C., W. Wu, Y. Feng, D. Zhao, and R. Yan. (2020). "Improving Matching Models with Hierarchical Contextualized Representations for Multi-turn Response Selection". In: *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1865–1868.

Tao, C., W. Wu, C. Xu, W. Hu, D. Zhao, and R. Yan. (2019a). "Multi-representation fusion network for multi-turn response selection in retrieval-based chatbots". In: *WSDM*.

Tao, C., W. Wu, C. Xu, D. Zhao, and R. Yan. (2019b). "One time of interaction may not be enough: Go deep with an interaction-over-interaction network for response selection in dialogues". In: *ACL*.

Tay, Y., A. T. Luu, and S. C. Hui. (2018a). "Hermitian Co-Attention Networks for Text Matching in Asymmetrical Domains". In: *IJCAI*. 4425–4431.

Tay, Y., L. A. Tuan, and S. C. Hui. (2018b). "Multi-cast Attention Networks". In: *SIGKDD*.

Thorat, S. A. and V. Jadhav. (2020). "A review on implementation issues of rule-based chatbot systems". In: *Proceedings of the International Conference on Innovative Computing & Communications (ICICC)*.

Tian, Z., W. Bi, D. Lee, L. Xue, Y. SONG, X. Liu, and N. L. Zhang. (2020). "Response-Anticipated Memory for On-Demand Knowledge Integration in Response Generation". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 650–659.

Tian, Z., W. Bi, X. Li, and N. L. Zhang. (2019). "Learning to abstract for memory-augmented conversational response generation". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 3816–3825.

Tian, Z., R. Yan, L. Mou, Y. Song, Y. Feng, and D. Zhao. (2017). "How to make context more useful? an empirical study on context-aware neural conversational models". In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*. 231–236.

Tong, X., Z. Fu, M. Shang, D. Zhao, and R. Yan. (2018). "One" Ruler" for all languages: multi-lingual dialogue evaluation with adversarial multi-task learning". *arXiv preprint arXiv:1805.02914*.

Tuan, Y.-L., Y.-N. Chen, and H.-y. Lee. (2019). "DyKgChat: Benchmarking dialogue generation grounding on dynamic knowledge graphs". *arXiv preprint arXiv:1910.00610*.

Tuan, Y.-L., W. Wei, and W. Y. Wang. (2020). "Knowledge Injection into Dialogue Generation via Language Models". *arXiv preprint arXiv:2004.14614*.

Ueyama, A. and Y. Kano. (2020). "Diverse dialogue generation with context dependent dynamic loss function". In: *Proceedings of the 28th International Conference on Computational Linguistics*. 4123–4127.

Vakulenko, S., S. Longpre, Z. Tu, and R. Anantha. (2020). "A Wrong Answer or a Wrong Question? An Intricate Relationship between Question Reformulation and Answer Selection in Conversational Question Answering". In: *Proceedings of the 5th International Workshop on Search-Oriented Conversational AI (SCAI)*. 7–16.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. (2017). "Attention is all you need". In: *Advances in neural information processing systems*. 5998–6008.

Verma, N., A. Sharma, D. Madan, D. Contractor, H. Kumar, and S. Joshi. (2020). "Neural Conversational QA: Learning to Reason vs Exploiting Patterns". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 7263–7269.

Vinyals, O. and Q. Le. (2015). "A neural conversational model". *arXiv preprint arXiv:1506.05869*.

Vougiouklis, P., J. Hare, and E. Simperl. (2016). "A Neural Network Approach for Knowledge-Driven Response Generation". In: *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 3370–3380.

Wallace, R. (2003). "The elements of AIML style". *Alice AI Foundation*. 139.

Wang, C., P. Liu, and Y. Zhang. (2021). "Can Generative Pre-trained Language Models Serve As Knowledge Bases for Closed-book QA?" In: *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Online: Association for Computational Linguistics. 3241–3251. DOI: 10.18653/v1/2021.acl-long.251.

Wang, H., Z. Wu, and J. Chen. (2019a). "Multi-turn response selection in retrieval-based chatbots with iterated attentive convolution matching network". In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 1081–1090.

Wang, W., S. Feng, D. Wang, and Y. Zhang. (2019b). "Answer-guided and Semantic Coherent Question Generation in Open-domain Conversation". 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)*. 5069–5079.

Wang, X., W. Shi, R. Kim, Y. Oh, S. Yang, J. Zhang, and Z. Yu. (2019c). "Persuasion for good: Towards a personalized persuasive dialogue system for social good". *arXiv preprint arXiv:1906.06725*.

Wang, Y., C. Liu, M. Huang, and L. Nie. (2018). "Learning to Ask Questions in Open-domain Conversational Systems with Typed Decoders". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 2193–2203.

Wang, Y., P. Ke, Y. Zheng, K. Huang, Y. Jiang, X. Zhu, and M. Huang. (2020a). "A large-scale chinese short-text conversation dataset". In: *CCF International Conference on Natural Language Processing and Chinese Computing*. Springer. 91–103.

Wang, Y., S. Joty, M. Lyu, I. King, C. Xiong, and S. C. Hoi. (2020b). "VD-BERT: A Unified Vision and Dialog Transformer with BERT". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 3325–3338.

Wei, J., M. Bosma, V. Y. Zhao, K. Guu, A. W. Yu, B. Lester, N. Du, A. M. Dai, and Q. V. Le. (2021). "Finetuned language models are zero-shot learners". *arXiv preprint arXiv:2109.01652*.

Weintraub, J. (1986). "History of the PC Therapist". URL: http://www.loebner.net/Prizef/weintraub-bio.html.

Weizenbaum, J. (1966). "ELIZA—a computer program for the study of natural language communication between man and machine". *Communications of the ACM*. 9(1): 36–45.

Whang, T., D. Lee, C. Lee, K. Yang, D. Oh, and H. Lim. (2019). "An effective domain adaptive post-training method for bert in response selection". *arXiv preprint arXiv:1908.04812*.

Whang, T., D. Lee, D. Oh, C. Lee, K. Han, D.-h. Lee, and S. Lee. (2020). "Do Response Selection Models Really Know What's Next? Utterance Manipulation Strategies for Multi-turn Response Selection". *arXiv preprint arXiv:2009.04703*.

Williams, J. D., A. Raux, and M. Henderson. (2016). "The dialog state tracking challenge series: A review". *Dialogue & Discourse*. 7(3): 4–33.

Williams, J. D. (2007). "Partially observable Markov decision processes for spoken dialogue management". *PhD thesis*. University of Cambridge.

Wolf, T., V. Sanh, J. Chaumond, and C. Delangue. (2019). "Transfertransfo: A transfer learning approach for neural network based conversational agents". *arXiv preprint arXiv:1901.08149*.

Wu, B., M. Li, Z. Wang, Y. Chen, D. F. Wong, J. Huang, B. Wang, *et al.* (2020a). "Guiding Variational Response Generator to Exploit Persona". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 53–65.

Wu, C.-S., S. C. Hoi, R. Socher, and C. Xiong. (2020b). "TOD-BERT: Pre-trained Natural Language Understanding for Task-Oriented Dialogue". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 917–929.

Wu, S., Y. Li, D. Zhang, and Z. Wu. (2020c). "Improving Knowledge-Aware Dialogue Response Generation by Using Human-Written Prototype Dialogues". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 1402–1411.

Wu, S., Y. Li, D. Zhang, Y. Zhou, and Z. Wu. (2020d). "Diverse and informative dialogue generation with context-specific commonsense knowledge awareness". In: *Proceedings of the 58th annual meeting of the association for computational linguistics*. 5811–5820.

Wu, S., Y. Li, D. Zhang, Y. Zhou, and Z. Wu. (2020e). "Topicka: Generating commonsense knowledge-aware dialogue responses towards the recommended topic fact". In: *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI*. 3766–3772.

Wu, W. and R. Yan. (2019a). "Deep Chit-Chat: Deep Learning for Chatbots". In: *Companion Proceedings of The 2019 World Wide Web Conference*. 1329–1329.

Wu, W. and R. Yan. (2019b). "Deep chit-chat: Deep learning for chatbots". In: *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 1413–1414.

Wu, W., Z. Guo, X. Zhou, H. Wu, X. Zhang, R. Lian, and H. Wang. (2019a). "Proactive human-machine conversation with explicit conversation goals". *arXiv preprint arXiv:1906.05572*.

Wu, Y., F. Wei, S. Huang, Y. Wang, Z. Li, and M. Zhou. (2019b). "Response generation by context-aware prototype editing". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33. No. 01. 7281–7288.

Wu, Y., W. Wu, C. Xing, M. Zhou, and Z. Li. (2017). "Sequential Matching Network: A New Architecture for Multi-turn Response Selection in Retrieval-Based Chatbots". In: *ACL*. 496–505.

Wu, Y., W. Wu, C. Xu, and Z. Li. (2018). "Knowledge enhanced hybrid neural network for text matching". In: *AAAI*.

Xing, C., Y. Wu, W. Wu, Y. Huang, and M. Zhou. (2018). "Hierarchical recurrent attention network for response generation". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. No. 1.

Xu, H., S. Moon, H. Liu, B. Liu, P. Shah, and S. Y. Philip. (2020a). "User Memory Reasoning for Conversational Recommendation". In: *Proceedings of the 28th International Conference on Computational Linguistics*. 5288–5308.

Xu, J., H. Wang, Z. Niu, H. Wu, and W. Che. (2020b). "Knowledge graph grounded goal planning for open-domain conversation generation". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 05. 9338–9345.

Xu, K., J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhudinov, R. Zemel, and Y. Bengio. (2015). "Show, attend and tell: Neural image caption generation with visual attention". In: *International conference on machine learning*. PMLR. 2048–2057.

Xu, R., C. Tao, D. Jiang, X. Zhao, D. Zhao, and R. Yan. (2020c). "Learning an Effective Context-Response Matching Model with Self-Supervised Tasks for Retrieval-based Dialogues". *arXiv preprint arXiv:2009.06265*.

Xu, X., O. Dušek, I. Konstas, and V. Rieser. (2018a). "Better Conversations by Modeling, Filtering, and Optimizing for Coherence and Diversity". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 3981–3991.

Xu, Y., E. Ishii, S. Cahyawijaya, Z. Liu, G. I. Winata, A. Madotto, D. Su, and P. Fung. (2021). "Retrieval-free knowledge-grounded dialogue response generation with adapters". *arXiv preprint arXiv:2105.06232*.

Xu, Z., N. Jiang, B. Liu, W. Rong, B. Wu, B. Wang, Z. Wang, and X. Wang. (2018b). "LSDSCC: a large scale domain-specific conversational corpus for response generation with diversity oriented evaluation metrics". In: *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. 2070–2080.

Xu, Z., B. Liu, B. Wang, C.-J. Sun, X. Wang, Z. Wang, and C. Qi. (2017). "Neural response generation via gan with an approximate embedding layer". In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 617–626.

Yan, R. (2018). ""Chitty-Chitty-Chat Bot": Deep Learning for Conversational AI". In: *IJCAI*.

Yan, R., Y. Song, and H. Wu. (2016). "Learning to respond with deep neural networks for retrieval-based human-computer conversation system". In: *SIGIR*.

Yan, R. and W. Wu. (2021). "Empowering Conversational AI is a Trip to Mars: Progress and Future of Open Domain Human-Computer Dialogues". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. No. 17. 15078–15086.

Yang, B., C. Han, Y. Li, L. Zuo, and Z. Yu. (2021a). "Improving Conversational Recommendation Systems' Quality with Context-Aware Item Meta Information". *arXiv preprint arXiv:2112.08140*.

Yang, L., J. Hu, M. Qiu, C. Qu, J. Gao, W. Croft, X. Liu, Y. Shen, and J. Liu. (2019a). "A Hybrid Retrieval-Generation Neural Conversation Model". *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*.

Yang, P., X. Sun, W. Li, S. Ma, W. Wu, and H. Wang. (2018). "SGM: Sequence Generation Model for Multi-label Classification". In: *Proceedings of the 27th International Conference on Computational Linguistics*. 3915–3926.

Yang, T., Z.-J. Zha, and H. Zhang. (2019b). "Making history matter: History-advantage sequence training for visual dialog". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2561–2569.

Yang, Y., Y. Li, and X. Quan. (2021b). "UBAR: Towards Fully End-to-End Task-Oriented Dialog System with GPT-2". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 35. No. 16. 14230–14238.

Yang, Z., W. Wu, C. Xu, X. Liang, J. Bai, L. Wang, W. Wang, and Z. Li. (2020). "StyleDGPT: Stylized Response Generation with Pre-trained Language Models". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 1548–1559.

Yang, Z., Z. Dai, Y. Yang, J. Carbonell, R. Salakhutdinov, and Q. V. Le. (2019c). "Xlnet: Generalized autoregressive pretraining for language understanding". *arXiv preprint arXiv:1906.08237*.

Yang, Z., X. He, J. Gao, L. Deng, and A. Smola. (2016). "Stacked attention networks for image question answering". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 21–29.

Yenicelik, D., F. Schmidt, and Y. Kilcher. (2020). "How does BERT capture semantics? A closer look at polysemous words". In: *Proceedings of the Third BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*. 156–162.

Yih, W.-t., X. He, and J. Gao. (2015). "Deep learning and continuous representations for natural language processing". In: *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Tutorial Abstracts*. 6–8.

Yih, W.-t., X. He, and J. Gao. (2016). "Deep learning and continuous representations for natural language processing". In: *IJCAI*.

Young, S., M. Gašić, S. Keizer, F. Mairesse, J. Schatzmann, B. Thomson, and K. Yu. (2010). "The hidden information state model: A practical framework for POMDP-based spoken dialogue management". *Computer Speech & Language*. 24(2): 150–174.

Yu, H., A. Li, R. Jiang, Y. Jia, X. Zhao, and W. Han. (2019a). "HDGS: A Hybrid Dialogue Generation System using Adversarial Learning". *2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC)*: 135–141.

Yu, T., Y. Shen, R. Zhang, X. Zeng, and H. Jin. (2019b). "Vision-language recommendation via attribute augmented multimodal reinforcement learning". In: *Proceedings of the 27th ACM International Conference on Multimedia*. 39–47.

Yuan, C., W. Zhou, M. Li, S. Lv, F. Zhu, J. Han, and S. Hu. (2019). "Multi-hop Selector Network for Multi-turn Response Selection in Retrieval-based Chatbots". In: *EMNLP*.

Yuma, T., N. Yoshinaga, and M. Toyoda. (2020). "uBLEU: Uncertainty-Aware Automatic Evaluation Method for Open-Domain Dialogue Systems". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*. 199–206.

Zamani, H., J. R. Trippas, J. Dalton, and F. Radlinski. (2022). "Conversational Information Seeking". *arXiv preprint arXiv:2201.08808*.

Zeng, M., Y. Wang, and Y. Luo. (2019). "Dirichlet latent variable hierarchical recurrent encoder-decoder in dialogue generation". 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)*. 1267–1272.

Zeng, X., D. Zeng, S. He, K. Liu, and J. Zhao. (2018). "Extracting relational facts by an end-to-end neural model with copy mechanism". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 506–514.

Zhang, B. and X. Zhang. (2019). "Hierarchy response learning for neural conversation generation". 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)*. 1772–1781.

Zhang, H., Y. Lan, L. Pang, J. Guo, and X. Cheng. (2019a). "ReCoSa: Detecting the Relevant Contexts with Self-Attention for Multi-turn Dialogue Generation". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 3721–3730.

Zhang, H., Z. Liu, C. Xiong, and Z. Liu. (2020a). "Grounded Conversation Generation as Guided Traverses in Commonsense Knowledge Graphs". In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2031–2043.

Zhang, J., C. Tao, Z. Xu, Q. Xie, W. Chen, and R. Yan. (2019b). "EnsembleGAN: Adversarial Learning for Retrieval-Generation Ensemble Model on Short-Text Conversation". In: *SIGIR*.

Zhang, J., Y. Yang, C. Chen, L. He, and Z. Yu. (2021). "KERS: A Knowledge-Enhanced Framework for Recommendation Dialog Systems with Multiple Subgoals". In: *Findings of the Association for Computational Linguistics: EMNLP 2021*. 1092–1101.

Zhang, L., Y. Yang, J. Zhou, C. Chen, and L. He. (2020b). "Retrieval-Polished Response Generation for Chatbot". *IEEE Access*. 8: 123882–123890.

Zhang, M. and J. Li. (2021). "A commentary of GPT-3 in MIT Technology Review 2021". *Fundamental Research*. 1(6): 831–833.

Zhang, R., Y. Zheng, J. Shao, X. Mao, Y. Xi, and M. Huang. (2020c). "Dialogue Distillation: Open-domain Dialogue Augmentation Using Unpaired Data". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 3449–3460.

Zhang, R., T. Yu, Y. Shen, H. Jin, C. Chen, and L. Carin. (2020d). "Reward Constrained Interactive Recommendation with Natural Language Feedback". *arXiv preprint arXiv:2005.01618*.

Zhang, R., J. Guo, Y. Fan, Y. Lan, J. Xu, and X. Cheng. (2018a). "Learning to control the specificity in neural response generation". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1108–1117.

Zhang, S., E. Dinan, J. Urbanek, A. Szlam, D. Kiela, and J. Weston. (2018b). "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 (Volume 1: Long Papers)*. 2204–2213.

Zhang, T., V. Kishore, F. Wu, K. Q. Weinberger, and Y. Artzi. (2019c). "Bertscore: Evaluating text generation with bert". *ICLR*.

Zhang, W., Y. Cui, Y. Wang, Q. Zhu, L. Li, L. Zhou, and T. Liu. (2018c). "Context-sensitive generation of open-domain conversational responses". In: *Proceedings of the 27th International Conference on Computational Linguistics*. 2437–2447.

Zhang, W., K. Song, Y. Kang, Z. Wang, C. Sun, X. Liu, S. Li, M. Zhang, and L. Si. (2020e). “Multi-Turn Dialogue Generation in E-Commerce Platform with the Context of Historical Dialogue”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 1981–1990.

Zhang, X. (2019). “MC^2: Multi-perspective convolutional cube for conversational machine reading comprehension”. In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 6185–6190.

Zhang, Y., S. Sun, M. Galley, Y.-C. Chen, C. Brockett, X. Gao, J. Gao, J. Liu, and W. B. Dolan. (2020f). “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*. 270–278.

Zhang, Y., X. Chen, Q. Ai, L. Yang, and W. B. Croft. (2018d). “Towards conversational search and recommendation: System ask, user respond”. In: *Proceedings of the 27th acm international conference on information and knowledge management*. 177–186.

Zhang, Z., J. Li, P. Zhu, H. Zhao, and G. Liu. (2018e). “Modeling Multi-turn Conversation with Deep Utterance Aggregation”. In: *COLING*. 3740–3752.

Zhao, T., R. Zhao, and M. Eskenazi. (2017). “Learning Discourse-level Diversity for Neural Dialog Models using Conditional Variational Autoencoders”. In: *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 654–664.

Zhao, W., M. Peyrard, F. Liu, Y. Gao, C. M. Meyer, and S. Eger. (2019a). “Moverscore: Text generation evaluating with contextualized embeddings and earth mover distance”. *arXiv preprint arXiv:1909.02622*.

Zhao, X., B. He, Y. Wang, Y. Li, F. Mi, Y. Liu, X. Jiang, Q. Liu, and H. Chen. (2021). “UniDS: A Unified Dialogue System for Chit-Chat and Task-oriented Dialogues”. *arXiv preprint arXiv:2110.08032*.

Zhao, X., C. Tao, W. Wu, C. Xu, D. Zhao, and R. Yan. (2019b). “A document-grounded matching network for response selection in retrieval-based chatbots”. In: *IJCAI*.

Zhao, X., W. Wu, C. Tao, C. Xu, D. Zhao, and R. Yan. (2019c). "Low-Resource Knowledge-Grounded Dialogue Generation". In: *International Conference on Learning Representations*.

Zhao, X., W. Wu, C. Xu, C. Tao, D. Zhao, and R. Yan. (2020a). "Knowledge-Grounded Dialogue Generation with Pre-trained Language Models". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 3377–3390.

Zhao, Y., C. Xu, and W. Wu. (2020b). "Learning a Simple and Effective Model for Multi-turn Response Generation with Auxiliary Tasks". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 3472–3483.

Zheng, C., Y. Cao, D. Jiang, and M. Huang. (2020a). "Difference-aware Knowledge Selection for Knowledge-grounded Conversation Generation". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. 115–125.

Zheng, Y., G. Chen, M. Huang, S. Liu, and X. Zhu. (2019a). "Personalized dialogue generation with diversified traits". *arXiv preprint arXiv:1901.09672*.

Zheng, Y., R. Zhang, M. Huang, and X. Mao. (2020b). "A pre-training based personalized dialogue generation model with persona-sparse data". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 34. No. 05. 9693–9700.

Zheng, Z., W. Wang, S. Qi, and S.-C. Zhu. (2019b). "Reasoning visual dialogs with structural and partial observations". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 6669–6678.

Zhong, P., C. Zhang, H. Wang, Y. Liu, and C. Miao. (2020). "Towards Persona-Based Empathetic Conversational Models". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 6556–6566.

Zhong, V. and L. Zettlemoyer. (2019). "E3: Entailment-driven Extracting and Editing for Conversational Machine Reading". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2310–2320.

Zhou, H., M. Huang, T. Zhang, X. Zhu, and B. Liu. (2018a). "Emotional chatting machine: Emotional conversation generation with internal and external memory". In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. No. 1.

Zhou, H., P. Ke, Z. Zhang, Y. Gu, Y. Zheng, C. Zheng, Y. Wang, C. H. Wu, H. Sun, X. Yang, *et al.* (2021a). "EVA: An open-domain chinese dialogue system with large-scale generative pre-training". *arXiv preprint arXiv:2108.01547*.

Zhou, H., T. Young, M. Huang, H. Zhao, J. Xu, and X. Zhu. (2018b). "Commonsense knowledge aware conversation generation with graph attention." In: *IJCAI*. 4623–4629.

Zhou, K., S. Prabhumoye, and A. W. Black. (2018c). "A Dataset for Document Grounded Conversations". In: *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 708–713.

Zhou, K., K. Zhang, Y. Wu, S. Liu, and J. Yu. (2019). "Unsupervised Context Rewriting for Open Domain Conversation". 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)*. 1834–1844.

Zhou, K., W. X. Zhao, S. Bian, Y. Zhou, J.-R. Wen, and J. Yu. (2020a). "Improving conversational recommender systems via knowledge graph based semantic fusion". In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1006–1014.

Zhou, K., Y. Zhou, W. X. Zhao, X. Wang, and J.-R. Wen. (2020b). "Towards Topic-Guided Conversational Recommender System". *arXiv preprint arXiv:2010.04125*.

Zhou, L., J. Gao, D. Li, and H.-Y. Shum. (2020c). "The design and implementation of xiaoice, an empathetic social chatbot". *Computational Linguistics*. 46(1): 53–93.

Zhou, P., K. Gopalakrishnan, B. Hedayatnia, S. Kim, J. Pujara, X. Ren, Y. Liu, and D. Hakkani-Tur. (2021b). "Think Before You Speak: Using Self-talk to Generate Implicit Commonsense Knowledge for Response Generation". *arXiv preprint arXiv:2110.08501*.

Zhou, X. and W. Y. Wang. (2018). "MojiTalk: Generating Emotional Responses at Scale". In: *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1128–1137.

Zhou, X., D. Dong, H. Wu, S. Zhao, D. Yu, H. Tian, X. Liu, and R. Yan. (2016). "Multi-view response selection for human-computer conversation". In: *EMNLP*. 372–381.

Zhou, X., L. Li, D. Dong, Y. Liu, Y. Chen, W. X. Zhao, D. Yu, and H. Wu. (2018d). "Multi-turn response selection for chatbots with deep attention matching network". In: *ACL*.

Zhu, Q., L. Cui, W.-N. Zhang, F. Wei, and T. Liu. (2019). "Retrieval-Enhanced Adversarial Training for Neural Response Generation". In: *ACL*.

Zhuang, Y., X. Wang, H. Zhang, J. Xie, and X. Zhu. (2017). "An Ensemble Approach to Conversation Generation". In: *NLPCC*.

Zou, M., X. Li, H. Liu, and Z.-H. Deng. (2018). "Memd: A diversity-promoting learning framework for short-text conversation". In: *Proceedings of the 27th International Conference on Computational Linguistics*. 1281–1291.