

# Can Question Rewriting Help Conversational Question Answering?

Etsuko Ishii\*, Yan Xu\*, Samuel Cahyawijaya\*, Bryan Wilie

The Hong Kong University of Science and Technology

{eishii, yxucb, scahyawijaya, bwilie}@connect.ust.hk

## Abstract

Question rewriting (QR) is a subtask of conversational question answering (CQA) aiming to ease the challenges of understanding dependencies among dialogue history by reformulating questions in a self-contained form. Despite seeming plausible, little evidence is available to justify QR as a mitigation method for CQA. To verify the effectiveness of QR in CQA, we investigate a reinforcement learning approach that integrates QR and CQA tasks and does not require corresponding QR datasets for targeted CQA. We find, however, that the RL method is on par with the end-to-end baseline. We provide an analysis of the failure and describe the difficulty of exploiting QR for CQA.

## 1 Introduction

The question rewriting (QR) task has been introduced as a mitigation method for conversational question answering (CQA). CQA asks a machine to answer a question based on the provided passage and a multi-turn dialogue (Reddy et al., 2019; Choi et al., 2018), which poses an additional challenge to comprehend the dialogue history. To ease the challenge, QR aims to teach a model to paraphrase a question into a self-contained format using its dialogue history (Elgohary et al., 2019a; Anantha et al., 2021a). Except for Kim et al. (2021), however, no one has provided evidence that QR is effective for CQA in practice. Existing works on QR often (i) depend on the existence of a QR dataset for every target CQA dataset, and (ii) focus more on generating high-quality rewrites than improving CQA performance, making them unsatisfactory for the justification of QR.

To verify the effectiveness of QR, we explore a reinforcement learning (RL) approach that integrates QR and CQA tasks without corresponding labeled QR datasets. In the RL framework, a QR model plays the role of “the agent” that receives

\* Equal Contribution

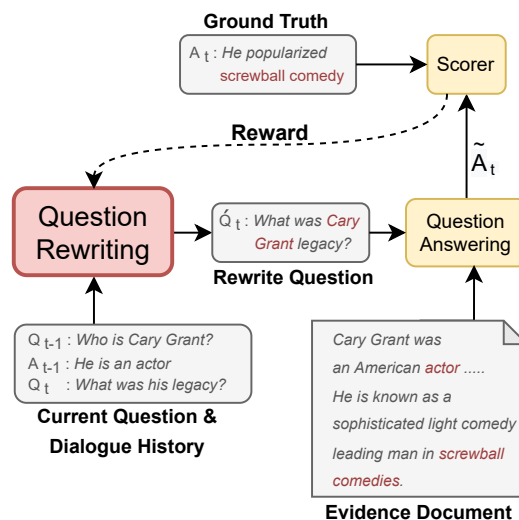


Figure 1: Overview of the RL approach. The current question  $Q_t$  and its dialogue history are reformulated into a self-contained question  $\hat{Q}_t$  by the QR model. Then,  $\hat{Q}_t$  is passed to the QA model to extract an answer span  $\hat{A}_t$  from the evidence document. We train the QR model by maximizing the reward obtained by comparing the predicted answer span  $\hat{A}_t$  with the gold span  $A_t$ .

rewards from a QA model that acts as “the environment.” During training, the QR model aims to maximize the performance on the CQA task by generating better rewrites of the questions.

Despite the potential and plausibility of the RL approach, our experimental results suggest an upper bound of the performance, and it is on par with the baselines without QR. In this paper, we provide analysis to (i) understand the reason for the failure of the RL approach and (ii) reveal that QR cannot improve CQA performance even with the non-RL approaches. The code is available at <https://github.com/HLTCHKUST/cqr4cqa>.

## 2 Related Work

The CQA task aims to assist users in seeking information (Reddy et al., 2019; Choi et al., 2018; Campos et al., 2020). The key challenge is to re-

Models	CoQA						QuAC			
	Overall F1	Child.	Liter.	M&H	News	Wiki.	F1	HEQ-Q	HEQ-D	
	end-to-end	84.5	<b>84.4</b>	82.4	<b>82.9</b>	86.0	86.9	<b>67.8</b>	63.5	7.9
QReCC	pipeline	82.9	82.9	80.9	81.5	84.4	84.8	66.3	62.0	6.6
	ours	<b>84.7</b>	<u>84.3</u>	<b>83.1</b>	<u>82.7</u>	<b>86.3</b>	<u>86.8</u>	<u>67.6</u>	<u>63.2</u>	<u>7.8</u>
CANARD	pipeline	82.8	83.4	80.1	80.8	84.4	85.6	66.5	62.5	7.4
	EXCORD <sup>†</sup>	83.4 (+0.6)	<b>84.4 (1.9)</b>	81.2 (+1.0)	79.8 (-0.3)	84.6 (+0.3)	<b>87 (0.0)</b>	<u>67.7 (+1.2)</u>	<b>64.0 (+1.6)</b>	<b>9.3 (+2.1)</b>
	ours	<u>84.4</u>	<u>84.1</u>	<u>82.7</u>	<u>82.6</u>	<u>86.0</u>	<u>86.7</u>	<u>67.4</u>	<u>62.7</u>	<u>8.1</u>

Table 1: Evaluation results of our approach and baselines on the test set. EXCORD<sup>†</sup> follows the results reported by Kim et al. (2021) and ( $\pm x.x$ ) indicate the improvement compared to their original baseline. **Bold** are the best results amongst all. Underlined represents the best score on each combination of the CoQA and QR datasets.

solve the conversation history and understand a highly-contextualized question. Most prior works focus on model structures (Zhu et al., 2018; Yeh and Chen, 2019; Zhang et al., 2021b; Zhao et al., 2021) or training techniques (Ju et al., 2019; Xu et al., 2021) to improve the performance. QR tasks have been proposed to further improve CoQA systems by paraphrase a question into a self-contained styles (Elgohary et al., 2019a; Petrén Bach Hansen and Søgaard, 2020; Anantha et al., 2021a). While many of the existing works on QR put more effort toward generating high-quality rewrites (Lin et al., 2020; Vakulenko et al., 2021), Kim et al. (2021) introduced a framework to leverage QR to finetune CoQA models with a consistency-based regularization. QR has also been studied in single-turn QA and other information-seeking tasks (Nogueira and Cho, 2017; Buck et al., 2018).

### 3 Methodology

We denote a CoQA dataset as  $\{\mathcal{D}^n\}_{n=1}^N$  and the dialogue history at turn  $t$  as  $\mathcal{D}_t = \{(Q_i, A_i)\}_{i=1}^t$ , where  $Q_t$  is the question and  $A_t$  is the answer. Along with the QA pairs, the corresponding evidence documents  $Y_t$  are also given.

As depicted in Figure 1, our proposed RL framework involves a QA model as an environment and a QR model as an agent. Let  $\hat{Q}_t = \{\hat{q}_l\}_{l=1}^L$  denote a generated rewritten question sequence of  $Q_t$ . The objective of the QR model is to rewrite the question  $Q_t$  at turn  $t$  into a self-contained version, based on the current question and the dialogue history  $\mathcal{D}_{t-1}$ . The agent takes an input state  $X_t = (\mathcal{D}_{t-1}, Q_t)$  and generates a paraphrase  $\hat{Q}_t$ . Then,  $\hat{X}_t = (\mathcal{D}_{t-1}, \hat{Q}_t)$  and an evidence document  $Y_t$  are provided to an environment, namely, the QA model  $f_\phi$ , which extracts an answer span  $\tilde{A}_t = f_\phi(\hat{X}_t, Y_t)$ . We aim for the agent, a QR model  $\pi_\theta$ , to learn to generate a high-quality para-

phrase of the given question based on the reward received from the environment.

The policy, in our case the QR model, assigns probability

$$\pi_\theta(\hat{Q}_t | X_t) = \prod_{l=1}^L p(\hat{q}_l | \hat{q}_1, \dots, \hat{q}_{l-1}, X_t). \quad (1)$$

Our goal is to maximize the expected reward of the answer returned under the policy, namely,

$$\mathbb{E}_{\hat{q}_t \sim \pi_\theta(\cdot | q_t)} [r(f_\phi(\hat{X}_t))], \quad (2)$$

where  $r$  is a reward function. We apply the token-level F1-score between the predicted answer span  $\tilde{A}_t$  and the gold span  $A_t$  as the reward  $r$ . We can directly optimize the expected reward in Eq. 2 using RL algorithms.

Prior to the training process, the QA model  $f_\phi$  is fine-tuned on  $\{\mathcal{D}^n\}$  and the QR model is initialized with  $\pi_\theta = \pi_{\theta_0}$ , where  $\pi_{\theta_0}$  is a pretrained language model. We apply Proximal Policy Optimization (PPO) (Schulman et al., 2017; Ziegler et al., 2019) to train  $\pi_\theta$ . PPO is a policy gradient method which alternates between sampling data through interaction with the environment and optimizing a surrogate objective function via stochastic gradient ascent. Following Ziegler et al. (2019), we apply a KL-penalty to the reward  $r$  so as to prevent the policy  $\pi_\theta$  from drifting too far away from  $\pi_{\theta_0}$ :

$$R_t = R(\hat{X}_t) = r(f_\phi(\hat{X}_t)) - \beta \text{KL}(\pi_\theta, \pi_{\theta_0}),$$

where  $\beta$  represents a weight factor and  $R_t$  is the modified reward of  $r$ .

## 4 Experiments

### 4.1 Setup

We use a pretrained RoBERTa (Liu et al., 2019) model as the initial QA model and adapt it to the

	Question	F1 Score		Question	F1 Score
$Q_t$	What is the Vat the <b>library</b> of?	1.0	$Q_t$	Where <b>did</b> the band The Smashing Pumpkins put on display?	1.0
$\hat{Q}_t$	What is the Vat the <b>Library</b> of?	0.22	$\hat{Q}_t$	Where <b>was</b> the band The Smashing Pumpkins put on display?	0.0
$Q_t$	What was <b>everybody</b> doing?	0.91	$Q_t$	Which company produced the movie <b>Island of Misfit Toys</b> ?	1.0
$\hat{Q}_t$	What was <b>everyone</b> doing?	0.0	$\hat{Q}_t$	Which company produced the movie, <b>The Island of Misfit Toys</b> ?	0.0

Table 2: Minor modification of questions may cause a drastic change in CQA performance.

CQA tasks. For the QR models, we leverage pre-trained GPT-2 (Radford et al., 2019) and first fine-tune them with QR datasets for better initialization. We attempt three settings: (a) directly fine-tune the QA model on the CQA datasets (end-to-end), (b) fine-tune the QA model with questions rewritten by the QR model (pipeline), and (c) train the QR model based on the reward obtained from the QA model. More details of the experiments can be found in Appendix A.

**Datasets** We conduct our experiments on two crowd-sourced CQA datasets, CoQA (Reddy et al., 2019) and QuAC (Choi et al., 2018). Since the test set is not publicly available for both CoQA and QuAC, following Kim et al. (2021), we randomly sample 5% of dialogues in the training set and adopt them as our validation set and report the test results on the original development set for the CoQA experiments. We apply the same split as Kim et al. (2021) for the QuAC experiments.

For the QR model pre-training, we use two QR datasets: QReCC (Anantha et al., 2021b) and CANARD (Elgohary et al., 2019b). CANARD is generated by rewriting a subset of the original questions in the QuAC datasets, and contains 40K questions in total. QReCC is built upon three publicly available datasets: QuAC, TREC Conversational Assistant Track (CAST) (Dalton et al., 2020) and Natural Questions (NQ) (Kwiatkowski et al., 2019). QReCC contains 14K dialogues with 80K questions, and 9.3K dialogues are from QuAC.

**Evaluation Metrics** Following the leaderboards, we utilize the unigram F1 score to evaluate the QA performance. In CoQA evaluation, the QA models are also evaluated with the domain-wise F1 score. In QuAC evaluation, we incorporate the human equivalence score HEQ-Q and HEQ-D as well. HEQ-Q indicates the percentage of questions on which the model outperforms human beings and HEQ-D represents the percentage of dialogues on which the model outperforms human beings for all questions in the dialogue.

## 4.2 Results

We report our experimental results in Table 1. We see that our RL approach yields 0.9–1.6 F1 improvement over the pipeline setting regardless of the dataset combinations and performs almost as well as the end-to-end setting. This partially supports our expectation that RL lifts the CQA performance. However, we find it almost impossible to bring significant improvement over the end-to-end baseline despite our extensive trials. One reason why we cannot provide as much improvement as reported in Kim et al. (2021) would be related to the inputs of the QA model. Their EXCORD feeds the original questions together with the rewritten questions, whereas we only use the rewritten questions. It is also noteworthy that their results are consistently lower than ours, even lower than our end-to-end settings.

Our inspection of the questions generated by the QR models reveals that the models learn to copy the original questions by PPO training, and this is the direct reason that our method cannot outperform the end-to-end baselines. Indeed, on average, 89.6% of the questions are the same as the original questions after PPO training, although this value is 34.5% in the pipeline settings. We also discover a significant correlation between the performance and how much the QR models copy the original question (the correlation coefficient is 0.984 for CoQA and 0.967 for QuAC) and the edit distance from the original question (the correlation coefficient is -0.996 for CoQA and -0.989 for QuAC).

## 5 Discussion

In this section, we provide an analysis to (i) raise a sensitivity problem of the QA model to explain the failure of RL and (ii) disclose that there is no justification for QR, even in the non-RL approaches.

### 5.1 Sensitivity of the QA model

It appears that the QA models are more sensitive to trivial changes than the reward models in other successful language generation tasks, and this could

Perturb	Sentiment Analysis		CQA	
	Amazon	Yelp	CoQA	QuAC
<b>Original</b>	95.8	98.2	84.5	67.8
<b>UPC</b>	95.8 (-)	96.7 <b>(-1.5)</b>	74.8 <b>(-9.8)</b>	57.4 <b>(-10.5)</b>
<b>SLW</b>	91.9 <b>(-3.9)</b>	97.0 <b>(-1.1)</b>	83.0 <b>(-1.6)</b>	66.7 <b>(-1.1)</b>
<b>WIF</b>	94.3 <b>(-1.5)</b>	97.7 <b>(-0.5)</b>	82.6 <b>(-2.0)</b>	65.6 <b>(-2.2)</b>
<b>SPP</b>	94.8 <b>(-1.0)</b>	97.7 <b>(-0.5)</b>	78.3 <b>(-6.2)</b>	65.5 <b>(-2.4)</b>

Table 3: Robustness test on Sentiment Analysis and CQA tasks. We apply four perturbations: **UPC** (upper casing), **SLW** (slang word), **WIF** (word inflection), and **SPP** (sentence paraphrasing).

Datasets	QuAC Model		CANARD Model	
	F1	EM	F1	EM
QuAC	67.7	51.5	62.9	46.8
CANARD	65.1	49.9	63.3	46.9

Table 4: Results of the supervised learning approach. “XX Model” denotes the QA model trained on XX, and EM the percentage of the predictions the same as the gold.

account for our lower performance on CQA. As can be seen from the examples in Table 2, a subtle alteration such as uppercasing or replacement with synonyms can significantly change F1 scores.

To quantify the sensitivity of the reward models, we compare model robustness between our QA models and sentiment analysis models that have been reported in Ziegler et al. (2019) to be effective for stylistic language generation. We adopt publicly available models that are fine-tuned sentiment analysis datasets: BERT-based trained on Amazon polarity (McAuley and Leskovec, 2013)<sup>1</sup> and RoBERTa-base trained on Yelp polarity (Zhang et al., 2015)<sup>2</sup>. To test the robustness of the models, we introduce small perturbations to the samples in the test set using the NL-Augmenter toolkit (Dhole et al., 2021), and compare F1 scores on each task (experimental details in Appendix B).

Based on the robustness test given in Table 3, the QA models are shown to be significantly less robust against most perturbations compared to the sentiment analysis models. It is conceivable that this sensitivity of the QA model leads to a sparse reward problem for the agent, which causes instability for the model learning the optimal policy. An important direction for future studies is to ease the sparse reward problem by, for example, enhancing the robustness of the QA models.

<sup>1</sup>[https://huggingface.co/fabriceyh/bert-base-uncased-amazon\\_polarity](https://huggingface.co/fabriceyh/bert-base-uncased-amazon_polarity)

<sup>2</sup><https://huggingface.co/VictorSanh/roberta-base-finetuned-yelp-polarity>

Datasets	CoQA		QuAC	
	F1	EM	F1	EM
end-to-end	<b>84.5</b>	<b>76.4</b>	<b>67.83</b>	51.47
QReCC	84.1	76.0	<b>67.83</b>	51.48
CANARD	83.7	75.8	67.81	<b>51.50</b>

Table 5: Results of the data augmentation approach. EM denotes the percentage of the predictions the same as the gold.

## 5.2 Can QR Help in Non-RL Approaches?

First, we evaluate with a simple supervised learning approach using rewrites provided by CANARD. Extracting the QuAC samples that have a CANARD annotation, we (i) evaluate the CANARD annotations with the QA model trained on QuAC (the model used in the main experiments) and (ii) train another QA model with the CANARD annotations. Training is under the same conditions of the QA model initialization as in the main experiments. As the results in Table 4 show, we can hardly observe the effectiveness of the CANARD annotations. This supports the claim in Buck et al. (2018) that better rewrites in the human eye are not necessarily better for machines and implies the difficulty of exploiting QR for CQA.

Moreover, we explore a data-augmentation approach to integrate QR and CQA. First, we generate ten possible rewrites using top- $k$  sampling (Zhang et al., 2021a) for all the questions of the CQA datasets. To guarantee the quality of the rewrites, we select the best F1 scoring ones from every ten candidates and use them to teach another QR model how to reformulate questions (experimental details in Appendix C). As the results in Table 5 show, we consistently get worse scores compared to the end-to-end settings in CoQA, and almost the same scores for QuAC, not finding justification to apply QR in the manner of the data augmentation approach.

## 6 Conclusion

In this paper, we explore the RL approach to verify the effectiveness of QR in CQA, and report that the RL approach is on par with simple end-to-end baselines. We find the sensitivity of the QA models would disadvantage the RL training. Future work is needed to verify that QR is a promising mitigation method for CQA since even the non-RL approaches perform unsatisfactorily.

## References

- Raviteja Anantha, Svitlana Vakulenko, Zhucheng Tu, Shayne Longpre, Stephen Pulman, and Srinivas Chappidi. 2021a. [Open-domain question answering goes conversational via question rewriting](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 520–534, Online. Association for Computational Linguistics.
- Raviteja Anantha, Svitlana Vakulenko, Zhucheng Tu, Shayne Longpre, Stephen Pulman, and Srinivas Chappidi. 2021b. [Open-domain question answering goes conversational via question rewriting](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 520–534.
- Christian Buck, Jannis Bulian, Massimiliano Ciaramita, Wojciech Gajewski, Andrea Gesmundo, Neil Houlsby, and Wei Wang. 2018. [Ask the right questions: Active question reformulation with reinforcement learning](#). In *International Conference on Learning Representations*.
- Jon Ander Campos, Arantxa Otegi, Aitor Soroa, Jan Derru, Mark Cieliebak, and Eneko Agirre. 2020. [DoQA - accessing domain-specific FAQs via conversational QA](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7302–7314, Online. Association for Computational Linguistics.
- Eunsol Choi, He He, Mohit Iyyer, Mark Yatskar, Wentaoh Yih, Yejin Choi, Percy Liang, and Luke Zettlemoyer. 2018. [QuAC: Question answering in context](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2174–2184, Brussels, Belgium. Association for Computational Linguistics.
- Jeffrey Dalton, Chenyan Xiong, and Jamie Callan. 2020. [Trec cast 2019: The conversational assistance track overview](#). *arXiv preprint arXiv:2003.13624*.
- Kaustubh D. Dhole, Varun Gangal, Sebastian Gehrmann, Aadesh Gupta, Zhenhao Li, Saad Mahamood, Abinaya Mahendiran, Simon Mille, Ashish Srivastava, Samson Tan, Tongshuang Wu, Jascha Sohl-Dickstein, Jinho D. Choi, Eduard Hovy, Ondrej Dusek, Sebastian Ruder, Sajant Anand, Nandgender Aneja, Rabin Banjade, Lisa Barthe, Hanna Behnke, Ian Berlot-Attwell, Connor Boyle, Caroline Brun, Marco Antonio Sobrevilla Cabezudo, Samuel Cahyawijaya, Emile Chapuis, Wanxiang Che, Mukund Choudhary, Christian Clauss, Pierre Colombo, Filip Cornell, Gautier Dagan, Mayukh Das, Tanay Dixit, Thomas Dopierre, Paul-Alexis Dray, Suchitra Dubey, Tatiana Ekeinhor, Marco Di Giovanni, Rishabh Gupta, Rishabh Gupta, Louanes Hamla, Sang Han, Fabrice Harel-Canada, Antoine Honore, Ishan Jindal, Przemyslaw K. Joniak, Denis Kleyko, Venelin Kovatchev, Kalpesh Krishna, Ashutosh Kumar, Stefan Langer, Seungjae Ryan Lee, Corey James Levinson, Hualou Liang, Kaizhao Liang, Zhexiong Liu, Andrey Lukyanenko, Vukosi Marivate, Gerard de Melo, Simon Meoni, Maxime Meyer, Afnan Mir, Nafise Sadat Moosavi, Niklas Muennighoff, Timothy Sum Hon Mun, Kenton Murray, Marcin Namysl, Maria Obedkova, Priti Oli, Nivranshu Pasricha, Jan Pfister, Richard Plant, Vinay Prabhu, Vasile Pais, Libo Qin, Shahab Raji, Pawan Kumar Rajpoot, Vikas Raunak, Roy Rinberg, Nicolas Roberts, Juan Diego Rodriguez, Claude Roux, Vasconcellos P. H. S., Ananya B. Sai, Robin M. Schmidt, Thomas Scialom, Tshephisho Sefara, Saqib N. Shamsi, Xudong Shen, Haoyue Shi, Yiwen Shi, Anna Shvets, Nick Siegel, Damien Sileo, Jamie Simon, Chandan Singh, Roman Sitelew, Priyank Soni, Taylor Sorensen, William Soto, Aman Srivastava, KV Aditya Srivatsa, Tony Sun, Mukund Varma T, A Tabassum, Fiona Anting Tan, Ryan Teehan, Mo Tiwari, Marie Tolkiehn, Athena Wang, Zijian Wang, Gloria Wang, Zijie J. Wang, Fuxuan Wei, Bryan Wilie, Genta Indra Winata, Xinyi Wu, Witold Wydmański, Tianbao Xie, Usama Yaseen, M. Yee, Jing Zhang, and Yue Zhang. 2021. [NL-augmenter: A framework for task-sensitive natural language augmentation](#).
- Ahmed Elgohary, Denis Peskov, and Jordan Boyd-Graber. 2019a. [Can you unpack that? learning to rewrite questions-in-context](#). 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)*, pages 5918–5924, Hong Kong, China. Association for Computational Linguistics.
- Ahmed Elgohary, Denis Peskov, and Jordan Boyd-Graber. 2019b. [Can you unpack that? learning to rewrite questions-in-context](#). 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)*, pages 5918–5924.
- Ying Ju, Fubang Zhao, Shijie Chen, Bowen Zheng, Xuefeng Yang, and Yunfeng Liu. 2019. [Technical report on conversational question answering](#). *arXiv preprint arXiv:1909.10772*.
- Gangwoo Kim, Hyunjae Kim, Jungsoo Park, and Jae-woo Kang. 2021. [Learn to resolve conversational dependency: A consistency training framework for conversational question answering](#). 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)*, pages 6130–6141, Online. Association for Computational Linguistics.
- Tom Kwiatkowski, Jennimaria Palomaki, Olivia Redfield, Michael Collins, Ankur Parikh, Chris Alberti, Danielle Epstein, Illia Polosukhin, Jacob Devlin, Kenton Lee, et al. 2019. [Natural questions: A benchmark for question answering research](#). *Transactions of the*

- Association for Computational Linguistics*, 7:452–466.
- Sheng-Chieh Lin, Jheng-Hong Yang, Rodrigo Nogueira, Ming-Feng Tsai, Chuan-Ju Wang, and Jimmy Lin. 2020. Conversational question reformulation via sequence-to-sequence architectures and pretrained language models. *arXiv preprint arXiv:2004.01909*.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Julian McAuley and Jure Leskovec. 2013. **Hidden factors and hidden topics: Understanding rating dimensions with review text**. In *Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13*, page 165–172, New York, NY, USA. Association for Computing Machinery.
- Rodrigo Nogueira and Kyunghyun Cho. 2017. **Task-oriented query reformulation with reinforcement learning**. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 574–583, Copenhagen, Denmark. Association for Computational Linguistics.
- Victor Petrén Bach Hansen and Anders Søgaard. 2020. **What do you mean ‘why?’: Resolving sluices in conversations**. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(05):7887–7894.
- Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.
- Siva Reddy, Danqi Chen, and Christopher D. Manning. 2019. **CoQA: A conversational question answering challenge**. *Transactions of the Association for Computational Linguistics*, 7:249–266.
- John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. **Proximal policy optimization algorithms**. *CoRR*, abs/1707.06347.
- Svitlana Vakulenko, Shayne Longpre, Zhucheng Tu, and Raviteja Anantha. 2021. Question rewriting for conversational question answering. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 355–363.
- Yan Xu, Etsuko Ishii, Genta Indra Winata, Zhaojiang Lin, Andrea Madotto, Zihan Liu, Peng Xu, and Pascale Fung. 2021. Caire in dialdoc21: Data augmentation for information seeking dialogue system. In *Proceedings of the 1st Workshop on Document-grounded Dialogue and Conversational Question Answering (DialDoc 2021)*, pages 46–51.
- Yi-Ting Yeh and Yun-Nung Chen. 2019. Flowdelta: Modeling flow information gain in reasoning for conversational machine comprehension. In *Proceedings of the 2nd Workshop on Machine Reading for Question Answering*, pages 86–90.
- Hugh Zhang, Daniel Duckworth, Daphne Ippolito, and Arvind Neelakantan. 2021a. **Trading off diversity and quality in natural language generation**. In *Proceedings of the Workshop on Human Evaluation of NLP Systems (HumEval)*, pages 25–33, Online. Association for Computational Linguistics.
- Xiang Zhang, Junbo Zhao, and Yann LeCun. 2015. **Character-level Convolutional Networks for Text Classification**. *arXiv:1509.01626 [cs]*.
- Zhuosheng Zhang, Junjie Yang, and Hai Zhao. 2021b. **Retrospective reader for machine reading comprehension**. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(16):14506–14514.
- Jing Zhao, Junwei Bao, Yifan Wang, Yongwei Zhou, Youzheng Wu, Xiaodong He, and Bowen Zhou. 2021. **RoR: Read-over-read for long document machine reading comprehension**. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1862–1872, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Chenguang Zhu, Michael Zeng, and Xuedong Huang. 2018. Sdnet: Contextualized attention-based deep network for conversational question answering. *arXiv preprint arXiv:1812.03593*.
- Daniel M Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. 2019. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*.