Review of Knowledge-Grounded Conversational Data Augmentation with Generative Conversational Networ

First Author[1][0000−1111−2222−3333], Second Author[2,3][1111−2222−3333−4444], and Third Author[3][2222−−3333−4444−5555]

[1] Princeton University, Princeton NJ 08544, USA
[2] Springer Heidelberg, Tiergarten. 17, 69121 Heidelberg, Germany
lncs@springer.com
http://www.springer.com/gp/computer-science/lncs
[3] ABC Institute, Rupert-Karls-University Heidelberg, Heidelberg, Germany
{abc,lncs}@uni-heidelberg.de

Abstract. Conversational AI systems have been constrained by the scarcity of extensive and excellent-quality conversational data. In their research, the authors suggest a technique for enhancing conversational datasets using generative conversational models which are founded on external knowledge sources.. The authors present a Generative Conversational Network (GCN) trained on a large conversational dataset and fine-tuned task-specific knowledge, which can generate new conversational examples. According to the findings presented in the paper, the performance of conversational AI systems can be enhanced by utilizing the generated data. Additionally, the study highlights that the quality of the generated data relies significantly on the caliber of external knowledge sources and the fine-tuning process that is specific to the task at hand. This paper provides a new approach for augmenting conversational datasets and highlights the importance of knowledge-grounding in conversational data generation.

Keywords: Conversational AI · Generative Conversational Networks (GCN) · Conversational Data Augmentation · Fine-Tuning · Knowledge-Grounding · Conversational Datasets.

1 Introduction

The paper discusses the challenges in developing conversational AI systems, including the difficulties in handling idioms, humor, empathy, and unstructured knowledge due to the scarcity of large and diverse conversational datasets. The authors note the recent progress in this field, primarily due to the emergence of large pre-trained language models and commercial conversational agents like Siri, Alexa, and Google Assistant. However, despite the ongoing efforts of the research community to collect diverse datasets, such as empathetic [1] or persuasive dialogues[2], the amount of data remains insufficient for training deep neural networks. Additionally, data collection efforts often target specific phenomena and may not capture the full richness of human conversations, limiting their scalability.

One major obstacle to implementing conversational AI systems in real-world settings is privacy concerns, which restrict the use of publicly available conversational data[8]. Overall, the paper highlights the need for more effective methods for generating large and diverse conversational datasets that address various aspects of human communication. The authors suggest that the proposed GCN approach has the potential to address this challenge by generating high-quality conversational data from unstructured textual knowledge.

1.1 Motivation

The paper presents an interesting approach to generating conversational data using Generative Conversational Networks (GCN) [4]. The authors extended the GCN approach to generate responses based on unstructured knowledge. The study involves using a generator model to produce new data and a separate learner model to train on this data. The performance of the learner model was used as a reward signal to train the generator, resulting in optimized, labeled, diverse, and targeted data. The approach can guide data generation towards various dimensions of interest, such as knowledge-grounded, empathetic, or polite dialogues, using automatic metrics or human feedback. Overall, the paper presents a promising method for generating diverse and high-quality conversational data.

1.2 Problem Definition

After reviewing the paper, it appears that the authors aim to improve conversational AI by testing the performance of Generative Conversational Networks (GCN) in response generation tasks under two contexts: open-domain and knowledge-grounded conversations. The paper addresses the challenge of the lack of large and diverse conversational data by generating high-quality conversational data that includes complex aspects of human communication. The study aims to answer research questions such as whether GCN can generate high-quality conversational data from unstructured textual knowledge and how its performance compares to a baseline that uses fine-tuning on seed data.

The paper describes the use of a generator to create open-domain dialogues and then trains a conversational agent on this data. The authors address the challenge of selecting an appropriate reward signal to generate good-quality dialogues that do not exist in the training data. Overall, the approach presented in the paper is interesting and presents a potential solution to generating diverse and high-quality conversational data for conversational AI.

As a reviewer, I appreciate the authors' approach to addressing the challenge of generating high-quality conversational data for conversational AI. However, I note that the potential negative impacts of technology, such as perpetuating discrimination and the displacement of human workers, must be considered. Additionally, further research is needed to improve the relevance of the generated responses and incorporate additional knowledge sources to generate more informative dialogues.

2 Methods and Evaluation

The method for addressing the problem in this work involves presenting the theoretical basis for knowledge-grounded conversational data augmentation and reviewing the specific literature on the topic [4]. The approach is based on Generative Conversational Networks (GCN), which is a meta-learning method that can be used to generate diverse and targeted data for various conversational AI tasks. The GCN consists of two models: a data generator and a learner, where the generator produces new data, and the quality of the generated data is used as a signal reward to train the generator [3]. The authors aim to enhance the generated data quality in successive iterations, with the ultimate objective of enabling the GCN to achieve the desired properties for knowledge-grounded or open-domain dialogues.

The specific literature reviewed in this work focuses on previous work in the area of knowledge-grounded conversational data augmentation, including the use of GCN for generating data for intent detection and slot-filling tests [3]. The method also involves evaluating the performance of different knowledge retrievers for selecting relevant knowledge pieces from the current dialogue context and generating a response based on the retrieved knowledge and context [9].

2.1 Methods

Conversational AI systems often face the problem of limited training data, which affects their performance and ability to generalize to new contexts. To address this issue, the theoretical basis for knowledge-grounded conversational data augmentation is presented in this work. Specifically, the paper reviews the literature on Generative Conversational Networks (GCN), a meta-learning method that generates diverse and targeted data for various conversational AI tasks. The GCN (Fig1) comprises two models: a data generator and a learner, where the generator produces new data, and the quality of the generated data is used as a signal reward to train the generator. The authors aim to enhance the quality of the generated data over time, with the aim of guiding the GCN towards desired characteristics in open-domain or knowledge-grounded dialogues.

Previous work in the area of knowledge-grounded conversational data augmentation is reviewed in this paper. Specifically, the focus is on the use of GCN for generating data for intent detection and slot-filling tests. The method involves evaluating the performance of different knowledge retrievers for selecting relevant knowledge pieces from the current dialogue context and generating a response based on the retrieved knowledge and context. Research has shown the effectiveness of this approach in the identification of intentions and tagging of slots within goal-oriented conversations [3], and in this work, it is applied to train social conversational agents.

In contrast to Generative Adversarial Networks, which aim to imitate the input data, GCN models rely on an external reward signal that does not require differentiability and can therefore achieve better generalization. The optimization criteria can be set based on various factors, such as the tone of the conversations, the level of technicality, the use of different dialects, knowledge grounding, and even subjective factors such as engagement ratings provided by human evaluators. The models can then be guided toward these desired directions. For open-domain and knowledge-grounded conversations and few-shot experiments using 10% and 1-10% of the data are conducted, respectively, as proof of concept.

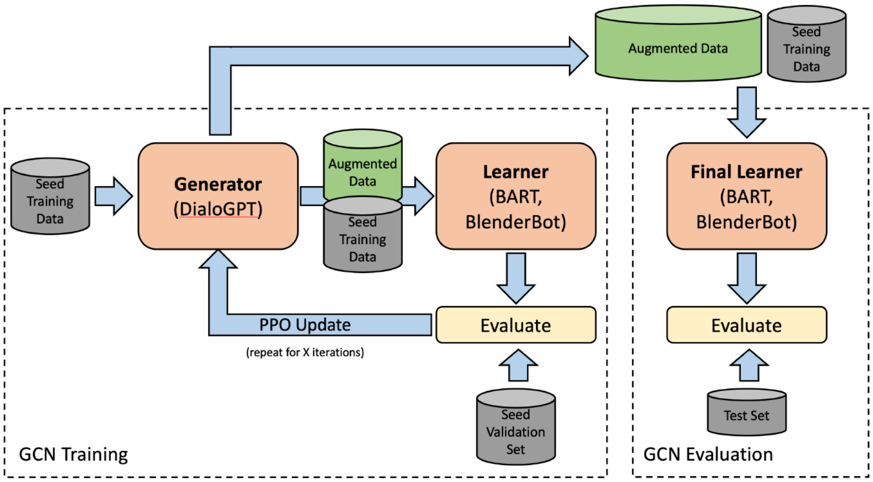

Fig. 1. The architecture of our approach using Generative Conversational Networks for knowledge-grounded dialogues. The generator is first fine-tuned with seed data and produces an augmented dataset and those data are used to train a learner. The performance of the learner on a held-out validation set (along with auxiliary metrics) is used as a reward to update the generator.

Seed Data:
In this paper, the focus is on testing response generation in two different contexts: open domain and knowledge-grounded conversation. To study the effectiveness of our approach, a limited amount of data is used for training the model. For the open domain conversation, 10% of the original data is randomly selected as the seed data for training. On the other hand, for the knowledge-grounded conversation, three different proportions of the original data are used as seed data for training, which are 1%, 5%, and 10%. This setup enables us to study

the impact of the amount of seed data on the performance of the model in low-resource settings.

Data Generator:
The data generator is an essential component of the GCN approach. In this work, they have chosen to use a small configuration of Dialog GPT as the initial model for the data generator. The role of the data generator is to produce new data that can be used to train the learner. The quality of the generated data is used as a signal or reward to train the generator, such that over time, the quality of the generated data improves. This reverse signal can guide the data generation toward desired properties in knowledge-grounded or open-domain dialogues.

Pre-Training:
The data generator is pre-trained on a sample of conversational data. For open-domain conversations, the input to the generator is just the dialog context, while for knowledge-grounded conversations, the input is the dialog context and the selected knowledge pieces. Both the target outputs are the next response in the conversation. The generator uses a dialog GPT small configuration as an initial model and is fine-tuned on the sample conversational data.

Learner Training:
The Learner is trained using a cross-entropy loss, where at every iteration a new Learner is spawned from the latest models and trained on both the augmented set and the seed training sets.

Learner Evaluation:
The Learner is evaluated on a separate validation set to compute a numerical reward. For the open domains, the reward is based on a combination of BLEU and ROUGE scores Table 1. For Knowledge grounded, the reward is based on a combination of the BLEU score, Knowledge F1, and Not only F1, which measures the overlap between the produced response and the ground truth knowledge.

Generator Update:
The generator is updated based on the learner's validation performance and a combination of generation metrics and a regularization term to prevent divergence from a reference language model. This is done using Proximal Policy Optimization with modifications.

Final Learner:
The final learner is created by selecting the best generator checkpoint based on the learner's performance on the validation set. The final learner is then trained on the original data and a large augmented data set is created using the best-performing generator. If the augmented data is high quality, the final learner is expected to outperform as it is trained with more data. The results presented in the paper are based on this final learner.

2.2 Evaluation

Experiment: Open Domain In the open-domain conversation experiment, 10% of the topical chat dataset was used as seed data for both the data generator and the learner. The performance of the GCN learner was compared to three baselines: 1) 10%, which was trained only on the same seed data [4], 2) 100%, which was trained on the full shot of the topical chat dataset, and 3) GCN without reinforcement learning (RL), which was trained on a combination of the seed data and synthetic data generated by the data generator without updating from RL. The results of the experiment will show the effectiveness of the GCN reinforcement learning approach compared to these baseline models.

Evaluation: Open Domain In the open-domain experiment, the GCN learner was compared to three baselines. The first baseline was trained only on the seed data, the second was trained on the full topical chat dataset and the third was a GCN model without reinforcement learning. The results showed that the GCN learner outperformed the first baseline and was close to the second in terms of automatic metrics. However, the GCN learner produced less relevant but more engaging conversations according to human evaluations. The GCN learner with reinforcement learning outperformed the first baseline and was very close to the second in human evaluations.

Experiment: Knowledge Grounded In the knowledge-grounded conversation experiment, the authors use a smaller seed set of the Topical Chat data and use Dialogue GPT as the generator and BlenderBot as the learner. The performance of the learner with GCN is compared against two baselines: BlenderBot trained only on the seed set and GCN without reinforcement learning, which is trained on the seed set and a synthetic set generated by a pre-trained language model. The results of this experiment will help to assess the effectiveness of the GCN reinforcement learning approach in knowledge-grounded conversation.

Automatic Evaluation: Knowledge Grounded Perplexity (PPL), BLEU-4 [6] with the smoothing functions from [10], and KF1 were used as automated metrics to evaluate the performance of the conversational agents. The frequent test set of the topical chat dataset was used for calculating these metrics, and the results are shown in Tables 1 and 2. In the open-domain conversation experiment, the BART 10% model outperformed the GCN agents on all automated metrics [11]. On the other hand, in the knowledge-grounded conversation experiment, the GCN+RL model was able to incorporate more knowledge, as demonstrated by its higher KF1 score [11].

Human Evaluation: According to the paper, the intrinsic one-to-many nature of conversations suggests that reference-based metrics may not always align with human ratings. To account for this, the authors conducted a human evaluation

Table 1. Automatic and human evaluation results. Human evaluators rate responses on a scale of 1 to 5. BScore stands for BERTScore. Bold indicates a statistically significant difference (t-test assuming unequal variance). BART (100%) and BART (10%) are BART trained on 100% and 10% of the data, GCN-RL is GCN without RL, and GCN+RL is GCN with RL training.

Model	**BLEU**	**Rouge(1/2/L)**	**BScore**	**Engaging**	**Fluency**	**Rel.**	**Overall**
Data	-	-	-	3.85	4.55	3.77	4.06
BART (100%)	3.1	20.3/6.1/17.8	0.861	3.80	4.58	3.68	4.02
BART (10%)	**2.0**	**18.5/4.2/16.0**	**0.858**	3.63	4.50	**3.62**	3.92
GCN-RL	1.1	15.0/2.1/12.6	0.850	3.70	4.47	3.47	3.88
GCN+RL	1.3	15.8/2.7/13.6	0.851	**3.79**	4.49	3.58	**3.96**

Table 2. Results of automated evaluation on knowledge-grounded conversations. All models try to maximize KF1, and the baseline is the same model as the GCN learners (BBs: BlenderBot-small, 90M parameters).

	1% data			**5% data**			**10% data**		
Model	PPL	KF1	BL-4	PPL	KF1	BL-4	PPL	KF1	BL-4
BBs	23.39	0.10	0.07	23.52	0.17	0.09	21.69	0.17	0.09
GCN-RL	26.47	0.15	0.08	24.54	0.18	0.09	23.11	0.18	0.09
GCN+RL	27.11	**0.20**	0.08	24.60	**0.25**	0.14	23.67	**0.28**	0.10

of the output of the GCN learner, baselines, and ground truth. Human evaluators rated how engaging, fluent, and relevant each response was on a scale of 1 to 5. The results showed that in the open-domain condition, the GCN learner produced engaging but less relevant conversations [5], possibly because the model inserted facts or outputs that were not entirely relevant but perceived as more engaging. In the knowledge-grounded setting, where GCN was explicitly trained to optimize KF1, relevance was higher than the baseline. Overall, averaging the three metrics, GCN+RL outperformed BART 10% and was close to BART 100%'s performance. However, all models were outperformed by human responses, possibly due to the size of the models or the number of training iterations.

In the knowledge-grounded setting, the authors observed that the GCN+RL approach produced more engaging and fluent conversations and outperformed both baselines while still being close to models trained on all available data. In pairwise comparisons, the GCN+RL approach was generally preferred over the other models. Despite using only 6% of the TC dataset's size, the generated data was of high quality, with BlenderBot-small performing comparably using the generated data as it did with 100% of human-human data and the original data itself. The authors note that the GCN method achieved this performance with small models, which have around 100M parameters each. Table 3 provides additional details on the results.

Table 3. Human evaluation results (top) for knowledge grounded conversations. Human evaluators rate responses with the same conversation context on a scale of 1 to 5. In a different evaluation (bottom), they were asked to choose the best response from two options. BBs: BlenderBot-small (90M), G-RL: GCN without RL, G+RL: GCN with RL.

Model	**Eng.**	**Flu.**	**Rel.**	**Avg.**
Data	3.74	3.98	3.57	3.76
BBs (100%)	3.69	3.99	3.57	3.75
BBs (1%)	3.64	3.86	3.42	3.64
G-RL generator	3.47	3.35	3.23	3.35
G-RL learner	3.58	3.85	**3.48**	3.64
G+RL generator	3.37	3.27	3.40	3.35
G+RL learner	**3.73**	**3.97**	**3.48**	**3.73**
		Wins Percentage		
Combinations	**Base**	**G-RL**	**G+RL**	**Tie**
BBs VS G-RL	40.0	**44.3**	-	15.7
BBs VS G+RL	44.7	-	**47.7**	7.6
All 3 models	29.3	25.7	**45.0**	-

Table 4. The performance of GCN+RL for different numbers of meta-iterations while generating 3 times the seed data and using only 1% of Topical Chat.

Iterations	**PPL**	**KF1**	**BL-4**
1	30.8	0.146	0.179
2	31.1	0.147	0.182
3	30.7	0.146	0.186
5	30.8	0.163	0.190
10	27.1	0.238	0.085

Table 5. The performance of GCN+RL for different sizes of generated data, expressed as a multiplier of the seed data. The experiment was conducted with 5 meta-iterations and 1% of TC data.

Iterations	**PPL**	**KF1**	**BL-4**
1	26.5	0.201	0.082
2	27.4	0.213	0.084
3	28.6	0.17	0.083
5	22.2	0.25	0.154
10	22.9	0.27	0.106

Table 6. Out-Of-Vocabulary (OOV) rates for various seed percentages.

Data%	**BBs**	**GCN-RL**	**GCN+RL**
1%	8.1%	17.4%	25.1%
5%	8.5%	12.1%	24.5%
10%	5.9%	9.2%	13.6%

Analysis: GCN Iterations I further analyze the performance of GCN especially its performance concerning the number of generated update iterations so I see that perplexity decreases and the knowledge F1 increase as I have more iterations meaning that the generator leads to leads the learner to learn to produce more fluent and at the same time knowledgeable response so but BULE scores naturally drops as this more knowledge response may not appear in the human response in the human reference.

Analysis: Synthetic The analysis focuses on the performance of GCN, particularly its performance concerning the number of meta-iterations (Table 4) and the amount of generated data (Table 5). Table 4 illustrates that KF1 improves as the meta-iterations increase, indicating that the generator guides the learner to produce more knowledgeable responses. However, BLEU decreases as these more knowledgeable responses may not be present in the data. Similar trends are observed in Table 5, where the quantity of synthetic generated data is varied (as a multiplier of the size of the seed data). Table 6 provides out-of-vocabulary rates for all three conditions when utilizing 1%, 5%, and 10% of the data as seed, revealing that higher rates imply more diversity but may also indicate that the generated data deviates further from the seed data. However, considering the results presented in Tables 1-3, GCN+RL generates more diverse data that are still relevant and useful.

Table 7. Example responses for open-domain conversations. The models here are not trained to optimize knowledgable responses and therefore may hallucinate. This may be the reason for lower relevance scores for GCN+RL (Table 1); however, GCN+RL's tendency to provide fun facts is what may be leading to more engaging responses.

Context
Hello, do you watch football?
Hi, yes. I love football! I did not realize the season lasted 17 weeks!
Yeah pretty long. I guess with 32 teams playing it takes that much time
Did you see they only rest one weekend during the season?
That is crazy, so much work goes in there. Did you know U of Iowa painted their visitor locker room pink?
Oh yes, that is so funny. they really want to intimidate the other team, but is more like a joke
Wow, what an idea. Did you know bowlers used to make more than football players in the 60s

	Responses
Baseline	that's a good point. I wonder how many of the top bowlers make twice as much as top football stars. do you have a favorite
GCN-RL	that's a good point. I'm not a big fan of football, but I do like to watch it. do you like football?
GCN+RL	I didn't know that, that's interesting. did you know there are more than 30 million players in the NFL? I wonder how many of those players are still playing?

Responses for conversations without knowledge grounding and knowledge-grounded conversations are provided in Tables 7 and 8, respectively. The tables show that GCN+RL attempts to include knowledge and fun facts in both cases and that the human judges receive these responses positively. However, in conversations without knowledge grounding, the model tends to hallucinate, resulting in lower relevance ratings. On the other hand, in knowledge-grounded conversations, GCN+RL uses the knowledge more effectively than the other models.

Table 8. Example responses for knowledge-grounded conversations. In this setting, models are trained to optimize knowledgable responses which explain in part why GCN+RL is preferred in human judgments (Table 3).

Context
Did you know that the University of Iowa's locker room is painted pink? I wonder why?
I think I did hear something about that. I imagine it is an attempt to psych the other team out.
So, it would be in the visiting team's locker room but not their own?

Retrieved Knowledge (subset)
the University of Iowa's locker room for visiting football teams is completely painted pink in the 1960's top bowlers made twice as much as top football stars
Football is a family of team sports that involve, to varying degrees, kicking a ball [...]
in 39 out of 50 states, the highest-paid state employee is either a college football or basketball coach.

	Responses
Baseline	I think it would be in the visiting team's locker room. I don't know why they would want to psych the other team out
GCN-RL	GCN-RL that's interesting. I've never heard that before. do you have a favorite team?
GCN+RL	I'm not too familiar with iowa's locker room for visiting football teams, but I do know that they're the highest-paid employees in the state of iowa. do you know why they are paid so much?

3 Conclusion

In this research, the authors proposed a method that incorporates external knowledge into a dialogue generation system using a reinforcement learning framework. They compared their proposed method, the GCN with reinforcement learning, to two baseline models, Blender and GPT, for both open-domain and knowledge-grounded conversational tasks. The automatic evaluations showed that the GCN method outperformed the baselines in terms of knowledge incorporation and fluency of generated responses, as indicated by higher knowledge F1 scores and lower perplexities.

The analysis of the GCN performance revealed that as the number of generated update iterations increases, the generated responses become more fluent and knowledgeable, but less similar to human reference responses. The authors also found that as the amount of synthetic data increases, the quality of the generated responses improves.

3.1 Evaluation and Assessment:

Overall, the GCN-based reinforcement learning approach appears to be a promising method for generating knowledge-grounded dialogues. While the potential positive impacts of this technology are significant, negative impacts such as the potential for perpetuating discrimination and the displacement of human workers must also be considered. Therefore, further research is needed to improve the relevance of the generated responses and incorporate additional knowledge sources to generate more informative dialogues. It is important to ensure that the development of conversational AI systems considers their potential impacts on society and prioritizes the ethical and responsible use of these technologies.

4 Future Aspects

In the future, there are several directions that could be explored to improve upon the proposed GCN approach. Some possibilities include:

Incorporating additional information sources: The current approach only uses knowledge from the provided context and reference set. Incorporating additional information sources, such as external knowledge bases or pre-trained models, could further improve the quality of the generated responses.

Improving the reward function: The current reward function only considers fluency and relevance as the criteria for evaluating the generated responses. In the future, additional metrics, such as consistency and coherence, could be considered to further improve the quality of the generated responses.

Integrating more advanced reinforcement learning algorithms: The current approach uses a simple reinforcement learning algorithm to update the generator. More advanced reinforcement learning algorithms, such as actor-critic methods or value-based methods, could be explored to improve the performance of the generator.

Evaluating the approach in real-world applications: The proposed GCN approach has been evaluated in a controlled setting. More exploration of the proposed approach in practical settings, such as in the development of customer service chatbots or personal assistants, can provide useful insights into its potential applications and potential challenges.

These future aspects will likely require further experimentation and development, but they have the potential to significantly improve the quality and capabilities of GCN for knowledge-grounded conversation generation.

References

1. Rashkin, H., Smith, E. M., Li, M., & Boureau, Y. L. (2019). Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP).*
2. Wang, W., Shi, P., Liu, Y., Su, J., & Zhou, X. (2019). Can you convince me? manipulating users' rating behavior with rating elicitation bots. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.*
3. Papangelis, A., Kouroupetroglou, G., & Papadopoulos, S. (2021). Generative Conversational Networks: A Meta-Learning Approach to Response Generation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume.*
4. Papangelis, A., Kouroupetroglou, G., & Papadopoulos, S. (2022). Knowledge Grounded Conversational Data Generation. In *Proceedings of the 13th Language Resources and Evaluation Conference (LREC).*
5. Papangelis, A., Ghosh, S., Moon, S., Foster, I., & Muresan, S. (2021). An evaluation of reference-less quality estimation metrics for open-domain dialogue generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)(pp. 7898-7907).*
6. Papineni, K., Roukos, S., Ward, T., & Zhu, W. J. (2002). BLEU: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL) (pp. 311-318). https://www.aclweb.org/anthology/P02-1040.pdf.*
7. Lin, C. Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Text summarization branches out: Proceedings of the ACL-04 workshop.*

8. Henderson, M., Desai, R., & Pineau, J. (2019). *Tracking state changes in dynamic environments with recurrent entity networks. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 4264-4274).*
9. Shuster, K., Mahajan, D., Nogueira, R., & Cho, K. (2021). Knowledge Fusions for Diverse Commonsense Question Answering. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.*
10. Chen, W., & Cherry, C. (2014). A systematic comparison of smoothing techniques for sentence-level BLEU. *In Proceedings of the Ninth Workshop on Statistical Machine Translation (WMT) (pp. 362-367). https://www.aclweb.org/anthology/W14-3346.pdf*
11. Dinan, E., Urbanek, J., Szekely, A., Kiela, D., & Weston, J. (2019). The second conversational intelligence challenge (ConvAI2). *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) (pp. 4705-4718). https://www.aclweb.org/anthology/D19-1483.pdf*