GENERATING A NARRATIVE WITH EXCELLENT STORY TELLING CAPABILITIES BY GIVING PROMPTS TO PRE-TRAINED MODELS

MOHAMMED AAMER
(STUDENT ID : PN1130101)

Research Proposal

MAY 2024

Abstract

With advent of advanced language models (PLM), significant advances have been made in text design problems. However, these models can only reflect a certain width of the output. PLM cannot create long stories because it does not understand narrative structure. Recent research on story generation has used the concept of clarity, which can create stories with multiple reasons for events. However, the scarcity of training data makes the optimization of PLM difficult. Even with appropriate treatment, definitive control is difficult to achieve. Therefore, it is not easy to create models that can create long stories. A recent study answers this question. This article shows a way to use rapid learning to create stories while maintaining quality control.

Table of Contents

Table of Contents 3
1. Introduction 4
2. Exisiting Research 5
3. Questions 6
4. Objectives 6
5. Significance 6
6. Scope of Study 7
7. Research Methodology 7
7.1 Dataset 7
7.2 Data Preparation 7
7.3 Algorithms 8
7.3.1 Pre-trained Language Models (PLMs) 8
7.3.2 Few-Shot Learning (FSL) 8
7.3.3 Prompt-Learning 9
7.4 Implementation 9
7.5 Assessment 10
8. Requirements 10
8.1 Software 10
8.2 Hardware 10
8.3 Dataset 11
9. Plan 11
9.1 Gantt Chart 11
9.2 Risk & Exigency 12
References 13

1. Introduction

The quality of machine authoring has increased in recent decades with the invention of large-scale pretraining models. Today they can produce texts that are very similar to human handwriting.

However, while large-scale PLM has demonstrated great potential in creating consistent and understandable documents, managing the creation remains a challenge (Keskar et al., 2019; Radford et al., 2019). A more comprehensive review of the published literature revealed issues such as conflicting topics and conflicting personalities (Fan et al., 2019; Bisk et al). These deficits are especially important for open reading activities that require collaboration, such as narrative construction.

Narratives created using these models have been shown to have absence of dialogue and. Cognitive skills. Although single sentences in the text may seem meaningful and flowing, when put together the whole narrative is not very cohesive (see et al., 2019; Goldfarb-Tarrant et al., 2020). Repetition of sentences in long texts will reduce the quality of the story (Yao et al., 2019). Preparation of content in different formats. Use the instructions. Use keywords and keywords using semantic frames. To get the most out of these projects, PLM often needs to fine-tune the data. One of the challenges in optimizing PLM is that in addition to the required information, the model also tends to examine events arising from the planned content and provide useful context from them (Fan et al., 2019). This causes a lack of diversity in the narratives created. In this context, special warnings are used to solve text problems. Study indicates that using guidelines, PLM can resolve current or next-generation tasks eliminating the need for optimization. But still some problems exist. Cues are specific to a task and difficult to adapt or reuse to new tasks. Even within the task, instructions may not be applicable to all situations in large groups.

2. Existing Research

The stories of the affected generation are examined from various perspectives. Researchers use ideas, genre, style, content, etc. to control story formation. They focused on the use of general concepts. Some studies try to use stories, narratives and plans for detailed management (Peng et al., 2018). This study was evaluated using the Five Element Stories dataset ROC Stories (Mostafazadeh et al., 2016). Later, some studies have experimented using long texts to control story formation (Fang et al., 2021).

Researchers also attempted to use QC control to guide the novella. (Fang et al., 2021) suggested creating stories based on the content of the story/sentence. (Rashkin et al., 2020) proposed a similar system with a special system and memory. (Sun et al., 2020) Create a storytelling by creating a summary for each part of the story. Each element is then expanded to form a complete story.

Most research in this area is on the basis of fine-tuning selected or constructed TX-based pre-learned language models (PLMs). In particular, GPT-2 has received great recognition in this field due to its unique structure for illicit texts. Recently, the use of GPT3 (Brown et al., 2020) for text has increased following the release of its API (Dou et al., 2021; Shakeri et al., 2021).

It is difficult to fine-tune PLM when there is not enough data. To solve this problem, researchers tried the plug-and-play method to control the creation of stories without proper editing (2022 Sen et al., 2022).

Another learning method is hint based that does not require proper correction. Some projects use multigenerational teaching (Brown et al., 2020). Others tried to obtain discrete instructions and continuous instructions. Some even attempted to use the workplace to generate cues for goal work (Vu et al., 2021).

3. Questions

In this research, we will try address the below points:

1. Story creation process along with quality measures needs to be optimized in PLM. Can this technique be used with rapid learning to create stories with fewer shots without requiring fine-tuning?

2. We have been using Prompt-based learning in many ways to generate text. Is it possible to extende this to story-making?

3. The previous system only used GPT2 as the base model. Can text be loaded using a new version (or newer) of GPT3?

4. Objectives

We try to apply the capabilities of GPT3 for long story generation which can be controllable.

Objectives:

- Perform a thorough review of available literature for long narrative generation and shot generation techniques.

- Attempt to develop a method for generation of long and short stories using few prompts

- To assess the generated narratives with the help of existing metrics and compare with them.

5. Significance

Story creation is an active area of research. Although the short story has been studied extensively, its longer history has been less researched. Although optimization has been used as a method in previous studies, there is no research on creating stories without optimization. This study also explores recent developments and some inflections in cognitive-based learning. Writers can benefit from this by getting assistance for new ideas or overcome the condition known as writer's block.

6. Scope of Study

Below is our scope for this study:

- Completion time is 18 weeks after submission of the proposal of research.
- Open-source tool and data will be used for this project
- Publicly available compute sources will be used for the project.

7. Research Methodology

The project focuses on generation of text and the capability to learn of PLM.

7.1 Dataset

We make the use of below datasets:

- **ROC Stories**: This document was introduced by (Mostafazadeh et al., 2016) and is approximately 98,000 5 sentences long, including story titles. This information is widely used in short-form generation projects.
- **Writing Prompts**: This collection was introduced by (Fan et al., 2018b) and is an introduction to story writing alongside stories written by approximately 300,000 groups. The story was captured from reddit. These are long, multi-sentence stories, therefore useful for complex tasks such as generation of long form stories.

7.2 Data Preparation

Plan should be an example of a sentence. While sending articles and sharing written information via ROC Stories, no information is prepared. For this reason, the process needs to be extracted from the story data set and then displayed on the link. This pair of descriptive sentences can be used in many conversations.

One of two forms can be taken by outline instances:

- **Summary:** An example of the structure here is the summary of the sentence. This article has been expanded from the abstract. For content extraction, it is recommended to use

TextRank (Mihalcea and Tarau, 2004) for extracting the words that give the most information from the sentence.

- **Key phrases / Keywords:** Examples of structure here are the main subject and the expressions in the sentence. Articles consist of these keywords/keywords. For the extraction process, it is recommended to use RAKE (Rose et al., 2010) to remove important words from sentence.

7.3 Algorithms

7.3.1 Pre-trained Language Models (PLMs)

The method of transfer learning is use by PLMs. It’s the method of utilizing exiting knowldge to new jobs. It always uses a lot of written content for supervision training.

Pretraining for self-tracking of multiple unlabeled objects has become the most popular learning transformation in deep learning. Pretraining differs from other methods in that it uses anonymous data for traiing itself and is useful for several tasks, from optimization to microlearning. These models are trained to self-monitor using a variety of inappropriate texts.

These models can be used in many natural language applications such as answering questions, creating text, classifying text. Beautiful designs. Some of the most popular language patterns are:

- GPT3 (Brown et al., 2020)
- BERT (Devlin al., 2018)

7.3.2 Few-Shot Learning (FSL)

Using a small sample size and experience, people can easily identify new categories in the data. This is also known as mets learning. Learning in Few Steps is form of meta-learning. This way, learners are taught many related tasks during the meta-study period, with the aim of making new (but related) tasks useful with some small examples during the meta-analysis. A good way to solve the problem of learning a pair is to learn for several tasks and then train the specific tasks. It’s a method that always monitors learning process, which requires a lot of recording information for the purpose of training.

7.3.3 Prompt-Learning

English Language Learning is a new course for teaching learning styles to ML models. When prompted, users say directly in the language of the task that they need previous language patterns to understand and complete. Whereas, Transformer always first performs the representation utilizing unsigned data and then uses the registered data to tune the model to use the necessary low-level operations. Warnings are user-written instructions that the model must follow. More inspiration may be needed depending on the complexity of the task being trained. Request engineering is the process of selecting appropriate requests or requests for required tasks. Sign-based learning has many advantages over pre-training and traditional methods. The main conclusion is that most of the lessons are done well with a small sample of labels.

7.4 Implementation

What we have implemented can be categorized into 2 steps:

1. Creating prompts for Few-Shot Learning – In this step, information is generated about a small set of sample pairs. Each example has a outline (o) and a paragraph (t).

Use examples of verbs with several conjugations to create incomplete sentences for new verbs; Verbs, examples with few conjugations, and sample questions are sent to the sample channel as emotional input. This formula returns the sentence generated by the given query.

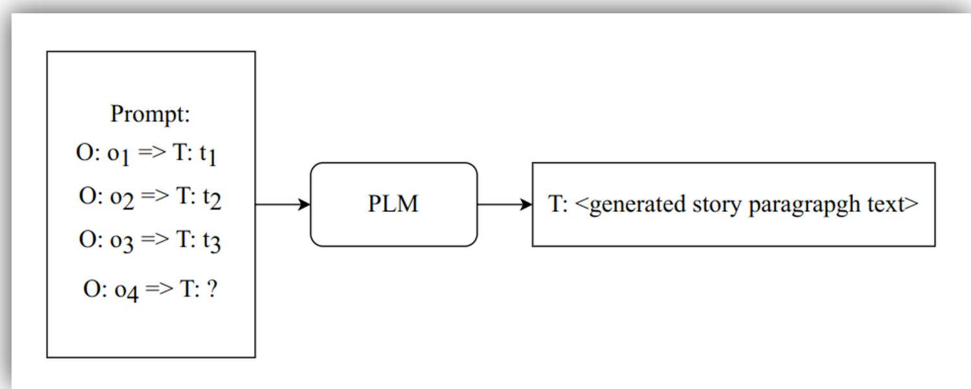


Fig. 7.4.1

7.5 Assessment

Resulting stories will be assessed with various measurements. The study focuses solely on assessement using indicators which are automated. Book review of resulting stories is beyond this study.

The evaluation metrics are as follows:

- **ROUGE** (Lin, 2004) – Interpretation is similar to score of BLEU.
- **Perplexity (PPL)** (Fang et al., 2021; Jin et al., 2022) – Its used for computing complexity at word level.
- **Self-BLEU** (Zhu et al., 2018) – Used for measuring lexical diversity (intra-story)
- **BLEU** (Papineni et al., 2002) – Used for measuring n-gram overlap b/w the ground truth and generated texts.
- **DIST/distinct-n** (Li et al., 2015) - It quanitifis diversity in generation as a ratio of all generated n-grams to distinct n-grams.

Benchmarking is done against the following base:

- Outline-2-Story (**O2S**) (Fang et, 2021)
- Prompt Transfer for Text Generation (**PTG**) (Li et al., 2022)
- Summarize and Outlin and Elaborate (**SOE**) (Sun et al., 2020)

8. Requirements

8.1 Software

Below is the list of software that we will be utilizing for this work:

- Internet Explorer (Preferable Chrome)
- Integrated Development Environment (Preferable PyCharm)

- Python for coding
- NVIDIA - CUDA libraries
- DL Lib (TensorFlow, PyTorch, HuggingFace)

8.2 Hardware

We will be using the following hardware:

- PC with internet access and coding application
- GPUs to train the models

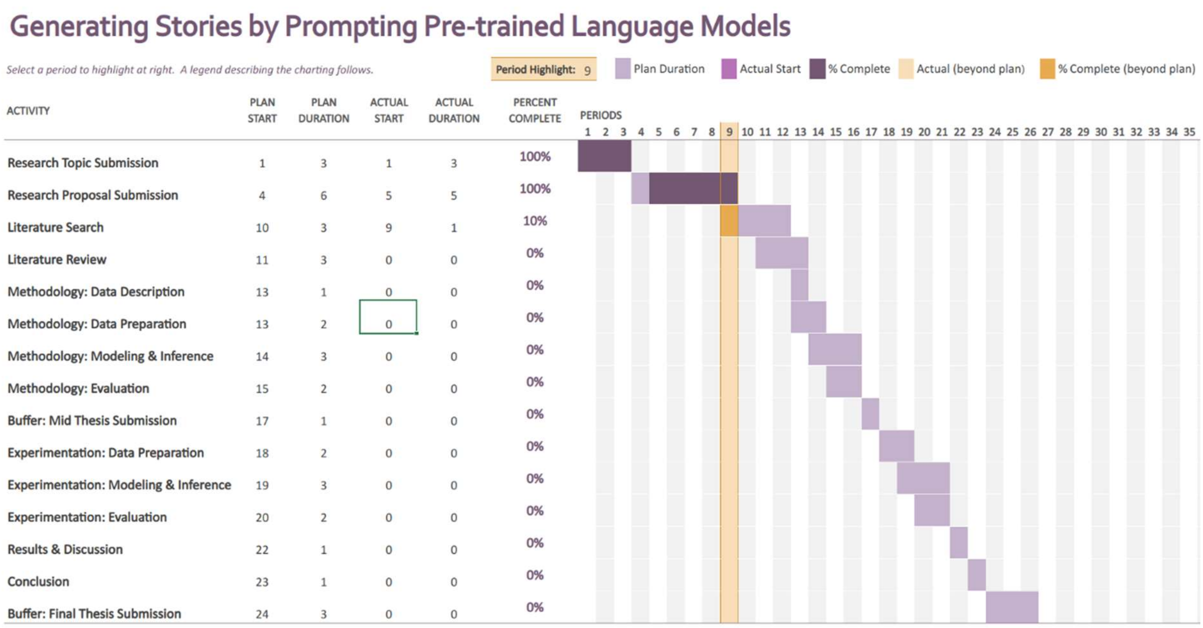
8.3 Dataset

We will be using the following datasets:

- ROC Stories dataset

9. Plan

9.1 Gantt Chart



ACTIVITY	PLAN START	PLAN DURATION	ACTUAL START	ACTUAL DURATION	PERCENT COMPLETE
Research Topic Submission	1	3	1	3	100%
Research Proposal Submission	4	6	5	5	100%
Literature Search	10	3	9	1	10%
Literature Review	11	3	0	0	0%
Methodology: Data Description	13	1	0	0	0%
Methodology: Data Preparation	13	2	0	0	0%
Methodology: Modeling & Inference	14	3	0	0	0%
Methodology: Evaluation	15	2	0	0	0%
Buffer: Mid Thesis Submission	17	1	0	0	0%
Experimentation: Data Preparation	18	2	0	0	0%
Experimentation: Modeling & Inference	19	3	0	0	0%
Experimentation: Evaluation	20	2	0	0	0%
Results & Discussion	22	1	0	0	0%
Conclusion	23	1	0	0	0%
Buffer: Final Thesis Submission	24	3	0	0	0%

Fig. 9.1.1 (1 Period = 7 days)

9.2 Risk & Exigency

Below is a list of risks and related situations in completing a thesis project.:

Table 9.2.1

Risk	Exigency
Candidates cannot conduct research studies due to health problems or personal problems that may affect the duration.	Add buffer time & contact university for any additional support
Non-availability of local compute	Use cloud graphical processing units

References

Anon (2022) *ROCStories and the Story Cloze Test*. [online] Available at: https://cs.rochester.edu/nlp/rocstories/ [Accessed 26 Oct. 2022].

Bakhtin, A., Deng, Y., Ott, M., Ranzato, M.' A. and Szlam, A., (2021) Residual Energy-Based Models for Text. *Journal of Machine Learning Research*, [online] 22, pp.1–41. Available at: http://jmlr.org/papers/v22/20-326.html. [Accessed 23 Oct. 2022].

Bisk, Y., Holtzman, A., Thomason, J., Andreas, J., Bengio, Y., Chai, J., Lapata, M., Lazaridou, A., May, J., Nisnevich, A., Pinto, N. and Turian, J., (2020) Experience Grounds Language. *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, [online] pp.8718–8735. Available at: https://arxiv.org/abs/2004.10151v3 [Accessed 23 Oct. 2022].

Bosselut, A., Celikyilmaz, A., He, X., Gao, J., Huang, P. sen and Choi, Y., (2018) Discourse-Aware Neural Rewards for Coherent Text Generation. *NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, [online] 1, pp.173–184. Available at: https://arxiv.org/abs/1805.03766v1 [Accessed 23 Oct. 2022].

Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., Mccandlish, S., Radford, A., Sutskever, I. and Amodei, D., (2020) Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, [online] 33, pp.1877–1901. Available at: https://commoncrawl.org/the-data/ [Accessed 23 Oct. 2022].

Chen, Z., Eavani, H., Chen, W., Liu, Y. and Wang, W.Y., (2019) Few-Shot NLG with Pre-Trained Language Model. [online] pp.183–190. Available at: https://arxiv.org/abs/1904.09521v3 [Accessed 25 Oct. 2022].

Clark, E., August, T., Serrano, S., Haduong, N., Gururangan, S. and Smith, N.A., (2021) All That's 'Human' Is Not Gold: Evaluating Human Evaluation of Generated Text. *ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, [online] pp.7282–7296. Available at: https://arxiv.org/abs/2107.00061v2 [Accessed 23 Oct. 2022].

Conneau, A. and Lample, G., (2019) Cross-lingual Language Model Pretraining. *Advances in Neural Information Processing Systems*, [online] 32. Available at: https://arxiv.org/abs/1901.07291v1 [Accessed 25 Oct. 2022].

Dathathri, S., Madotto, A., Lan, J., Hung, J., Ai, U., Frank, E., Molino, P., Yosinski, J. and Liu, R., (2019) Plug and Play Language Models: A Simple Approach to Controlled Text Generation. [online] Available at: https://arxiv.org/abs/1912.02164v4 [Accessed 25 Oct. 2022].

Devlin, J., Chang, M.W., Lee, K. and Toutanova, K., (2018) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human*

Language Technologies - Proceedings of the Conference, [online] 1, pp.4171–4186. Available at: https://arxiv.org/abs/1810.04805v2 [Accessed 23 Oct. 2022].

Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., Gao, J., Zhou, M. and Hon, H.-W., (2019) Unified Language Model Pre-training for Natural Language Understanding and Generation. *Advances in Neural Information Processing Systems*, [online] 32. Available at: https://github.com/microsoft/unilm. [Accessed 25 Oct. 2022].

Dou, Y., Forbes, M., Koncel-Kedziorski, R., Smith, N.A. and Choi, Y., (2021) Is GPT-3 Text Indistinguishable from Human Text? Scarecrow: A Framework for Scrutinizing Machine Text. [online] pp.7250–7274. Available at: https://arxiv.org/abs/2107.01294v3 [Accessed 23 Oct. 2022].

Dziri, N., Madotto, A., Zaiane, O. and Bose, A.J., (2021) Neural Path Hunter: Reducing Hallucination in Dialogue Systems via Path Grounding. *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, [online] pp.2197–2214. Available at: https://arxiv.org/abs/2104.08455v2 [Accessed 23 Oct. 2022].

Fan, A., Lewis, M. and Dauphin, Y., (2018a) Hierarchical Neural Story Generation. *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, [online] 1, pp.889–898. Available at: https://arxiv.org/abs/1805.04833v1 [Accessed 23 Oct. 2022].

Fan, A., Lewis, M. and Dauphin, Y., (2018b) Hierarchical Neural Story Generation. *ACL 2018 - 56th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, [online] 1, pp.889–898. Available at: https://aclanthology.org/P18-1082 [Accessed 13 Sep. 2022].

Fan, A., Lewis, M. and Dauphin, Y., (2019) Strategies for Structuring Story Generation. *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, [online] pp.2650–2660. Available at: https://arxiv.org/abs/1902.01109v2 [Accessed 23 Oct. 2022].

Goldfarb-Tarrant, S., Chakrabarty, T., Weischedel, R. and Peng, N., (2020) Content Planning for Neural Story Generation with Aristotelian Rescoring. *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, [online] pp.4319–4338. Available at: https://arxiv.org/abs/2009.09870v2 [Accessed 23 Oct. 2022].

Guan, J., Huang, F., Zhao, Z., Zhu, X. and Huang, M., (2020) A knowledge-enhanced pretraining model for commonsense story generation. *Transactions of the Association for Computational Linguistics*, [online] 8, pp.93–108. Available at: https://direct.mit.edu/tacl/article/doi/10.1162/tacl_a_00302/43540/A-Knowledge-Enhanced-Pretraining-Model-for [Accessed 23 Oct. 2022].

Hu, Z., Yang, Z., Liang, X., Salakhutdinov, R. and Xing, E.P., (2017) *Toward Controlled Generation of Text*. Available at: https://proceedings.mlr.press/v70/hu17e.html [Accessed 25 Oct. 2022].

Hua, X. and Wang, L., (2020) PAIR: Planning and Iterative Refinement in Pre-trained Transformers for Long Text Generation. *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, [online] pp.781–793. Available at: https://arxiv.org/abs/2010.02301v1 [Accessed 23 Oct. 2022].

Ji, H. and Huang, M., (2021) DiscoDVT: Generating Long Text with Discourse-Aware Discrete Variational Transformer. *EMNLP 2021 - 2021 Conference on Empirical Methods in Natural Language Processing, Proceedings*, [online] pp.4208–4224. Available at: https://arxiv.org/abs/2110.05999v1 [Accessed 23 Oct. 2022].

Ji, H., Ke, P., Huang, S., Wei, F., Zhu, X. and Huang, M., (2020) Language Generation with Multi-Hop Reasoning on Commonsense Knowledge Graph. *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, [online] pp.725–736. Available at: https://arxiv.org/abs/2009.11692v1 [Accessed 23 Oct. 2022].

Jin, Y., Kadam, V. and Wanvarie, D., (2022) Plot Writing From Pre-Trained Language Models. *INLG 2022*. [online] Available at: https://arxiv.org/abs/2206.03021 [Accessed 24 Aug. 2022].

Keskar, N.S., Mccann, B., Varshney, L.R., Xiong, C., Socher, R. and Research, S., (2019) CTRL: A Conditional Transformer Language Model for Controllable Generation. [online] Available at: https://arxiv.org/abs/1909.05858v2 [Accessed 23 Oct. 2022].

Li, J., Galley, M., Brockett, C., Gao, J. and Dolan, B., (2015) A Diversity-Promoting Objective Function for Neural Conversation Models. *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference*, [online] pp.110–119. Available at: https://arxiv.org/abs/1510.03055v3 [Accessed 29 Oct. 2022].

Li, J., Tang, T., Nie, J.-Y., Wen, J.-R. and Zhao, W.X., (2022) Learning to Transfer Prompts for Text Generation. *NAACL 2022*, [online] pp.3506–3518. Available at: https://arxiv.org/abs/2205.01543v2 [Accessed 11 Oct. 2022].

Li, J., Tang, T., Zhao, W.X., Wei, Z., Yuan, N.J. and Wen, J.R., (2021) Few-shot Knowledge Graph-to-Text Generation with Pretrained Language Models. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, [online] pp.1558–1568. Available at: https://arxiv.org/abs/2106.01623v1 [Accessed 25 Oct. 2022].

Li, X.L. and Liang, P., (2021) Prefix-Tuning: Optimizing Continuous Prompts for Generation. *ACL-IJCNLP 2021 - 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, [online] pp.4582–4597. Available at: https://arxiv.org/abs/2101.00190v1 [Accessed 25 Oct. 2022].

Lin, C.-Y., (2004) *ROUGE: A Package for Automatic Evaluation of Summaries*. Available at: https://aclanthology.org/W04-1013 [Accessed 29 Oct. 2022].

Lin, Z. and Riedl, M.O., (2021) Plug-and-Blend: A Framework for Controllable Story Generation with Blended Control Codes. *NAACL | NUSE | 2021*, [online] pp.62–71. Available at: https://aclanthology.org/2021.nuse-1.7 [Accessed 31 Aug. 2022].

Liu, X., Zheng, Y., Du, Z., Ding, M., Qian, Y., Yang, Z. and Tang, J., (2021) GPT Understands, Too. [online] Available at: https://arxiv.org/abs/2103.10385v1 [Accessed 25 Oct. 2022].

Mao, H.H., Majumder, B.P., McAuley, J. and Cottrell, G.W., (2019) Improving Neural Story Generation by Targeted Common Sense Grounding. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference* *on Natural Language Processing, Proceedings of the Conference*, [online] pp.5988–5993. Available at: https://aclanthology.org/D19-1615 [Accessed 31 Aug. 2022].

Mihalcea, R. and Tarau, P., (2004) TextRank: Bringing Order into Texts.

Mori, Y., Yamane, H., Shimizu, R. and Harada, T., (2022) Plug-and-Play Controller for Story Completion: A Pilot Study toward Emotion-aware Story Writing Assistance. [online] pp.46–57. Available at: https://aclanthology.org/2022.in2writing-1.6 [Accessed 1 Oct. 2022].

Mostafazadeh, N., Chambers, N., He, X., Parikh, D., Batra, D., Vanderwende, L., Kohli, P. and Allen, J., (2016) A Corpus and Cloze Evaluation for Deeper Understanding of Commonsense Stories. *2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL HLT 2016 - Proceedings of the Conference*, [online] pp.839–849. Available at: https://aclanthology.org/N16-1098 [Accessed 13 Sep. 2022].

Papineni, K., Roukos, S., Ward, T. and Zhu, W.-J., (2002) Bleu: a Method for Automatic Evaluation of Machine Translation. *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, [online] pp.311–318. Available at: https://aclanthology.org/P02-1040 [Accessed 29 Oct. 2022].

Pascual, D., Egressy, B., Bolli, F. and Wattenhofer, R., (2020) Directed Beam Search: Plug-and-Play Lexically Constrained Language Generation. [online] Available at: http://arxiv.org/abs/2012.15416 [Accessed 31 Aug. 2022].

Pascual, D., Egressy, B., Meister, C., Cotterell, R. and Wattenhofer, R., (2021) A Plug-and-Play Method for Controlled Text Generation. *Findings of the Association for Computational Linguistics, Findings of ACL: EMNLP 2021*, [online] pp.3973–3997. Available at: https://aclanthology.org/2021.findings-emnlp.334 [Accessed 31 Aug. 2022].

Peng, N., Ghazvininejad, M., May, J. and Knight, K., (2018) Towards Controllable Story Generation. [online] pp.43–49. Available at: https://aclanthology.org/W18-1505 [Accessed 25 Oct. 2022].

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. and Sutskever, I., (2019) Language Models are Unsupervised Multitask Learners. [online] Available at: https://github.com/codelucas/newspaper [Accessed 23 Oct. 2022].

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W. and Liu, P.J., (2020) Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*, [online] 21, pp.1–67. Available at: http://jmlr.org/papers/v21/20-074.html. [Accessed 25 Oct. 2022].

Rashkin, H., Celikyilmaz, A., Choi, Y. and Gao, J., (2020) PlotMachines: Outline-Conditioned Generation with Dynamic Plot State Tracking. *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, [online] pp.4274–4295. Available at: https://aclanthology.org/2020.emnlp-main.349 [Accessed 31 Aug. 2022].

Rashkin, H., Smith, E.M., Li, M. and Boureau, Y.L., (2018) Towards Empathetic Open-domain Conversation Models: a New Benchmark and Dataset. *ACL 2019 - 57th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, [online] pp.5370–5381. Available at: https://arxiv.org/abs/1811.00207v5 [Accessed 23 Oct. 2022].

Rose, S., Engel, D., Cramer, N. and Cowley, W., (2010) Automatic Keyword Extraction from Individual Documents. *Text Mining: Applications and Theory*, [online] pp.1–20. Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/9780470689646.ch1 [Accessed 29 Oct. 2022].

le Scao, T. and Rush, A.M., (2021) How Many Data Points is a Prompt Worth? *NAACL-HLT 2021 - 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, [online] pp.2627–2636. Available at: https://arxiv.org/abs/2103.08493v2 [Accessed 25 Oct. 2022].

See, A., Pappu, A., Saxena, R., Yerukola, A. and Manning, C.D., (2019) Do Massively Pretrained Language Models Make Better Storytellers? *CoNLL 2019 - 23rd Conference on Computational Natural Language Learning, Proceedings of the Conference*, [online] pp.843–861. Available at: https://arxiv.org/abs/1909.10705v1 [Accessed 23 Oct. 2022].

Shakeri, H., Neustaedter, C. and DiPaola, S., (2021) SAGA: Collaborative Storytelling with GPT-3; SAGA: Collaborative Storytelling with GPT-3. *Companion Publication of the 2021 Conference on Computer Supported Cooperative Work and Social Computing*. [online] Available at: https://doi.org/10.1145/3462204.3481771 [Accessed 1 Oct. 2022].

Shen, T., Lei, T., Barzilay, R., Jaakkola, T. and Csail, M., (2017) Style Transfer from Non-Parallel Text by Cross-Alignment. *Advances in Neural Information Processing Systems*, [online] 30. Available at: https://github.com/shentianxiao/language-style-transfer. [Accessed 25 Oct. 2022].

Shin, T., Razeghi, Y., Logan, R.L., Wallace, E. and Singh, S., (2020) AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, [online] pp.4222–4235. Available at: https://arxiv.org/abs/2010.15980v2 [Accessed 25 Oct. 2022].

Song, K., Tan, X., Qin, T., Lu, J. and Liu, T.Y., (2019) MASS: Masked Sequence to Sequence Pre-training for Language Generation. *36th International Conference on Machine Learning, ICML 2019*, [online] 2019-June, pp.10384–10394. Available at: https://arxiv.org/abs/1905.02450v5 [Accessed 25 Oct. 2022].

Su, Y., Wang, X., Qin, Y., Chan, C.-M., Lin, Y., Wang, H., Wen, K., Liu, Z., Li, P., Li, J., Hou, L., Sun, M. and Zhou, J., (2021) On Transferability of Prompt Tuning for Natural Language Processing. *NAACL 2022 - 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, [online] pp.3949–3969. Available at: http://arxiv.org/abs/2111.06719 [Accessed 25 Oct. 2022].

Sun, X., Fan, C., Sun, Z., Meng, Y., Wu, F. and Li, J., (2020) Summarize, Outline, and Elaborate: Long-Text Generation via Hierarchical Supervision from Extractive Summaries. [online] Available at: https://arxiv.org/abs/2010.07074v2 [Accessed 23 Oct. 2022].

Tan, B., Yang, Z., Al-Shedivat, M., Xing, E.P. and Hu, Z., (2020) Progressive Generation of Long Text with Pretrained Language Models. *NAACL-HLT 2021 - 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, [online] pp.4313–4324. Available at: https://arxiv.org/abs/2006.15720v2 [Accessed 23 Oct. 2022].

Vaswani, A., Brain, G., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., (2017) Attention is All you Need. *Advances in Neural Information Processing Systems*, 30.

Vu, T., Lester, B., Constant, N., Al-Rfou, R., Cer, D. and Research, G., (2021) SPoT: Better Frozen Model Adaptation through Soft Prompt Transfer. [online] pp.5039–5059. Available at: https://arxiv.org/abs/2110.07904v2 [Accessed 25 Oct. 2022].

Xu, J., Ren, X., Zhang, Y., Zeng, Q., Cai, X. and Sun, X., (2018) A Skeleton-Based Model for Promoting Coherence Among Sentences in Narrative Story Generation. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, EMNLP 2018*, [online] pp.4306–4315. Available at: https://arxiv.org/abs/1808.06945v2 [Accessed 23 Oct. 2022].

Xu, P., Patwary, M., Shoeybi, M., Puri, R., Fung, P., Anandkumar, A. and Catanzaro, B., (2020) MEGATRON-CNTRL: Controllable Story Generation with External Knowledge Using Large-Scale Language Models. *EMNLP 2020 - 2020 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, [online] pp.2831–2845. Available at: https://arxiv.org/abs/2010.00840v1 [Accessed 23 Oct. 2022].

Yao, L., Peng, N., Weischedel, R., Knight, K., Zhao, D. and Yan, R., (2019) Plan-and-Write: Towards Better Automatic Storytelling. *Proceedings of the AAAI Conference on Artificial Intelligence*, [online] 3301, pp.7378–7385. Available at: https://ojs.aaai.org/index.php/AAAI/article/view/4726 [Accessed 23 Oct. 2022].

Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., Choi, Y. and Allen, P.G., (2019) Defending Against Neural Fake News. *Advances in Neural Information Processing Systems*, [online] 32. Available at: https://rowanzellers.com/grover [Accessed 23 Oct. 2022].

Zhang, Y., Sun, S., Galley, M., Chen, Y.-C., Brockett, C., Gao, X., Gao, J., Liu, J. and Dolan, B., (2019) DialoGPT: Large-Scale Generative Pre-training for Conversational Response Generation. [online] pp.270–278. Available at: https://arxiv.org/abs/1911.00536v3 [Accessed 23 Oct. 2022].

Zhao, J.J., Kim, Y., Zhang, K., Rush, A.M. and Lecun, Y., (2018) Adversarially Regularized Autoencoders. [online] pp.5902–5911. Available at: https://proceedings.mlr.press/v80/zhao18b.html [Accessed 25 Oct. 2022].

Zhu, Y., Lu, S., Zheng, L., Guo, J., Zhang, W., Wang, J. and Yu, Y., (2018) Texygen: A benchmarking platform for text generation models. *41st International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2018*, pp.1097–1100.

Ziegler, Z.M., Melas-Kyriazi, L., Gehrmann, S. and Rush, A.M., (2019) Encoder-Agnostic Adaptation for Conditional Language Generation. [online] Available at: https://arxiv.org/abs/1908.06938v2 [Accessed 25 Oct. 2022].

Zou, X., Yin, D., Zhong, Q., Yang, H., Yang, Z. and Tang, J., (2021) Controllable Generation from Pre-trained Language Models via Inverse Prompting. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp.2450–2460.