

Group 4 Proposal

Problem and Task Definition:

Storytelling is a fundamental aspect of how humans communicate. It allows us to share ideas, convey emotions, and develop culture. For my project, I plan to train a neural language model to perform the task of short story (text) generation in the horror/thriller genre.

Key Concepts Explored:

Although this may change, I am currently planning on using a pre-trained state of the art transformer model like GPT2/3/BERT and fine tuning that architecture on a custom corpus. I may also explore RNNs however, the literature seems to suggest state of the art transformer models will likely perform best for this task.

Dataset:

The custom corpus will be short stories from the following subreddits:

- r/nosleep
- r/scarystories
- r/supershortscarystories
- r/creepyencounters
- r/truescarystories
- r/Glitch_in_the_Matrix
- r/Paranormal
- r/Ghoststories

This is not a definitive list as I may add or remove sources once I start parsing the stories.

Key Packages:

- Pytorch/Tensorflow
- Transformers (huggingface)
- Nltk/spacy
- BeautifulSoup/selenium

Model Performance:

I will use a combination of BLEAU/ROGUE/METEOR. However, BERTScore may be better evaluators as well.

Tentative Schedule:

- 3/21 week: Research and planning & scrape data
- 3/28 week: Scrape data & clean
- 4/4 week: Feature extraction & load pretrained model
- 4/11 week: Fine tune pretrained model on dataset
- 4/18 week: Evaluation & further finetune
- 4/25 week: Finalize and prepare presentation/report

- 5/2 week: Final presentation