

Introduction

Motivation

LLMs have been criticized for lacking creativity in its text generation. We want to fine tune a model to generate engaging and relevant titles for short form fictional stories such as .

Literature survey

Previous related works in creative title generation acknowledge that most title generation is very summarization based. To counteract this, many use probabilistic methods to increase creativity.

Problem Definition

The goal is to create a natural language processing solution capable of generating titles for short-form content, specifically focusing on short horror stories written and posted on reddit. The model should be able to extract the core ideas from the given text as well as get a general idea of the theme of the content in order to create a title that isn't just summarizative, but adds a creative twist to it.

Novel contribution:

Our work will be focusing on enhancing the creativity of titles, where "creativity" is measured by these attributes:

- 1. Is it novel or new?
- 2. Is it surprising?
- 3. Is it valuable?

(i.e. Useful in classifying the stories, giving a good idea of what the story is, and makes you want to read the story?)

Ideally, these generated titles will prove to be equally or more interesting than the actual human titles.

Methods

Fine Tuned HuggingFace GPT2 model on reddit horror stories dataset with data-points structured as follows:

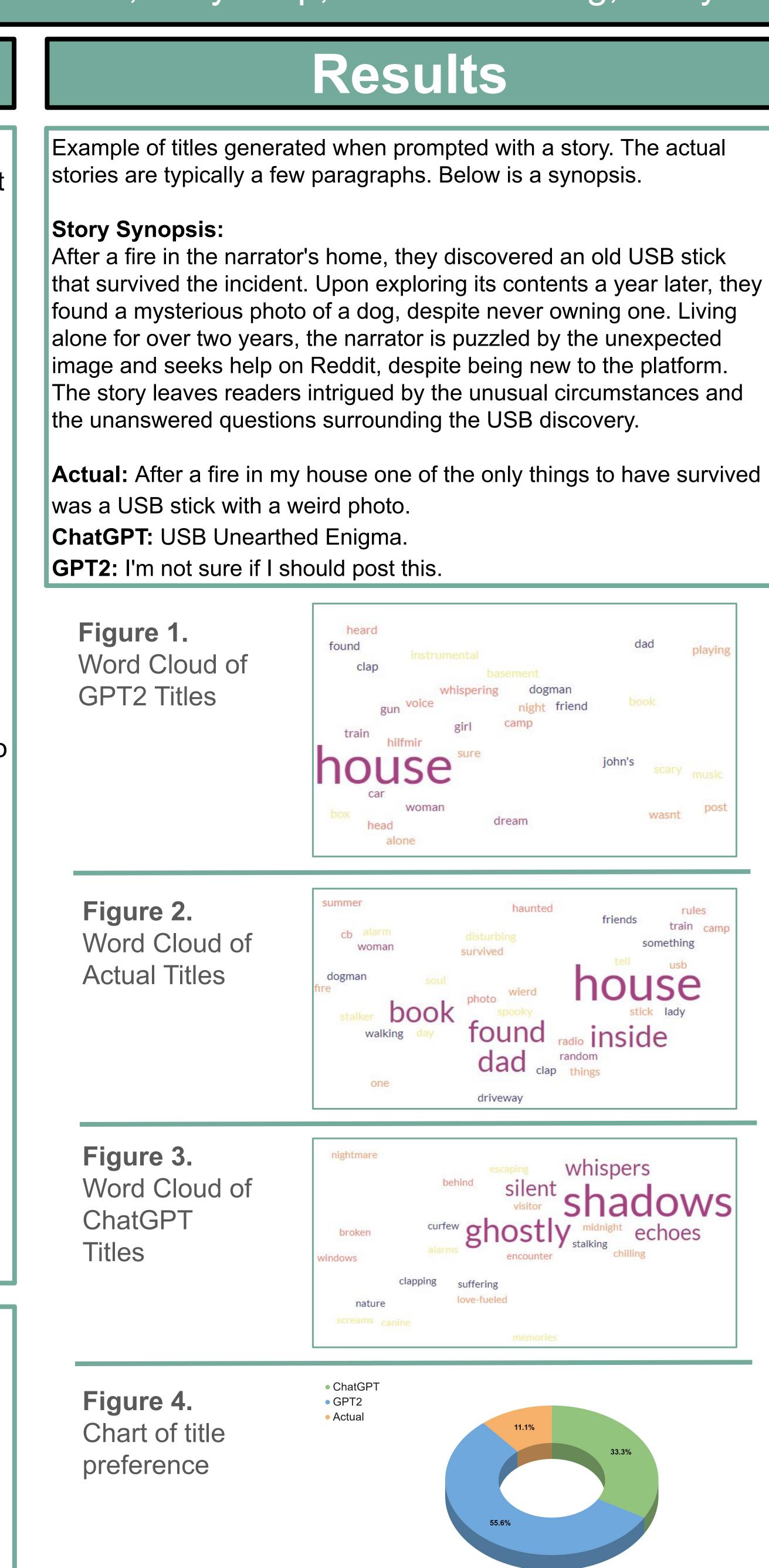
story-text + trigger + story-title

where trigger = "Title:"

Generated titles with the fine tuned model by providing prompt:

story-text + trigger

Title Generation for Fictional Stories Team Title Fight Jacob Malin, Tony Diep, Oscar Wiestling, Cody Cayetano



Analysis

Evaluation:

The actual title of a story was put side-by-side with ChatGPT generated titles as well as our GPT2 generated titles. These were voted on by our group members based on the three attributes defined in the "novel contribution" section. These votes were then aggregated to see which titles were the most creative, resulting in ChatGpt and GPT2 generating more "creative" titles than the original posts, with GPT2 coming out slightly ahead of ChatGPT.

Main findings:

- The actual reddit titles were scored the lowest among the 3 groups.
- Our GPT2 was most frequently selected for best titles.
- ChatGPT titles' were often verb + adj + noun, often had colons, or list of adjectives (felt formulaic)
- GPT2 titles often take on first person perspectives

Plan for final report:

- Further experiment with generation parameters
- Clean and sort the dataset
- Experiment with a larger model GPT-NEO-125M
- Develop refinement model: a separate GPT2 model fine tuned with data-points set up as follows:

generated-title + trigger + actual-title where trigger = "Title:"

- Combine all steps in one title generating pipeline
- Blind testing titles against each other

References

P. Mishra, C. Diwan, S. Srinivasa and G. Srinivasaraghavan, "Automatic Title Generation" for Text with Pre-trained Transformer Language Model," 2021 IEEE 15th International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 2021, pp. 17-24, doi: 10.1109/ICSC50631.2021.00009.

Reddit Horror Story Dataset: https://huggingface.co/datasets/intone/horror stories reddit

