**Project Proposal: Story Completion Using** [**ROCStories Dataset**](https://www.cs.rochester.edu/nlp/rocstories/)

**By Matthew Frank**

I aim to develop a model capable of predicting the missing final sentence in a five-sentence story using the ROCStories dataset. Story completion is challenging as it requires models to understand narrative flow, commonsense reasoning, and world knowledge. To accurately predict the final sentence, the model must comprehend both explicit and implicit cues within the story, drawing on real-world knowledge and narrative progression, which increases the complexity of the task.

The dataset contains 100,000 stories with consistent narrative structures. I’ll use pre-trained language models such as GPT, BERT, and LLAMA fine-tuning them for story completion. We will compare models based on perplexity and accuracy against the Story Cloze Test. This project will implement and build upon existing algorithms for text generation and comprehension.

Additionally, inspired by the rise of short-form content on platforms like Instagram Reels and TikTok, I envision future expansions where original stories can be generated and turned into text-to-speech and video formats. This would create fast-paced, engaging content suitable for modern digital consumption.

**References:**

* [**C.S Rochester:**](https://www.cs.rochester.edu/nlp/rocstories/)
  + [A Corpus and Evaluation Framework for Deeper Understanding of Commonsense Stories](https://arxiv.org/abs/1604.01696)
  + [Story Cloze Evaluator: Vector Space Representation Evaluation by Predicting What Happens Next](https://aclanthology.org/W16-2505.pdf)
* [Papers With Code:](https://paperswithcode.com/dataset/rocstories)
  + [MVP: Multi-task Supervised Pre-training for Natural Language Generation](https://arxiv.org/pdf/2206.12131v3)
  + [Few-shot Learning with Multilingual Language Models](https://aclanthology.org/W16-2505.pdf)