## **Summer of Innovation-2024**

## StoryForge Project Report

## Team Name: AI Noobs

## Team Members:

## - Kulkarni Atharva

## - Priyanshu Mishra

## Dataset Used:

## - Link: [Writing Prompts Dataset on Kaggle]

## (<https://www.kaggle.com/datasets/ratthachat/writing-prompts>)

## - Description: The dataset contains 15,000 stories along with their corresponding prompts, providing a comprehensive resource for fine-tuning our model. These prompts range across various genres and themes, ensuring a diverse training set for the storytelling model.

## Fine-Tuning Process:

## Base Model:

## - Model: [NousResearch/Llama-2-7b-chat-hf]

## (<https://huggingface.co/NousResearch/Llama-2-7b-chat-hf>)

## - Framework: Pytorch

## Data Preprocessing:

## To ensure compatibility with the LLaMA 2 model, the data was preprocessed into the following structure:

## <s> [INST] <<SYS>>

## System prompt

## <</SYS>>

## Prompt [/INST] answer </s>

## This format ensures that the model correctly interprets the prompts and generates appropriate responses.

## Quantization:

## We employed qLoRA for quantization to optimize model performance, reducing the memory footprint without significantly sacrificing accuracy.

## Training and Saving the Model:

## The fine-tuned model was trained on the preprocessed dataset and saved on Hugging Face for easy accessibility and deployment:

## - [Llama-2-7b-storyteller-v2-finetune]

## (https://huggingface.co/kulkarni-atharva/Llama-2-7b-storyteller-v2-finetune)

## Additional Components:

## Text-to-Speech:

## - Library: We used the Google Text-to-Speech (gTTS) library to convert the generated text stories into audio. This step involved processing the text outputs from the fine-tuned model and generating clear, understandable speech.

## Image Generation:

## - Model: For visual representation, we utilized the [Stable Diffusion Model - CetusMix\_v4](https://huggingface.co/redstonehero/cetusmix\_v4). This model was used to generate relevant images based on the themes and elements of the stories produced by the LLaMA 2 model.

## Audio and Video Combination:

## - Library: The MoviePy library was employed to combine the generated audio and images into cohesive video outputs. This process involved synchronizing the audio narration with the corresponding images to create an engaging storytelling video.

## Results:

## The outputs of the StoryForge project include:

## **Text Stories**:Generated by the fine-tuned LLaMA 2 model, showcasing coherent and engaging narratives.

## 2. **Audio Narration**: Converted from text using the gTTS library, providing a vocal rendition of the stories.

## 3. **Visual Representations**: Images generated by the Stable Diffusion model, adding a visual dimension to the storytelling.

## 4. **Storytelling Videos**: Integrated videos combining audio and visual elements, created using MoviePy.

## Attached are images showcasing examples of these outputs, highlighting the effectiveness of the integrated AI components.

## Work Distribution:

## - Priyanshu Mishra: Focused on dataset selection and preprocessing the data to match the required format for fine-tuning.

## - Kulkarni Atharva: Led the fine-tuning of the model and designed the overall pipeline for converting text stories into multimedia outputs.

## Conclusion:

## The StoryForge project successfully integrates advanced AI techniques to create a comprehensive and immersive storytelling system. By fine-tuning the LLaMA 2 model on a diverse dataset of writing prompts, we enabled the generation of high-quality stories. The additional integration of text-to-speech and image generation models further enriched the storytelling experience, culminating in complete multimedia stories. This project demonstrates the potential for AI in creative and entertainment applications, providing a foundation for future enhancements and innovations.

## Future work could involve:

## - Further refining the model’s storytelling capabilities to enhance narrative depth and coherence.

## - Improving the quality and relevance of generated images.

## - Enhancing the synchronization and overall quality of the text-to-video pipeline for a more seamless and immersive storytelling experience.