Overview

In this project I developed a generative AI system that produces illustrated stories. The system generates text and then produces images describing sections of text.

This project is inspired from my own love of reading & dark humour and being a novice writer. So I wanted to experiment with mixing 2 different styles of writing (poems and prose) and 2 different genres of content, namely comedy (Nonsense Poems by Edward Lear) and horror (Haunted Stories by Charles Dickens) in order to attempt to inject horror into funny poems to produce a kind of dark comedy written in rhymes. This is achieved using 2 models, GPT2 (open source predecessor of ChatGPT) and Stable Diffusion (diffusion model provided by Stability AI).

Files linked to this project can be found here. This is the directory structure of inputs & outputs:

- Sem2_CC
  - Content to Finetune Stories
  - GPT2 Model Samples Generated While Training
  - GPT2 Generated Stories
  - Stable Diffusion Generated Images
  - Illustrated Stories (PDF)

Generative Models

1. OpenAI's GPT2

   GPT2 is a transformer based large language model. This project uses the small 124M parameters model variant implemented by Max Woolf [repo]. The model is pre-trained on 8 million web pages to learn to predict the next word in a sequence. The self attention mechanism enables the model to focus on segments of text to maintain context while generating sentences.

   I fine-tuned this model on Nonsense Poems by Edward Lear, Edgar Allen Poe's poems, Haunted Stories by Charles Dickens and a collection of ghost stories from Gutenburg for 500 epochs. The training progress can be seen here in short samples; the model is learning to be a little spooky and even a little poetic.

   The trained model outputs a sequence of 1024 tokens which may be in the form of poems or prose. On close examination of the output it seems nonsensical but the overall theme is more interesting and sometimes a little creepy. The text becomes more nonsensical the longer it goes, a known problem of GPT2 but it is still interesting to read. Some of the generated samples can be seen here and here, where each sample is separated by `==========` (see Colab Notebook where finetuning was done).

2. Stable Diffusion by Stability AI

   This is a text to image deep learning model based on diffusion models using OpenAI's CLIP (Contrastive Pre-Training). CLIP embeds images and text in the same latent space in order to establish semantic relationships between them. So the (cosine) distance between a piece of text describing an image will have a small distance in this embedding space between the text and the image. CLIP was trained on 400 million tagged images to learn these relationships by encoding a text prompt with CLIP and is converted to an image. Latent diffusion models take a noisy latent representation of images and train to denoise it using U-Nets.

The diffusion model was fed sections of text to generate images with various extra prompts for styling (this was done on my laptop as I exhausted the GPU hours provided by Colab). I liked the images generated using the `abstract art` prompt but `Greg Rutkowski` styled pictures seemed to capture the most meaning from the text. They also have a more eerie theme, which is what I am more interested in showing. The `like a joke` prompt had more text in images being created and even though they were illegible, it was interesting to see that comic strips can be created using specific prompts, if one wants to.

The code from lab 9 was used for creating the stable diffusion pipeline which generates the images.

The analysis of these outputs can be seen in this document.

Generative Process

Figure 1 demonstrates the process of generating illustrated content done in this notebook.

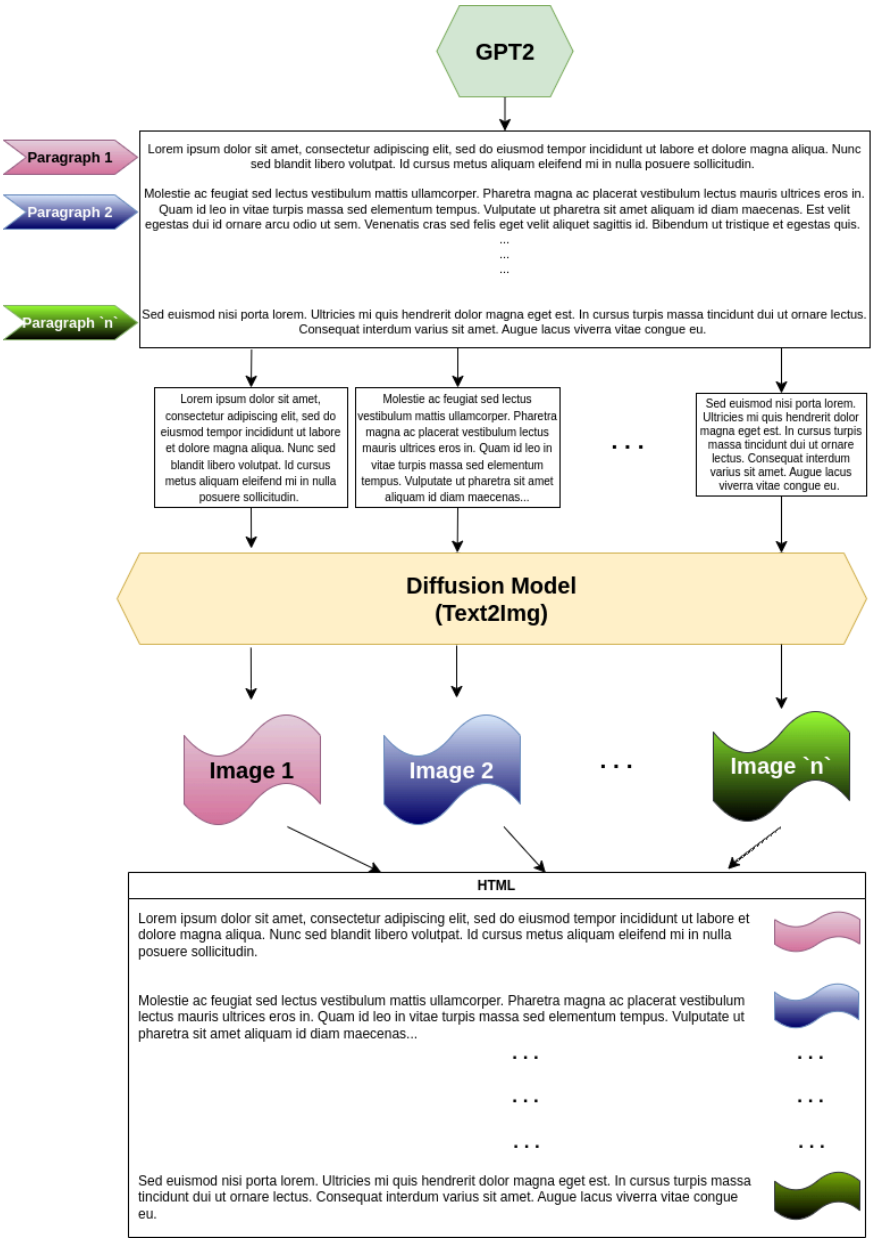

Figure 1: Data Flow Process

For the final output I seeded the text generation for reproducibility with the parameter top_k=20. This refers to sampling for next word generation according to the conditional probability distribution for the top 20 words. The generated story can be seen here.
The GPT2 result was cleaned for presentation purposes and divided in paragraphs.
Each paragraph was passed into the stable diffusion model pipeline to generate an image describing the text, styled like the digital artist Greg Rutkowski. The images can be seen in this directory.

Evaluation

1. GPT2
   The text generated by the trained GPT2 model was evaluated for cosine similarity with the training dataset by embedding the text using Gensim's Doc2Vec. The training data's embeddings was generated and stored to compare with embeddings on the new text. On average the similarity score is very high and this must be because of a lack of diversity in the training dataset. The model is a bit obsessed with *THE PHANTOM WOMAN* from `Ghost Stories8.txt` and uses it in many places as the title.

2. Diffusion Model
   The results of this model were evaluated visually for variety and meaning. Even though the text fed to the model doesn't make complete sense, the model is able to pick up key actions and paint a picture based on it.

My comments on the outputs and sample evaluation are done here.

Value Add & Conclusion

I fine tuned the GPT2 model on a small set of diverse textual content types to analyse what GPT2 learns in terms of overall structure, sentence construction and theme of the content. I was able to help the model produce humour in ghostly stories albeit in an awkward manner. The model definitely struggles as the stories get longer but they are fun to read.
The main highlight of this project is that it attempts at creating multimodal output (text and images). The imperfect sentences really test the diffusion models to describe the text and apply artistic prompts.
Some samples have Miss Piggy's images which look exactly like the original and that raises copyright questions. Also, since the GPT2 model produces content that seem to be very much like its source content, we must question if the models are creative at all even if they do not generate gibberish text or images having illegible text. This exposes us to an ongoing ethical debate about artists and their contents' rights. This topic definitely warrants an in depth analysis and discussion with the wider artistic & AI community.