

Capstone Project Proposal

Project Title:

Social Media Content Generator

Introduction and Objective:

This project will allow users to generate social media content provided an input image and platform type, such as Instagram, Twitter, or blog post. The model will analyze the image and then generate creative text to accompany the image on a post, streamlining the social media process. This would offer small businesses a faster process to generate social media posts, provide scalability for frequent posts, and consistency on brand.

Open-Sourced Pretrained Model:

The architecture of this project will contain two key models pre-trained models:

- BLIP: To caption the input image
- GPT-2: To convert that caption and other parameters into a social media post

BLIP is chosen for being one of the leading models in image captioning. Social media images will be diverse and require clear labeling, so this pre-trained model can be fine-tuned on images specific to social media to get an accurate analysis of the input photo. Natively, this caption is not creative and usually provides concise and straightforward captions, so it will then be handed off to GPT-2 to generate the post as a text-to-text input. GPT-2 will be used due to its availability as a pre-trained, open-sourced model and its training on a variety of data while maintaining scalability in its reduced size. Nonetheless, other LLMs may be explored

Project Definition and Use Cases:

These models will be utilized for content generation. This domain offers creativity and increased efficiency for all, but in particular small businesses and personal brands will benefit the most. The BLIP vision-language model will be paired with the GPT-2 large-language model to offer a simplified solution for generating social media posts. There are many use cases for social media, but today it serves as one of the best marketing channels for many businesses. Different platforms require different types of content, and maintaining consistent and frequent content can be challenging. Not only that, but being creative with social media posts is what onboards consumers onto a product or service. This project will offer custom social media content generation tailored to different platforms such as Instagram, Twitter, or blog posts. By supplying an image with a platform type, the combined model will output either a caption for Instagram or Twitter, or in the case of a blog post, a short article or message related to the image provided. Initially, these three platforms will be explored, but the solution may be applicable to other social media avenues such as LinkedIn, mail distribution lists, and more.

Implementation Plan:

Technology stack:

- Hugging face transformers (for both BLIP and GPT-2)
- PyTorch
- Google Colab
- Streamlit/React
 - I have never used Streamlit, but from a high-level view the tool seems easy to use and quick to deploy the application. Alternatively, React could be used for more customized development but at the cost of increased front-end workload.

Development steps:

1. Import pretrained models into a Colab notebook
2. Gather a dataset for fine-tuning models on social media posts
3. Adjust model architecture to allow for input parameters based on platform and content styles
4. Fine-tune the models and hyperparameters
5. Evaluate and repeat as necessary
6. Develop front-end code in Streamlit
7. Integrate models into the web application
8. Final end-to-end testing
9. Final evaluation

Model Evaluation Criteria:

The following metrics will be used to evaluate the model:

- Inference time
- Human evaluation between real and generated posts
- BLEU and ROUGE scores for similarity to human output

Expected Outcomes and Challenges:

This project is expected to form a complete and working web application in which a user can generate realistic and creative social media content for an image. Since the models are pretrained, it is expected that the output of the combined architecture will be coherent text relatable to the image given, however initial tests without fine-tuning show non-coherent output with repeated tokens over and over again. This leads into potential challenges:

- Modifying model architecture and weights
 - Fine-tuning and getting coherent and descriptive text outputs from the model may pose a challenge. From a standalone manual test with the pretrained models and no fine-tuning, the BLIP model generated an accurate caption, however, the GPT-2 model output random sequences of words and ended with repeated tokens. This model will likely need the most adjustment to get not only readable text, but creative text fitting to the social media post
 - Generating emojis and other special characters frequently used in social media may be required to make the application usable for a small business
- Finding appropriate social media datasets for fine-tuning

- Twitter datasets are commonly available, but doing quick searches on the web and Kaggle has not provided any immediate datasets yet. Some generic image to text datasets could be used, but ideally datasets specific to Instagram are used

Other potential challenges:

- Streamlit development with no prior experience
- Slow generation time and limited compute resources
- Project adaptation for additional features as needed

Required Resources:

The following lists the required resources, but remains a work in progress as more items may be added throughout the project:

- Python, PyTorch, Hugging Face Transformers, NLTK
- Google Colab T4 GPU runtimes (free but limited)
- WPI Turing Cluster
- Streamlit

Conclusion:

Altogether, this project has a very clear design implementation with a distinct output unique from many existing LLMs and applications. With the baseline use of two pretrained models, one for image-to-text and another text-to-text, the architecture will be developed and enhanced around social media content so that a user can input an image and appropriate parameters and the application will output text to accompany the image in a social media post on various platforms.

As of right now, I have tested some of the preliminary ideas and concepts for this project, but the full implementation is subject to change. Depending on how much fine-tuning and model adjustment needs to be done to the pretrained models, additional features can be explored such as incorporating additional parameters and styles to posts, the actual image generation, connecting directly to social media platforms, and even generating full social media profiles. Nonetheless, this project is open enough for variation along the way while still fulfilling the guidelines outlined in this proposal.