November 2025

Group Project

Proposed project I: Fine-Tuning GPT Models for Automated News Generation

Context:

In the rapidly evolving landscape of natural language processing (NLP), generative models such as OpenAI's GPT have become powerful tools for content generation. However, while these models can produce coherent text, fine-tuning them for domain-specific tasks, such as news writing, presents unique challenges. News articles follow specific styles, tones, and structures depending on the topic, outlet, and target audience. The aim of this project is to fine-tune GPT models to generate high-quality, accurate, and stylistically appropriate news pieces.

Project Objective:

This project involves a multi-step process, focusing on data collection, model fine-tuning, and prompt engineering to adapt a GPT model for the task of news generation. Students will work in groups to gather a dataset of news articles, fine-tune a GPT model on their local machines, and explore methods of enhancing news generation using prompt engineering or fine-tuning pre-existing GPT agents.

Project roadmap:

1. Data Collection & literature review:

- Gather a diverse and high-quality dataset of news articles from various reliable sources. The dataset should cover multiple categories (e.g., politics, technology, sports, etc.) to train the model on different writing styles and content.
- Preprocess the dataset to remove noise, ensure consistency, and structure the data appropriately for the fine-tuning task.
- Review what are the state-of-the-art techniques for this task

2. Fine-Tuning GPT Model:

- Fine-tune a pre-trained GPT model using the collected news dataset on local machines. The fine-tuning process should involve adapting the model to the specific language patterns, sentence structures, and tone common in journalistic writing.
- Evaluate the performance of the fine-tuned model using relevant metrics.

3. Prompt Engineering & Agent Fine-Tuning:

- Implement advanced prompt engineering or retrieval techniques to guide a GPT agent in generating more targeted and relevant news content. Experiment with different prompt styles and instructions to optimize the output.
- Optionally, fine-tune a GPT agent or API to improve contextual understanding and reduce hallucinations in generated articles, ensuring the output adheres to journalistic standards.

4. Ethical Considerations:

- Address the ethical challenges of using AI in news generation, including potential biases in data, factual inaccuracies, and the risks of misinformation.
- Develop strategies to mitigate these risks, such as incorporating verification steps or using external APIs to fact-check generated content

Proposed project II: Exploring Transfer Style in Music Generation

Context:

In recent years, deep learning techniques have made significant advancements in various creative domains, including music generation. One exciting area of research in this field is the transfer of musical styles, where the characteristics of one piece of music are applied to generate new compositions. This project aims to deep dive into the fascinating world of transfer style in music generation using advanced AI techniques

Project Objective:

The primary objective of this project is to develop a deep learning-based system that can transfer the style of one piece of music onto another while maintaining structural and melodic coherence. This involves understanding and extracting the underlying musical elements that define a particular style and transferring them to a different composition

Project roadmap:

1. Literature review:

- Conduct a short literature review on music generation, style transfer, and related AI techniques (focus mostly on techniques and model architectures).
- Identify key research papers, tools, and libraries in the field.

2. Data collection and preprocessing:

- Collect a diverse dataset of MIDI or audio files encompassing various musical styles.
- Preprocess the data to ensure consistency and compatibility for model training

3. Prompt Engineering & Agent Fine-Tuning:

- Explore and evaluate different deep learning architectures suitable for music style transfer.
- Choose or design an appropriate model that aligns with project objectives.
- Train the selected model on the collected dataset.
- Experiment with different hyperparameters and techniques to optimize performance

4. Ethical Considerations:

- Develop a method to extract style embeddings from music samples.
- Implement the style transfer mechanism to apply the extracted style onto other compositions

For both:

Project deliverables:

There are three main deliverables of the project

1. Written document including the project's codebase, methodologies, and findings.
2. Code including the implementation of the project
3. Presentation to showcase the project's outcomes and learnings

Annex

Some suggested sources of music from different styles that may be useful when building the dataset:

Classical music:

https://www.kaggle.com/datasets/imsparsh/musicnet-dataset/data

Pop music:

https://www.kaggle.com/datasets/kritanjalijain/music-midi-dataset

heavy metal:

https://metal-midi.grahamdowney.com/midi.html