Task-01 Report – Text Generation with GPT-2

This task involved fine-tuning the GPT-2 language model on a sample dataset to generate human-like text. The objective was to understand the fine-tuning process and observe how the model adapts to new training data.

Tools & Environment
- Platform: Google Colab (Free GPU)
- Libraries: transformers, datasets, torch
- Model Used: GPT-2 (Pre-trained by OpenAI)
- Dataset: Sample text dataset provided during execution

Steps Performed
1. Environment setup with required Python libraries.
2. Loaded GPT-2 model & tokenizer.
3. Created and loaded a sample dataset.
4. Tokenized the dataset.
5. Fine-tuned the model for 3 epochs using Hugging Face Trainer.
6. Saved the fine-tuned model.
7. Generated sample outputs with multiple prompts.

Sample Outputs

Prompt: Once upon a time
Once upon a time, the world of the gods was once a place of utter darkness, of the cold and the wet, of the damp and the wintry. But now it is filled with life and beauty, and of the infinite beauty of

Prompt: Artificial Intelligence will
Artificial Intelligence will have a unique ability to understand the way we use the internet, and how we interact with it. The software will also be able to help with the use of our social media accounts, as well as to create a more efficient system

Prompt: In the year 2050
In the year 2050, the number of people living in poverty will exceed the number of people living in the country at any given moment. The current poverty rate is 5.4 per cent, down from 6.3 per cent in 2014.

Prompt: The secret to happiness is
The secret to happiness is the determination to have a good life. Happiness is not about being happy, it is about being happy. Happiness is about being happy. The happiest person in the world is the one who lives in harmony with the universe

Prompt: My favorite food is
My favorite food is always in the freezer. I love to eat it right away. It's my favorite food to eat every day. I love it on my plates. I love to serve it in my car. It's my favorite food to eat

Observations
- The model produces text influenced by the sample dataset style.
- More data and training improve coherence.
- Sampling parameters like temperature and top_p control creativity.

Challenges
- Limited dataset caused repetitive phrasing.
- GPU time limitations on Colab.
- Occasional incomplete sentences.

Conclusion
Fine-tuning GPT-2 allows customization of generated text to fit a desired style or domain. This task demonstrated how to prepare data, fine-tune, and test a transformer-based language model.