Assignment 6

MGMT 675: Generative AI for Finance

Exercise 21: Train nanoGPT on Shakespeare. Follow the instructions in the slides to train a character-level GPT on Shakespeare's complete works using Karpathy's nanoGPT. Run the training on CPU with the suggested hyperparameters. After training, generate a sample of pseudo-Shakespearean text. Submit a screenshot of the training loss over iterations (`Exercise21-Screenshot.png`), a generated text sample of at least 500 characters (`Exercise21-GeneratedText.txt`), and a half-page reflection on what the model learned and where it fails (`Exercise21-Reflection.pdf`).

Exercise 22: Agent with Streamlit UI. Using Claude Code or a Python script, build an agent with a Streamlit web interface. The agent should have access to at least one tool (e.g., a SQL tool using SQLite with financial data, or a Python code execution tool). The Streamlit app should display a chat window where the user can type questions and see the agent's responses. If using a database agent, display query results in a formatted table. Deploy the app locally and submit the code (`Exercise22-Code.zip`), a screenshot of the running app (`Exercise22-Screenshot.png`), and a brief description of how the agent loop integrates with Streamlit (`Exercise22-Description.pdf`).

Exercise 23: Headline Sentiment Classification. Collect 20 recent financial news headlines (from sources like Reuters, Bloomberg, or Yahoo Finance). Using Claude or the OpenRouter API, classify each headline as positive, negative, or neutral for the stock mentioned. Also have Claude rate the magnitude (1–5) and relevance (low/medium/high). Compare Claude's classifications with the actual stock price movement on the day of the headline. Submit the 20 headlines, Claude's classifications, the actual price movements, and a summary of Claude's accuracy (`Exercise23.xlsx`). Submit a screenshot of Claude's classification output for at least five headlines (`Exercise23-Screenshot.png`).

Exercise 24: FinBERT vs. LLM Comparison. Using Google Colab, run FinBERT (from HuggingFace's `ProsusAI/finbert` model) on 10 financial headlines. Then classify the same headlines using an LLM (Claude or GPT via the API). Compare the two models' sentiment predictions on accuracy, nuance, and speed. Identify at least two headlines where the models disagree and explain which model's classification is more reasonable. Submit the notebook (`Exercise24.ipynb`), the comparison table and your analysis (`Exercise24-Analysis.pdf`),

and a screenshot of the Colab notebook showing FinBERT's output alongside the LLM's classifications (`Exercise24-Screenshot.png`).

Submission. Upload a zip file containing each file requested above with the filename shown in parentheses. Label the zip file "Assignment6' '.