AI Task 4: Model Evaluation and Comparison

4.1 Model Outputs Comparison

Question	Before Fine-Tuning	After Fine-Tuning	RAG Naive	RAG Advanced
What about COVID-19?	COVID-19 is a genus of "Coviridae".	COVID COLU-VERNATION	COVID was a bad thing in many ways	COVID-19
Summarize Trump's statements on the COVID-19 pandemic, focusing on his views on government response, China, and measures promoted.	Donald Trump's statements focusing on government response, China, and measures promoted.	Trump's comments centered on government response to China's role in COVID.	President Trump said China caused $60 trillion damage and millions of deaths.	Trump said China caused $60 trillion damage and millions of deaths.
How's the economy?	Good time to buy a home.	Blank statement	Economy is very bad	Very bad economy currently.
Summarize Trump's position on the U.S. economy, including claims about growth, unemployment, and stock market.	Comments on growth, unemployment, and stock market with transcript references.	President Donald Trump's position on the U.S. economy.	Tariffs and trade caused economy's growth.	Not trading, will impose large tariffs.
What about tariffs?	Taxation on imports/exports is key for economy.	Tariffs are inconsistently included.	Tariffs are bringing in business.	Word "tariff" is beautiful.
Summarize Trump's stance on tariffs, especially toward China, including goals, justifications, and expected outcomes.	Trump's stance on tariffs, goals, and outcomes.	Long-term impact on China, he said.	Wants to remove tariffs on China to cover inflation; use tariffs for strength.	Remove tariffs on China, end tariffs, impose big ones when needed.
What about immigration?	Immigration is a criminal offense, leading to deportation.	Migrant	Jobs taken by illegally imported migrants	Jobs and factories will return home.
Summarize Trump's position on immigration policy, border security, wall, and restrictions.	Position on immigration, border security, wall, restrictions.	Trump's position on immigration.	Strong stance, wants wall along border.	Strong record, highlighting border security, wall, restrictions with quotes.
What about national security?	Federal responsibility to protect citizens' privacy and security.	National security measures citizens' safety; U.S. as model.	Industrial imbalances put national security at risk.	Government's highest duty is protecting citizens, especially during emergencies.
Summarize Trump's views on national security, military strength, terrorism, and foreign policy.	Transcript of President-elect's address.	Trump's comments on national security.	U.S. President shows great strength and courage.	President Trump strongly committed to national security.

4.2 Fine-Tuned Model Analysis

The fine-tuned model performed worse than the baseline. Many answers are gibberish or repeated words, demonstrating overfitting. While occasionally relevant words appear, the model fails to generalize or preserve Trump's style. This likely results from using each phrase as both input and output during training, causing memorization rather than comprehension. Fine-tuning is limited when the dataset lacks real question-answer pairs.

4.3 RAG Analysis

Naive RAG outperformed all other methods, producing fewer hallucinations and more sensible outputs, even with a small model. Advanced RAG did not improve results, likely due to a weak tokenizer. Overall, RAG significantly improved output quality and better captured Trump's style. Prompt engineering helped, but smaller models benefit most from simpler RAG pipelines rather than complex query reformulation.

4.4 Conclusions

- LoRA fine-tuning was the least effective option in terms of performance vs. time.
- Prompt engineering may help larger models with more complex datasets.
- Both RAG implementations are far more effective, leveraging a vector database to retrieve relevant phrases, even for small models.