**Overview**

This repository contains the codes to run the presidential chatbot. Please follow the instructions to run the code and replicate results.

**Running the App**

**Option 1: Direct App Access**

1. **Download:** Download the chroma\_db folder “[chromadb\_combined\_data](https://github.com/smitpancholi313/Final-Project-Group2/tree/main/Code/chromadb_combined_data" \o "chromadb_combined_data)” and the fine-tuned model folder “[fine\_tuned\_president\_5\_epochs](https://github.com/smitpancholi313/Final-Project-Group2/tree/main/Code/fine_tuned_president_5_epochs)” to your project directory
2. **Run the App:** Run the “[FinalStreamlitApp.py](https://github.com/smitpancholi313/Final-Project-Group2/blob/main/Code/FinalStreamlitApp.py)” Using the streamlit run command “streamlit run FinalStreamlitApp.py --server.address=0.0.0.0 --server.port=8888”. When viewing the app in the browser, replace localhost with the public ipv4

**Option 2: Fine-Tuning and Database Creation**

1. Download the data files “[speeches\_russian\_PM.xlsx](https://github.com/smitpancholi313/Final-Project-Group2/blob/main/Code/speeches_russian_PM.xlsx)” and “[speeches.xlsx](https://github.com/smitpancholi313/Final-Project-Group2/blob/main/Code/speeches.xlsx)” to your project directory
2. **Fine-Tune the Model:** Run the “[FineTuning.py](https://github.com/smitpancholi313/Final-Project-Group2/blob/main/Code/FineTuning.py)” file. This file trains a GPT-2 model on the dataset, and stores the model to your project directory*. Note: This step may take approximately 1 hour.*
3. **Create Chroma Database:** Run the “[Chroma\_creation\_db.py](https://github.com/smitpancholi313/Final-Project-Group2/blob/main/Code/Chroma_creation_db.py)” file. This code chunks the data and stores their vector embeddings to a chroma vector database in your project directory *Note: This step may also take approximately 30 minutes.*
4. **Run the App:** You can run the app now