Email Parsing Using LLMs

Description

This project employs LLMs to parse emails and obtain the sender's signature in JSON format. Two models were compared: **LLaMA-3.1 Chat (8B)** and **GPT-3.5 Turbo** and the findings summarised in [prompts.md](./prompts.md).

Custom instructions (the main prompt) are defined in [system\_content.txt](./system_content.txt). The emails to be parsed are provided as user input while running the application.

The models may be evaluated by inspecting real time inputs when running the application interactively. Alternatively, you may analyse the logs produced in [email\_parser.log](./email_parser.log)

Three prompt iterations were done, and **GPT-3.5 Turbo** showed the most desirable results for each iteration. It strictly followed the instructions and only returned JSON while **LLaMA-3.1 Chat (8B)** was verbose and included additional content in its responses.

Usage

Clone this repository and run the following commands in a unix shell (or Git Bash in Windows):

pip install -r requirements.txt  
API\_KEY="\*\*\*\*" python email\_parser.py

API\_KEY refers to an API key you can obtain from <https://aimlapi.com>. You may set it as a system environment variable or provide it when running the script as shown above.