OPENAI API key not working due to insufficient quota

Looking at

* RAG
* LLAMA
* Mixture Of Experts
* LLMs
* Vector-Databases
* Embeddings – help to search through the datasets
* Inference and its definition
* Evaluating Different LLMs
* Justification behind choosing a certain architecture of embeddings model and LLM
* Knowledge Graphs for Retrieval Augmented Models
* RAG METRICS
* GuardRails around LLMs which makes them answer questions which pertain only to the specific topics they’re trained to serve on.

Shortcomings:

* Lack of attribution, figuring out which document is responsible for answer is not possible.
* They’re not knowledge bases, they’re probabilistic models of knowledge base.
* Implicit world knowledge (in LLM) can interfere with retrieved Documents to cause hallucination

What can we suggest: Retrieval Augmented Model

A screenshot of a computer

Description automatically generated  
  
  
A screenshot of a chat

Description automatically generated

Useful Links

<https://www.canva.com/design/DAFw0D8y038/5Yh9MA2XXd2Lfr2thcsuLA/edit>

<https://colab.research.google.com/drive/1TDiWZtb6gsM9wVXCLQrR-7OEPaQ2n-JA?usp=sharing#scrollTo=vAwklqQCgVi->