Cyber LLMs

Aritran Piplai
Assistant Professor, UTEP
Where is the major international conflict happening now?

GPT 3.5

Russia and Ukraine

My apologies, I am a language model not the New York Times

wrong its in Israel and Palestine
Dealing with new information

Prompt: Where is the major international conflict happening now?

Response: Russia and Ukraine

Response: Israel and Palestine
In cybersecurity

Prompt: How can I make myself safe from Pegasus malware?

Response: Just don’t click on any links (Pre 2020)

Response: Disable certain apps (Post 2020)
How do you solve this problem?

• Option 1 (Fine-tuning)
  • Get more recent data, train and update the model
  • Computationally expensive, the smallest LLMs have 7 Billion parameters!

• Option 2 (RAG)
  • Give the model access to recent data, use the old model
  • Computationally more efficient, model retains old information
Option-2 (Retrieval Augmented Generation)
Prompt: How can I make myself safe from Pegasus malware?

Response: Just don’t click on any links (Pre 2020)
Response: Disable certain apps (Post 2020)

Seek relevant information

External knowledge base

Cyber threat reports
Research outcomes

• Does RAG always help?
• Is there a minimum data requirement in the knowledge base?
• [optional] Is the textbook that you use in your class sufficient to answer all the questions in exams/quizzes?
Timeline

Week 1
- Choose the data (I have some data, but look for more sources)
- Clean the data

Week 2
- Get familiar with huggingface (starter code available)
- Create the vector DB (Knowledge base)

Week 3
- Build the RAG system
- Evaluate the performance

Week 4
- Repeat the experiment with another dataset (bigger or smaller)
- Evaluate the performance

Week 5
- Generate conclusions and report the findings
Questions?