

LLM's for Development

Or how to outsource your thinking for fun and profit

What am I trying to communicate?

- What are LLM's? What does it mean to model language?
- What LLM tools are available right now?
- How do LLM's work? Why can machines suddenly think?
- How can you apply LLM's to software development tasks?
 - What are they good at?
 - What are they bad at?
 - How can you improve the quality of LLM responses.
- What is the trajectory of LLM innovation?



So like what is a language model?

- LLM's are next-word prediction engines.
 - Good an answering questions like: "Mary had a little [WHAT?]"
- Text prediction has existed for a long time, but hasnt been particularly impressive.
 - "I went to the Calgary Stampede and met a nice [day]" (phone keyboard)
 - "I went to the Calgary Stampede and met a nice [cowboy]." (GPT-3.5)
- LLM's are able to accomplish this feat by training on a absurd amount of data
 - "Attention is all you need"
- **A good language model is a model of the world**



But ChatGPT doesn't just complete sentences...

- There are many different kinds of language models
 - Text Classification, Token Classification, Table Question Answering, Question Answering, Zero-Shot Classification, Translation, Summarization, Conversational, Text Generation, Text2Text Generation, Fill-Mask.
- Base models are trained by ingesting massive amounts of data, these models are fine-tuned into final applications.
- OpenAI can outperform smaller firms due to the capital required to **RLHF** their models (**reinforcement learning human feedback**)
 - GPT-4 inference is run on GPU clusters with 320 - 640gb of vram



Ok so how can I use this for software?

- Refactoring
 - Ask it to functionalize an idea out of spaghetti code
- Testing
 - Automatically generate test cases for rendering engine
- Feature development
 - Quickly develop script to process and get insights from dataset
- Debugging
 - Quickly print out complex data structures
- Researching
 - Explain and interpret the technical jargon outputted by cellular modem commands
- Textual manipulation
 - Converting unorganized web-scraped data into CSV



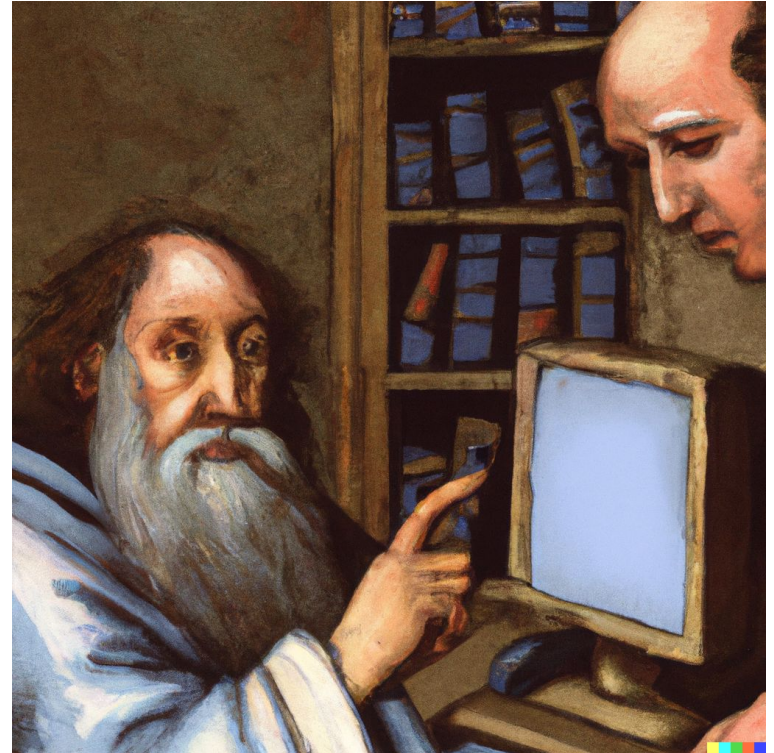
Case Study: CBOR Decode Script

1. Input raw CBOR annotated by cbor.me
 - a. “write a python script to decode the above commented cbor, have it take input cbor over the cli. Think carefully about cbor and decoding input generally”
 - b. GPT-4 outputs a python script that can decode CBOR
2. Input .h files that define message types
 - a. “make python enums from the above c header”
 - b. *after some encouragement* GPT-4 outputs a script that decodes the CBOR with named symantic names



How can I improve the output from LLM's?

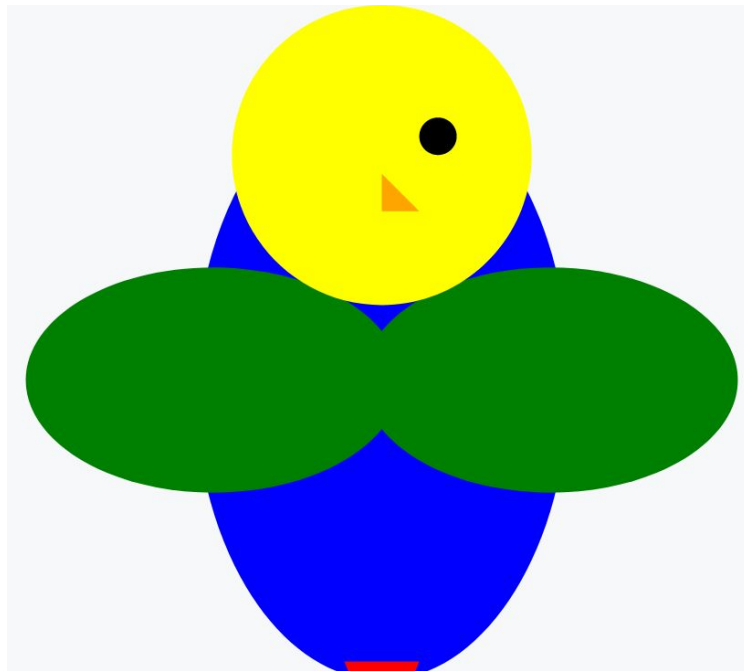
- Ask the model to **“think step by step to get the correct answer”**
 - By allowing the model to explain itself it can come up with more insightful output
- Ask the model to imitate an expert in the field
 - The model has been trained on all sorts of information, not all of it is expert advice
- Give the model simple, achievable goals, systematically increment of this output in testable ways
 - You can quickly confuse a model by asking for too much in one go



Misc tips and tricks

- Python is the home language of GPT-4
 - For complex tasks you are better off using python than bash
- GPT-4 is excellent at natural and computer language translation tasks
- You can use LLM's to generate useful bash one-liners
 - `DATE="2023-07-07"; git log --reflog --author="Gavin Guinn" --stat --source --decorate --since="$DATE 00:00:00" --until="$DATE 23:59:59"`
- GPT-4 was trained of image data and has some concept of how things *look*

PARROT



Last thoughts

- Giving LLM's more tools will go a long way
 - Code Interpreter enables GPT-4 to implement and **iterate** on python scripts that process uploaded data
 - Function Calling enables GPT-4 to call user definable functions by outputting user definable JSON
- Open-source models are improving rapidly but are not yet competitive with the reasoning GPT-4 is capable of
 - This gap will likely continue to narrow as model compression improves
- Are we in for the singularity in 10 - 100 years
 - *idk maybe*

