

①

GenAI for Everyone

GenAI:- AI systems that can produce high quality content specifically text, images & audio.
e.g. chatgpt

GenAI is build on

= LLM's are built by using supervised learning (A \rightarrow B) to repeatedly predict the next word

Supervised learning

= When we train a very large AI system on a lot of data (hundreds of billions of words) we get a LLM like chatgpt.

Example:-
Writing

(app)

Reading

Giving large info
& having it generate
short output.

or

Customer review
is that a complaint
or not
(LLM based
software app)

(app)

chatting

②

③

GenAI does not work well with structured data (tabular)

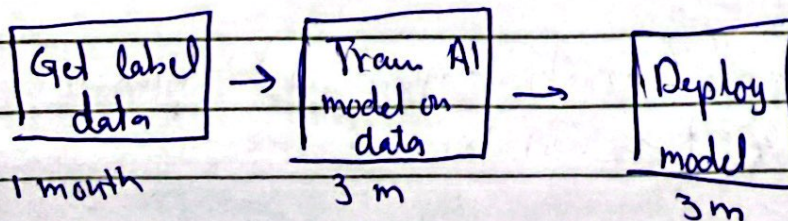
- supervised technique would be better
- If you give the data without labels it won't work it's better to give it labeled

Tips for prompting:-

- Be detailed & specific

Week #2 Software applications

- Traditional Restaurant review system



Prompt based:-

prompt = """ Classify the review +ive or -ive ~~or both~~] instruction set
The banana pudding was really tasty """] review text

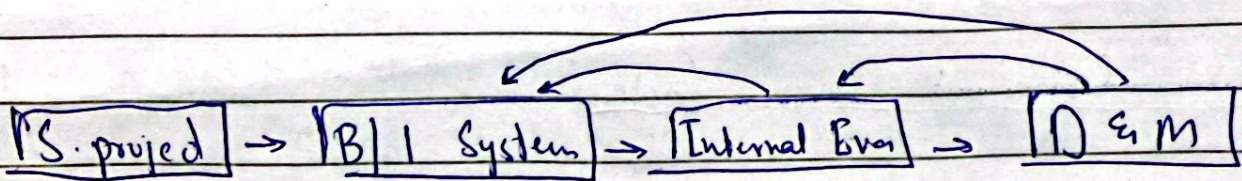
response = llm_response(prompt)
print(response)

} code to call LLM
- code to print output

(4)

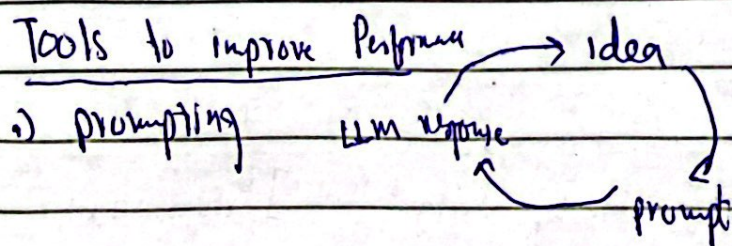
= Life Cycle of a generative AI project

- Scope project: what we want this project to do
- Build / improve system: - Initially a prototype that we will improve over time.
- Internal evaluation:-
- Deploy & monitor



Building GenAI is a highly experimental process. we repeatedly find & fix mistakes.

Tools to improve Performance



1) prompting

4) Pretrain models
• Train LLM from scratch

2) Retrieval augmented generation

Give LLM access to external data sources

3) Fine tune models

- Adapt LLM to your task

⑤

RAG:- It give additional knowledge beyond what it may have learned from data on the internet or open sources

e.g

General ChatBot

ChatBot with RAG

→ Is there parking for Employees?

→

→ I need more specific info about your workplace

→ Yes you can park on level 1 & 2 of office lot.

How it does?

Step #1 : first look through a collection of documents that may have the answer

e.g. Benefits, leave policy, facilities, payroll.

Step #2 : - Incorporate retrieved text into an updated prompt

- We retrieve the relevant context or relevant info & augmenting the prompt with that additional text

e.g. Use the following pieces of context to answer the question at the end.

Parking policy:

Is there parking for employees?

⑥

Step #3: Generate answer from the new prompt with additional context-

Pretrained: process of training LLM on lot of data is often called pre-training.

Fine-tune: It is just the priced way to tell LLM that what you want.

Why fine-tune?

To carry out a task that isn't easy to define in a prompt.

Example: Summarize in certain style or structure

To get a smaller model to perform a task.

- Lower cost / latency to deploy
- Can run on mob / laptop (edge devices)

⑦

Week #3 GenAI & business

Augmentation:- Help human with a task

e.g. recommend a response for a customer service agent to edit / approve.

Automation:- Auto perform a task.

e.g. auto transcribe & summarize records of customer interaction

- For some task, businesses will start with augmentation & gradually move towards automation.

- AI doesn't auto job it auto tasks.

The potential for Augmenting / Automating depends mostly on two things:

① Technical feasibility

Can AI do it?

= How costly is it to build

an AI system to do it?

= If unsure, try prompting an LLM to see if you can get it to do it.

= An AI engineer can help assess if RAG, fine tuning or other techniques can help.

② Business Value

How valuable is it for AI to augment or automate this task?

- How much time is spent on this task?

- Does doing this task significantly faster, cheaper or more consistently create substantial value.

Teams to build GenAI software:-

- Software engineer

- Responsible for writing software application. (making sure it runs robustly)
- ideally someone who has learned basics of LLM/prompting

- MLE

- Responsible for implementing AI system
- ideally familiar with LLM/prompting, RAG, fine-tuning

- Product manager

- Responsible for identifying & scoping the project

- Prompt engineer?

- usually not hired as a dedicated role. usually MLEs do this job.