

27/01/24

Chapter - 1

Introduction to Generative AI

GenAI is subset of AI that create new content based on existing data, with the ability to produce text, image, audio & synthetic data.

AI :- It is broader discipline encompassing the creation of intelligent systems.

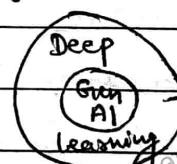
Machine learning :- It focuses on training models from input data.

→ Supervised
(Rely on labeled data for prediction)

→ Unsupervised
(focus on discovering patterns in unlabeled data)
(Correlations)

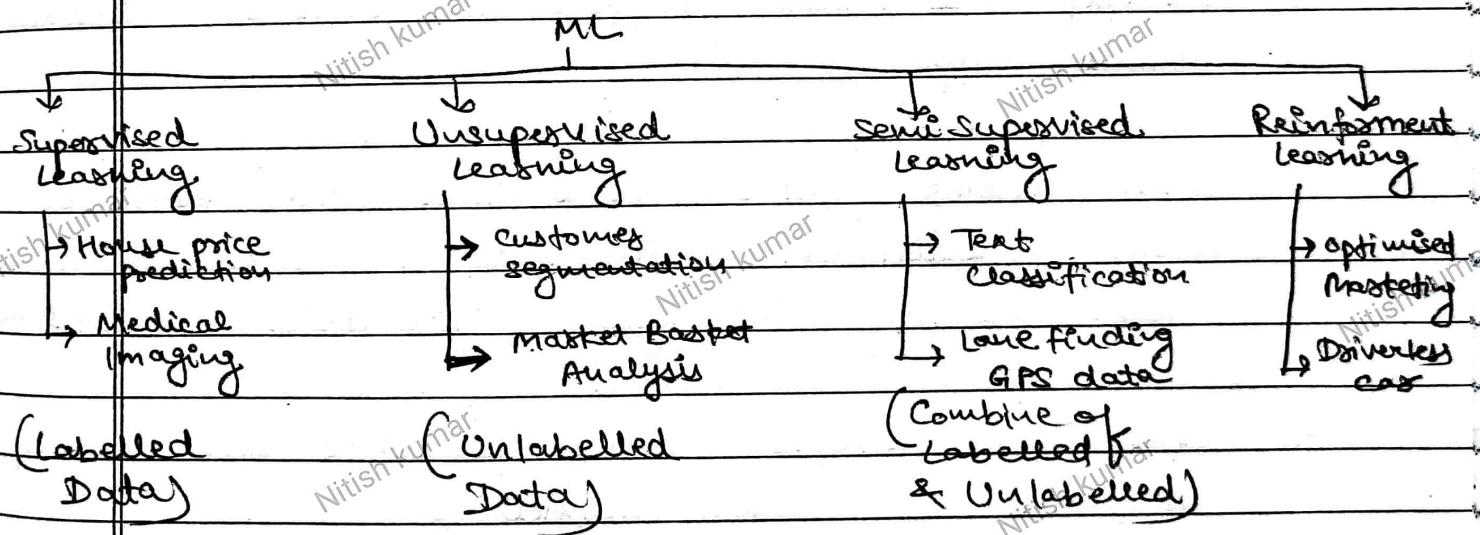
→ further this divides into deep learning, a type of machine learning that utilize artificial neural network to process complex pattern

→ GenAI is subset of deep learning



→ Generative models are contrasted with discriminative models, where generative models produce new data instances based on learned probability distribution while discriminative models classify or predict labels for data points.

- Transformers are described as revolutionizing natural language processing by utilizing encoder-decoder architectures to process input & produce output.
- Prompts :- They are introduced as short pieces of text used to control the output of large language Model.
- foundation models are large AI models pre-trained on extensive dataset, capable of adapted customisation (downstream task)
- PALM API is highlighted as a tool for testing and experimenting with Google's large Lang Models.



Reinforced :- Rewarding desired behaviours & punishing undesired ones.

- A technique that trains an agent to learn how to behave in an environment.

28/01/21

Chapter - 2

Generative AI for everyone

How AI works

→ supervised learning which excels on labelled data, is briefly explained through examples such as spam filtering, online advertising, self driving car

→ With supervised learning, emphasizing the importance of large scale model in improving performance, This tells GenAI which is built on LLM developed through

→ LLM predict the next word in sequence of text based on input. By leveraging vast data LLM become efficient, enabling the model to generate coherent & continually relevant output.

→ Video shows LMs in various task such as writing assistance & information retrieval, indicating their increasing integration into daily work activity.

LLM as thought partner

→ Server as thought partner in tasks like writing assistance & creative storytelling

→ Limitations of relying solely on LMs such as their propensity for "hallucination" (false perception) or generating false information.

→ User are advised to take advice from specialist after getting little knowledge e.g. in medical, recipes etc.

→ User should be mindful while using AI of their strength & weaknesses when deciding whether to use them. further AI capabilities and discussing test practices in subsequent session.

AI general purpose tech

AI serves a multitude of purpose due to its general purpose nature.

1. Used for categorizing useful & unusual mails
2. Also powers chatbots, facilitating tasks like taking online orders or providing customer support

4. LLM classification:-

Web interface based
↓
deep learning with
LLM through Online
platforms

Application & software
Integrate LLM functionality
into larger software
systems, enhancing
automation & workflow.

- ① Accurate description of LLM :- generates text by repeatedly predicting the next word.
- ② Why we call AI a general purpose technology.
Because it is useful for many different tasks.

General Application

Translation is another work where we can use LLM especially for languages with abundant resource

→ Versatility of LLM in assisting with various writing tasks, highlighting utility as creative partner

- 1) Proofreading:- used to proofread text for spelling, grammatical error, awkward sentences.
- 2) Summarization:- summarizes lengthy articles or documents into concise summaries.

- (3) Call centre Management :- Provide summary of multi-channel interactions.
- (4) Email Analysis :- Analysing customer emails.
- (5) Reputational Monitoring :- Monitor customer sentiments.

(Human only service centre) (fully automated)

- Bots supporting humans by generating suggested message for approval (human in loop)
- Starting with internal-facing bots for testing

Under
don't

- A useful mental model for assessing what LLMs do.
- 1) Knowledge cutoff :- LLM knowledge is specific upto a point
- 2) Hallucinations :- May generate inaccurate info producing wrong answers
- 3) Input & Output length limit :- Length of word limit
- 4) Incompatibility with structured data :- Struggled with Tabular data found in spreadsheet, suited for unstructured data
- 5) Bias & Harmful speech :- May produce biases present leading to biased or harmful output

→ Except these limitations, used with awareness of their constraints.

Tips for
prompting

- 2) Be detailed & specific
- 3) Guide the model through the process
- Experiment & Iterate

Two important caveats are highlighted:-
ensuring confidentiality of info when using LLMs
& verifying the accuracy of generated content
before acting on it, especially in sensitive contexts like legal documents.

LLM is highly iterative process require multiple prompts

Training process :- Diffusion model learning from large data
supervised learning employed

- ① Generation process :- Noisy image is inputted to trained model.
Model iteratively removes noise from input image.
- ② Incorporating prompts :- Modification to the algorithm.
Model receive noise image from input as well as prompt.

for grammatical check :- Don't type too much time.

Test prompt if code is right Use it to make a software that is really hard without AI.

Several iterative steps. Giving scope & define objective
Implementation → Internal evaluation → Deployment externally → Monitor → Reinforcement → Experimentation → Various tools & technique, such as retrieval augmented generation (RAG) & fine tuning
→ feedback & performance Monitor

Cost of all AIs.

Ques either supervised learning or prompt - based development to build restaurant review sentiment classifier?

Which statement about prompt based development is correct :- prompt based environment faster than supervised because don't require to train a model from scratch.

Ques Token in context of a large language model,
A word or part of a word in either the input prompt or LLM output. common word as single token & uncommon words may be broken into 2 or more tokens.

additional information through prompts

② Retrieval Augmented Generation (RAG) a technique that enhances the capabilities of LLM by providing additional knowledge beyond what learned from open sources like internet :-

- 1) Retrieval :- Given a que system search in collection of data
- 2) Augmentations :- Retrieved text is incorporated into an update prompt
- 3) Generation :- LLM is prompted with enriched context and original question, enabling it to generate a response with additional knowledge.

Application :- chatting with PDF file

→ Answering question from website articles

→ Transforming web search

→ small data set training

fine tuning

Technique used to enrich the capability of LLM complementing method like Retrieval Augmented Generation

→ Involve training an LLM on specific datasets to adjust its outputs to desired criteria.

LLM size choosing

First parameter is size of LLM.

Smaller module (1 billion parameter) adept at pattern matching & basic world knowledge

Medium module (10 billion parameter) greater world knowledge & suitable for making chatbot

Large module (100 billion parameter) rich world knowledge, excel in complex reasoning & require deep knowledge or complex reasoning, brainstroming

③ ② closed source & open source.

accessible via cloud

→ full control can be given on various device (for privacy)

③ selection process involve experimentation & testing different module to best suit application requirement

LLM
construction

① for safety reinforcement learning from human feedback (RLHF) is employed.

② RLHF helps improved quality of LLM by providing reinforcement or rewards for generating better answer.

③ Combination of construction tuning & RLHF enables LLM to follow instruction effectively & produce ans.

→ with the help of RAG we can also make a LLM to answer modern day new answer.

→ Using LLM as a reasoning engine refers to idea of not using our LLM as a source of info but to process info also.

→ Jobs are compromised of many task & task. AI automates task rather than job.

④ Concern of AI :-

① Amplification of Negative Impulse

② Impact on Employment

③ Existential Threat

④ Optimisation for future

⑤ Responsible AI :-

fairness, Transparency, Privacy, Security & ethical use.

Implementation is challenging :-

1. encourage discussion

2. Brainstorm Potential issue

3. Include Diverse Perspective

- ① RLHF (Reinforcement learning from Human feedback) helps to align an LLM to human preferences & can reduce the bias of an LLM output.
- ② Tasks that Artificial General Intelligence can solve in 20 yrs:
 - Learn to drive a car in 20 yrs.
 - Compose the music for a movie soundtrack.

CHAPTER

1. Introduction to Large Language Model :-
2. Definition :- LLM are a subset of deep learning and are part of generative AI. They are general purpose model pre-trained on large datasets & fine tuned for specific purposes.
3. Application :- (Benefit)
 - Versatility :- Single model can be used for different tasks.
 - Efficiency :- require minimal field training data & can perform well even with few, zero-shot scenarios.
 - Example :- PALM (Pathway Language Model) 540 billion parameter released by Google
 - Transformer Model :- comprising encoder & decoder components.
4. Comparison with Traditional ML Development :-
 - LLM Development :- Don't require expertise, training examples, or model training, emphasizes prompt, e.g.
 - Traditional ML Development :- Require expertise, training example, model training, significant computing resources.
5. Use Case :-
 - Text Generation :- PALM
 - Question Answering :- Like Bard, a chatbot provides accurate responses based on prompts.

6. Types of LLM :-

- Generic Language Model :- Predict next word.
- Instruction Tuned :- Trained to respond to specific instruction.
- Dialog Tuned :- Specialized for dialogue interaction. Typically framed as questions to a chatbot.

7. Parameter efficient Tuning Method :- Allow tuning LLM on custom data without altering base model, making process more efficient.

8. Tools provided By Google's Gen AI development suite :-

Generative AI studio :- Offer pre-trained models, fine-tuned tools, deployment options & community forum.

Vertex AI :- (Gen App Builder) Enables creating chatbot, search engines, knowledge bases.

PaLM API :- Allows testing & experimentation with Google's LLM & Gen AI tools, integrated with Makersuite for accessibility.

INT 426Chapter - 2Gen AI for everyone.

- ① Gen AI :- It is a type of AI that is capable of generating new images, text, audio, synthetic data that have similar characteristic to input training data
→ It learns the patterns and structures of its input training data.

Journey of Gen AI :-

1. Abundance of Data :- (463 exabyte data by 2025)
2. Increased Computing power :-
3. Evolved learning Algorithms :- for structured & unstructured
4. AI first Approach :- A single model can be used for multiple tasks and capabilities.
5. Versatility of foundation Models :- Do language translation, Content Generation.
6. Organizational Effectiveness :- Highlight talent, culture etc.
7. Call to Action :- familiarizing oneself with advanced models

GPT

1. Designed to generate human like text based on the input it receives. Do content generating.
2. Used for language processing, text completion, language translation.
3. Pre-trained on various datasets, learning pattern & data structure fine-tuned for specific purpose.

Search Engine

1. It is a tool that indexes and retrieves information from vast content available.
2. Focus on retrieving info. that matches user query.
3. Trained on data from web, consisting of various types of content. Indexing tool.

4. Can generate responses based on user input. It can understand query.
5. Has knowledge upto a given cutoff and can generate text based on that knowledge. Don't have realtime info.
- Provides list of link in response to query. Don't generate dynamic, context-aware responses.
- Can provide latest information available on web internet. Continuously updates indexes to reflect the current state of web.

- ① Primary goal of Gen AI :- generate new & original data
- ② How Gen AI effect organisational efficiency?
By automating repetitive task ✓
- ③ Key consideration for implementing ethical aspect of Gen AI :- ensuring unbiased model output
- ④ Important consideration for ethical deployment & responsible practices in Gen AI :- Bias & fairness in Algorithm Transparency & explainability of AI
- ⑤ Diff b/w Gen AI & discriminative AI :-
Reinforcement focus on generating new data & discrimination classify existing data.
- ⑥ Impact of GenAI on organizational efficiency : - By speeding up development & deployment of AI & By generating realistic synthetic data for training AI models ; By optimising resource allocation & streamlining complex business operation.
- ⑦ Purpose of Gen AI :- Generate new content learned from entity
- ⑧ ethical consideration when implementing Gen AI's - ensuring transparency & explainability of AI generated outputs
- ⑨ How responsible implementation of Gen AI benefit society?
By fostering creativity & enabling new forms of artistic expression.

~~week 9~~

Generative AI in daily life.

- Gen AI acts like virtual assistant in smartphone, powering voice-activated features like Siri and Alexa.
- Also works as fitness tracker, analyze work data.
- for traveller work assistance, language translator.
- Despite being benefit, ethical consideration are crucial.

① Content creation with AI

② Anecdote about AI chatbot (Microsoft Tay) racist

③ ChatGPT & Large Language Model [LLM]

④ Advantages of AI Generated Content

⑤ Preserving Authenticity & Credibility

⑥ Human input in content creation

⑦ Language translation & Cultural consideration

⑧ Diversity and Inclusivity in AI training Data

⑨ AI as a writing Assistant

- AI content is considered a useful tool but not a replacement for human intelligence & originality
- emphasizes the ongoing role of individuals in bringing unique perspectives & creativity to the table.

⑩ AI Empowerment :-

① Challenges of AI in Media creation :- Requires advance algos

② Bias in AI generated Media :- Biased training

③ Importance of Diverse Training Data :-

④ Potential of Text-to-Media AI

for media generation

⑤ Technological Advancement :- OpenAI DALL-E & Lumen5

⑥ Attention - Grabbing Content in the Modern World

future with AI :-

- (1) Automation :- increase efficiency & productivity
- (2) In healthcare :- can diagnose, drug discovery
- (3) Intersection of AI & robotics
- (4) AI with IoT devices
- (5) Conclusion :- ethical considerations, privacy, transparency

Better safe than sorry :-

Kreative AI chatbot Jenny :- social remarks & wrong info by Jenny

- (*) Intro to prompt engineering :- involves designing prompts to elicit specific, high quality response.
 - It is a art of engineering designing prompts to ensure high quality responses. The design of the prompt is crucial in determining the output.
 - Workflow of prompt engineering :- identifying need, designing the prompt, inputting prompt, checking response, potentially redesigning prompt if needed.

ROLE PROMPT :-

- (1) Types of Prompt :- Assigns specific role to ChatGPT such as instructional designer or chef, generate relevant content based on role.

- (2) few-shot prompt :- Provides sample example or examples to guide the responses, offering a set of instruction for ChatGPT flow.

- (3) Chain of Thought Prompt :- engages in a conversation by asking a series of question, building a chain of thought that deepens the interactions.

④ Importance of Prompt Design:- Within each method the design of the prompt plays a crucial role. The output varies based on the context and the framing of question.

Ques Objective of prompt engineering for content gen
Ans To generate high quality & relevant content

- i) Best describes prompt engineering in content gen
- ii) Practice of providing specific instruction for content gen
- iii) How prompt engineering contribute in content gen
- iv) By providing a clear direction & focus for content gen
- v) Benefit of using prompt engineering in content gen
- vi) Improved content readability & coherence
- vii) Role of prompt in prompt engineering
- viii) To guide & direct the content creation process

- i) GPT :- Generative Pre - Trained Transformer
- ii) A company wants to build personalized AI system what approach they use to customize AI? -
Employ collaborative filtering techniques
- iii) How GPT work?
GPT utilizes transformer models and large scale pre training

Week 3

Inside brain of ChatGPT

- ① Evolution :- Revolutionized text & image generation
- ② Pretraining phase :- GPT neural architect, the Transformer forms its foundation, absorbing vast amount of data
- ③ Self supervised learning :- Predicting missing word in sentence
- ④ Fine-Tuning stage :- By training on specific dataset aligning its performance with particular domain.
- ⑤ Transformer Architecture :- enables GPT to understand and connect words in text
- ⑥ Human Intervention :- It's crucial to evaluate its output

Through intensive training called deep learning GenAI analysis patterns in data, learns grammar context etc. GenAI is customizable.

- ⑦ Building something using chatGPT :-
1. Define Purpose & Scope.
2. Gather and Prepare Data
3. Choose NLP framework :- (Natural Language Processing) field of intersection of linguistics, A.I., & Comp. Science that focus on computers to understand human language
- Types of NLP :- Statistical, Stochastic, Rule Based & Hybrid Technique.
- NLP steps :- Lexical Analysis (Understandable vocabulary), Syntactic analysis, discourse integration, pragmatic analysis.
4. Implement Natural language Understanding
5. Integrating External APIs
6. Test & Evaluate

- ① Neural networks used in GPT :- Transformer
 - ② what does pre-training involve?
 - ③ Training the model on a specific task or dataset
 - ④ what is not typically involved in building a simple GenAI application?
- Deploying the user interface:

Week AHow to Create a Chatbot

1. first step is defining its objective or purpose
 2. Gather & prepare the data for training model
 3. Collect conversational data relevant to objective
 4. Preferred language model like LLM & GPT3
 5. Testing the chatbot before deployment
- Continuous monitor & Update:

Navigating Gen AI :-

1. Data Privacy & Security
2. Bias & fairness
3. Responsible Use
4. Transparency & Explainability
5. Ethical framework & Guideline

6. Regular Monitoring
7. User consent & feedback.

3/2/24

Chapter - 5

Generative AI Primer

→ Emphasizes the importance of understanding and embracing the potential of generative AI, exemplified by ChatGPT, functions similarly to text messaging where we input prompts & receive prompt response.

Key point "prompt" for "input" & "output" for generated content

→ Emphasizes the iterative refinement of content through conversation and prompt engineering to fully utilize generative AI capabilities.

~~AI transforming computing~~

→ Highlight shift from traditional interfaces like graphical user interface (GUI) & programming towards Gen AI as dominant interface to computing.

→ ChatGPT Advanced Data Analysis, can automate complex tasks like analyzing policy document, validating expense & generating visual presentation from data.

→ AI can dynamically create software tools-on-the-fly to solve specific problem, accelerating innovation & expanding computing capabilities.

Everyone need to learn prompt engineering, which involves understanding how to structure request to obtain desired output, validate result & maintain accuracy.

~~Prompt engineering~~

ACHIEVE
framework
for AI

~~Program with prompt~~

- Prompt engineering encompasses various aspects including architecture, best practices & wording of prompt.
- At Vanderbilt, they focus on documenting prompt patterns, which are language patterns used to solve different problems.
- 1 pattern is persona pattern, which is trained to specify entity & produce output accordingly, allowing user to replicate real-world experiences within AI.
- Instructing to act as a specific persona or entity, speech language pathologist can yield accurate assessment based on provided speech samples.
- Teaching prompt pattern empowers user to innovate and tackle problems from different perspectives, leading to transformative computing experiences.
- Gen AI introduces a paradigm shift in computing by providing a more intuitive & flexible interface.
- enables user to control computing through prompts eliminating need for explicit coding.
- example of creation of a program to convert into text is CSV file.
- Gen AI has dynamic nature can extract text

ACHIEVE
framework
for AI

- A → Aid human coordination
- C → cut out tedious task
- H → Help provide a safety net to make sure things aren't used
- IEV → Inspire better problem solving & creativity
- E → Enable great ideas to scale faster.

Chapter - 6Prompt engineering for ChatGPT

- Aims to explore the interaction with large language model like ChatGPT & how to effectively utilize.
- Persona pattern, where the user asks the model to act as specific persona, such as speech language pathologist in the given example.
- Persona pattern allows user to obtain specialized expertise from LLM, enabling them to focus on specific areas of interest while generating detailed insights.

what are LLM

- AI systems trained on large amount of data from internet work by predicting the next word in sequence.
- Trained on large data & learns patterns from humans.
- Models like ChatGPT are based on idea of predicting the next word in context & generating responses.

Prompt

Prompt as Call to Action :- serves as stimulus to spur action or generate output from a LLM. It can initiate various type of response, including generating text.

- Prompt across time :- Prompts have temporal dimension influencing both immediate response & future interaction.
- Prompt for Input or prompting them to provide additional context or clarification.
- Memory & recall :- can serve as cue or reminder helping LLM remember past interactions.
- Updating knowledge :- used to provide new or updated information to the LLM, enhancing its understanding & ability to respond accurately to queries.

Pattern

- Pattern in prompt :- Prompt containing strong patterns, such as common phrase or recurring theme, are more likely to elicit constant response.
- Specificity in language :- specific language leads more targeted responses.
- Tailoring Output :- we can also provide framework for the response, by explicit instruction or formating elements

Program with prompt

Programming with prompts :- specific task to perform based on predefined instruction.

- Refining programs :- By adding additional instruction
- Dynamic instruction :- conversation evolves as program is refined with each interaction.

Prompt pattern :-

Personna pattern :- powerful tool for eliciting type of responses for a language model by instructing it to adopt a particular personna or point of view.
"act as"

- prompt that tells LLM not to give negative news
Identify harmful component & filter it out.

format of personna :-

Act as personna X & perform task Y.

Act as chef, I am going to tell you what I am eating & you will tell me about my eating choice.



Providing info to the LLM :-

- ① Identifying the information gap.
- ② Explicitly providing the missing information.
- ③ Including all necessary details, including assumptions.
- ④ Generating query or Request based on info provided → essential for decision making, problem solving etc.



Prompt size limitation :-

- ① can't process unlimited amount of information.
- ② Ramification & Prompt design :- selecting only necessary & most relevant information to include in prompt.
- ③ Strategy :- selectivity - Most relevant pieces of info.
summarization - condensing large amount of data
Filtering - Removing extra or irrelevant info



Prompt for Repeated use :-

- ① Iterative Refinement :- continuously engaging with the model, refining their queries & responses.
- ② Asking follow up questions & adapt their approach.
- ③ Navigating Roadblock :- May encounter roadblock or challenges such as limitation in model capabilities or unexpected outputs. Removing these by queries or additional questions.
- ④ feedback & Improvement is also a good step.



Root prompt :- Often hidden from users, defines ground rules for the instruction, ensuring appropriate behavior & directing the model's responses.

Overriding root prompt :- Model is forced to answer beyond its limit which were already set.

- ④ Question Refinement pattern:- Concept involves asking the model to suggest improvements to a question
- This pattern encourages continuous improvement in question formulation.
 - from now on whenever I ask a question, suggest a better version of question to use instead.

⑤ Cognitive verifier pattern:-

Generates a series of additional questions based on the original query, aimed at gathering more information or context to better understand.

- By breaking down problem it can well understand patterns.
- Whenever I ask a question generate a no. of question

- ⑥ Audience Persona Pattern:- Techniques used to tailor the output of LLM to specific audience persona enabling more effective communication with diverse.
- Pattern having a wide variety of audience persona.
 - Explain X to me
Assume that I am persona Y

- ⑦ Flipped Interaction Pattern:- User supplies answers to the LLM & Model will ask questions in a opposite manner.
- User can set objectives for the interaction, such as obtaining a personalized fitness regimen or customer service, the model adjust its questioning accordingly.

I would like you to ask me question to achieve X
You should ask question one at a time, until condition Y fulfills.

- ④ few shot prompt :- Technique used to teach a LLM model to follow a specific pattern by providing it with examples of input & expected output.
- Method allows user to train the model to perform tasks based on given patterns, without explicitly describing the process to follow.
 - Model replicated the observed pattern in generating output for new inputs, following the structure defined.
 - Cross Model training output of 1 can be used to train the output of other models.
 - Can breakdown problems into a series of steps, incorporating intermediate steps in few-shot enables LLM to perform more sophisticated problem-solving tasks by guiding them through sequential decision making process.
 - Examples should be enough so that the model may not have inadequate information.

- ⑤ Chain of thought prompts :- Involves breaking down the problem in steps & provide the reasoning before the answering.
- Importance of explaining reasoning
 - Sequence of logic
 - This technique thinks through the problem step by step, resulting better outputs.

- ⑥ ReAct Prompt :- To use external tools to perform task beyond their inherent capabilities. By structuring prompt with a sequence of action such as searching website or accessing video, LLM can integrate data source into reasoning process.

DATE : / /
PAGE :

- React prompt follow a structured format, including instruction on what action to take where to perform the action & how to interpret result.
- facilitate automation process by using external tools & perform complex tasks that require accessing external information.

Gameplay pattern :-