Training Day 2 Report

Date: June 24, 2025

Generative AI or Gen AI is a type of artificial intelligence that can generate new content (images, text, videos, codes etc.).

In previous era, we were solely focusing on classifications based on predictions.

Gen AI Tools:

ChatGPT	Gemini
DeepSeek	Dall e
Copilot	Grok

Internet Architecture:

Input → Processor → Output

Generative AI Architecture:

Prompt → Model → Generated Content

LLM (Large Language Model): It is a type of artificial intelligence (AI) program that can understand and generate human-like language (like English, Hindi, etc.). It is trained on massive amounts of text data (books, websites, articles) to answer questions, write essays, translate languages, chat with people, and more.

AI Models:

Every AI Models are based on LLM.

- GPT 3.5/GPT 4.0
- Gemini 1.5
- Claude 3

- LLaMA
- Whisper
- Codex

LLM (Large Language Model)

LLM is a type of AI trained to understand and generate human like text.

It uses vast datasets (books, websites and conversations) to learn patterns in language.

Think of it like a super charged auto complete that understands context deeply.

Eg. Chatgpt, google-gemini, Claude, LLaMA.

Key terms of LLM

Token – Smallest unit (word or word part).

Parameter – Adjustable part of model (like a brain cell).

Prompt – The input or question you give the model.

Fine tuning – customizing a model on specific data.

Inference – The models response or output.

Evolution of language models

Year	Model	Parameter	Creators
2018	GPT 1	117M	Open AI
2019	BERT	110M	Google
2020	GPT 3	175B	Open AI
2023	LLaMA , Claude	~70B-100B	Meta, anthropic
2024	GPT 4o,Gemini	~200B	Open AI, Google

Eg. You typed a prompt: “What is the capital of France?”

Model tokenizes your prompt: [“What” , “is” , “the” , “capital” , “of”’ , “France” , “?”]

The transformer processes it using layers of attention.

It predicts the last next token – “Paris”.

Output: A smart context aware response.

Inside the LLM – Transformers

Text data ----Tokenizer----Language Model-----output

Self-attention (model finds which word relate to each other)

Feed Forward Network (learns deeper features)

Positional Encoding (remember word order)

Training LLMS (behind the scenes)

- Pretraining: read tons of text -----predict the next word
- Fine tuning: refine on specialized data
- RLHF (reinforcement learning with human feedback): people keep it learn better responses.

Eg: Teaching a parrot basic words----later refining to speak in sentences.

Applications

- Chabot's
- Education
- Healthcare
- Legal
- Writing

Limitations

- Hallucination
- Bias
- No real understanding
- Context length

Ethical concerns

- Misinformation generation
- Data privacy

- Deep fakes
- AI responsibility

Future of LLMs

Multimodal models

Autonomous agents

Smaller + fast open source models on devices LLMs