DAY - 2

INTRODUCTION TO GENERATIVE AI

On Day 2 of the training, we explored the concept of Generative Artificial Intelligence (Generative AI). We learned that Generative AI refers to a class of models that can create new content—such as text, images, audio, code, and more—based on the data they were trained on. These models are capable of learning patterns and generating human-like outputs, making them powerful tools for creativity, automation, and productivity.

Examples of generative outputs include:

- Text (e.g., essays, code, scripts)
- Images (e.g., artworks, visual designs)
- Audio (e.g., synthetic voices, music)
- Video (e.g., animation, scene generation)

UNDERSTANDING LLMs, DIFFUSION MODELS & MULTIMODAL MODELS

We were introduced to the basic architecture and function of the following foundational model types:

- **Large Language Models (LLMs):** These are deep learning models trained on massive amounts of text data to understand and generate human language. Examples include GPT (Generative Pre-trained Transformer), Claude, Gemini, and LLaMA.
- **Diffusion Models:** These are mainly used for image generation tasks. They start with random noise and gradually transform it into a coherent image. DALL·E by OpenAI uses similar mechanisms to create realistic images from text prompts.
- **Multimodal Models:** These models can process and combine multiple types of data inputs—such as text, images, and audio. Gemini and GPT-4o (Omni) are examples of such models, as they can handle text and visual inputs simultaneously.

TOOLS & PLATFORMS INTRODUCED

We explored several cutting-edge tools and platforms powered by generative AI. Each of these tools serves a specific purpose:

a. ChatGPT (OpenAI)

- Mainly used for document generation, summaries, ideation, and conversational AI.

- Model in use: GPT-4o (Omni) – a powerful, multimodal model capable of understanding and generating both text and images, with improved speed and cost-efficiency.

b. Gemini (by Google DeepMind)

- A strong LLM with capabilities in reasoning, coding, and multimodal input.
- Model in use: Gemini 1.5 Flash – optimized for speed and cost, suitable for prompt engineering and real-time responses.

c. DALL·E

- Specialized in image generation from textual descriptions.
- Also supports inpainting (editing parts of an existing image).

d. Co-Pilot (GitHub + OpenAI)

- AI assistant designed for code suggestions and completion.
- Integrated into development environments like VS Code for real-time help.

e. CodeX

- Developed by OpenAI for code generation and natural language-to-code transformation.
- Ideal for automating repetitive coding tasks.

f. Claude (Anthropic)

- A conversational AI known for safe, ethical, and nuanced responses.
- Built for high-quality reasoning, summarization, and assistance.

g. LLaMA (Meta AI)

- Lightweight open-source models for various tasks, suitable for local deployment.
- Popular in academic research and customized AI applications.

h. Whisper & Google Speech Recognition

- Whisper (OpenAI): A speech-to-text model capable of transcribing audio with high accuracy.
- Google Speech-to-Text: A robust, cloud-based API used for real-time voice recognition and transcription tasks.

CONCLUSION

Day 2 provided a well-rounded overview of foundational AI models and the tools built upon them. We not only learned how different architectures like LLMs and diffusion models function, but also gained hands-on awareness of the tools that leverage these technologies in the real world.

The exposure to models such as GPT-4o, Gemini 1.5 Flash, and tools like DALL·E and Whisper gives us a strong starting point for building creative, technical, and practical applications powered by generative AI.