Generative AI and Beyond

Unit 3 - Summary

- **Generative AI:** Generative AI is a subset of artificial intelligence that uses advanced techniques to create human-like content by learning patterns from the input data.

- **Large Language Models:** Large Language Models (LLMs) are utilized in Generative AI to generate text based on input prompts.

- **Foundation Models:** Foundation models are the basis for the creation and evolution of Generative AI systems, trained on massive unlabeled datasets.

- **Enterprise Challenges:** Generative AI can be used to solve enterprise challenges by creating new data based on patterns and trends learned from training data.

- **Beyond Generative AI:**
  - **Agentic AI**: AI systems that can act independently, with memory and reasoning capabilities, interacting with other agents.
  - **Physical AI**: AI systems integrated with physical systems, such as robots, enabling them to perceive, learn, and interact autonomously in real-world environments.
  - **AI Factories**: AI factories are designed to handle the increasing demands for compute resources as AI models scale up in complexity.

What are Foundation Models?

Foundation models are the basis for the creation and evolution of Generative AI systems. They consist of AI neural networks trained on massive unlabeled datasets, generally with unsupervised learning.

What are Large Language Models (LLMs)?

LLMs are utilized in Generative AI to generate text based on input prompts. They use the transformer architecture to establish relationships between words.

How are GPUs used in Generative AI?

GPUs train and run large models in Generative AI. They provide performance speed-up, efficiency, and scalability.

Why are AI Factories important as AI evolves into agentic and physical systems?

AI Factories provide integrated infrastructure for autonomous AI systems, enabling scalable enterprise intelligence and addressing challenges such as design complexity, deployment, cost, and time to value.