

Subject Code	Subject Name	Credits
26CS508	GENERATIVE AI AND LARGE LANGUAGE MODELS	4

**Course Objectives:**

1. To introduce the fundamentals of Generative AI and deep learning-based models.
2. To understand the architecture and working of Large Language Models (LLMs).
3. To explore prompt engineering and fine-tuning techniques.
4. To study applications of generative models in real-world scenarios.

**Learning Objectives:**

After completion of the course the students will be able to:

1. Understand core concepts of Generative AI and probabilistic modeling.
2. Explain transformer architecture and LLM working principles.
3. Apply prompt engineering techniques effectively
4. Develop and fine-tune generative AI models.

**Unit I - Introduction to generative AI (12 Hrs.)**

Overview of AI, ML, and Deep Learning - Generative AI vs Discriminative Models - Probabilistic models and generative processes - Types of Generative Models - Applications of Generative AI - Limitations and challenges

**Unit II - Foundations of large language models (12 Hrs.)**

Natural Language Processing - Language modeling and probability distributions - N-grams and statistical language models - Neural Language Models - Word embeddings: Word2Vec, GloVe - Contextual embeddings and sequence modeling

**Unit III - Transformer architecture and llms (12 Hrs.)**

Transformer architecture: encoder-decoder model - Self-attention and multi-head attention mechanism - Positional encoding - Pre-training and fine-tuning paradigms - Overview of popular LLMs (GPT, BERT, etc.)

**UNIT IV - Prompt engineering and model customization (12 Hrs.)**

Prompt design principles and techniques - Zero-shot, one-shot, and few-shot learning - Chain-of-thought prompting - Fine-tuning LLMs - Parameter-efficient tuning (LORA, adapters) - Evaluation metrics for generative models - Tools and frameworks (Hugging Face, OpenAI APIs) - Deployment considerations

**UNIT V - Advanced techniques in large language models (12 Hrs.)**

Scaling laws in LLMs - Pre-training and dataset construction - Fine-tuning and RLHF - Retrieval-Augmented Generation (RAG) architecture and pipelines - Model optimization - Inference and deployment - Evaluation benchmarks: BLEU, ROUGE, perplexity, and human evaluation

**Reference Books:**

1. Ian Goodfellow, Yoshua Bengio, Aaron Courville — Deep Learning, MIT Press, 2016.
2. Generative AI with Python and TensorFlow 2 – Joseph Babcock, Raghav Bali, 2021.
3. Jacob Eisenstein — Introduction to Natural Language Processing, 2019.
4. Yoav Goldberg — Neural Network Methods for NLP, 2017.
5. Lewis Tunstall et al. — Natural Language Processing with Transformers, O'Reilly, 2022.
6. Sebastian Raschka — Machine Learning with PyTorch and Scikit-Learn, 2022.
7. Denis Rothman — Transformers for Natural Language Processing, 2021.
8. Tom Taulli — Generative AI: How ChatGPT and Other AI Tools Will Revolutionize Business, 2023.