Course Title: Generative AI: Math Foundations to Multimodal Modeling

Course Level: Intermediate, Advanced

Course Description: This course provides a comprehensive introduction to the exciting field of generative AI. Students will learn the fundamental concepts, explore various generative models, and gain hands-on experience in developing and applying these models to real-world problems.

Prerequisites:

- Fluency with Python programming
- Basic knowledge of machine learning concepts
- Basic understanding of probability and statistics

Course Objectives: Upon successful completion of this course, students will be able to:

- Understand the fundamental concepts and principles of generative AI.
- Grasp the core concepts of machine learning relevant to generative modeling.
- Explain the differences between various generative models (GANs, VAEs, diffusion models, etc.).
- Implement and train generative models using popular frameworks (e.g., TensorFlow, PyTorch).
- Apply generative AI techniques to solve problems in different domains (e.g., image generation, text generation, music composition).
- Evaluate the performance of generative models and understand their limitations.
- Discuss the ethical implications and societal impact of generative AI.

Course Outline:

Module 1: Mathematical Foundations for Machine Learning

- Linear Algebra
    - Vectors, matrices, eigenvalues, eigenvectors, matrix operations
    - Applications in machine learning (e.g., dimensionality reduction)
- Calculus
    - Derivatives, gradients, chain rule, optimization
    - Applications in machine learning (e.g., gradient descent)
- Probability and Statistics
    - Probability distributions, sampling, Bayes' theorem
    - Statistical inference, hypothesis testing
- Information Theory
    - Entropy, cross-entropy, KL divergence
    - Applications in generative modeling (e.g., VAE loss functions)

Module 2: Machine Learning Fundamentals

- Supervised Learning
    - Classification, regression, model evaluation
    - Common algorithms (e.g., linear regression, logistic regression, decision trees)
- Unsupervised Learning
    - Clustering, dimensionality reduction
    - Common algorithms (e.g., k-means, PCA)

- Deep Learning Basics
    - Neural networks, activation functions, backpropagation
    - Deep learning frameworks (TensorFlow/PyTorch)
    - Building Blocks of Neural Networks
        - Forward propagation: How data flows through a neural network.
        - Loss functions: Measuring the error of a model.
        - Optimization algorithms: Gradient descent and its variants.
        - Backpropagation: Calculating gradients to update network weights.
    - Convolutional Neural Networks (CNNs)
        - Image data and computer vision tasks.
        - CNN architecture: Convolutional layers, pooling layers, feature maps.
        - Applications: Image classification (e.g., MNIST digit recognition).
        - Data preprocessing: Image augmentation, normalization.
        - Training process: Epochs, batches, learning rate.
        - Evaluation metrics: Accuracy, precision, recall.
        - Overfitting and regularization techniques.
    - Recurrent Neural Networks (RNNs) & Long Short-Term Memory (LSTM) Networks
        - Sequential data and natural language processing (NLP).
        - RNN architecture: Hidden state, recurrent connections.
        - Applications: Text classification (e.g., sentiment analysis).
        - Vanishing gradient problem in RNNs.
        - LSTM architecture: Gates and memory cells.
        - Applications: Sequence prediction, language modeling.
    - Word Embeddings and Text Processing
        - Representing words as vectors: Word2Vec, GloVe.
        - Text preprocessing: Tokenization, stemming, lemmatization.
        - Applications: Text classification with word embeddings.
- Introduction to Transformers
    - Attention mechanism: Focusing on important parts of the input.
    - Transformer architecture: Encoder and decoder.
    - GPT Models: Generative pre-training for text generation (e.g., story writing, code generation).
    - BERT Models: Bidirectional encoding for language understanding (e.g., sentiment analysis, question answering).
    - Applications: Machine translation, text summarization.

Module 3: Introduction to Generative AI

- What is Generative AI?
    - Definition, history, and key concepts

    - Distinguishing generative models from discriminative models
    - Applications and impact of generative AI
- Generative Adversarial Networks (GANs)
  - Introduction to GANs
    - Architecture and training process (generator and discriminator)
    - Loss functions and optimization
  - Variations of GANs
    - DCGAN, CycleGAN, StyleGAN, Progressive Growing of GANs
  - Applications of GANs
    - Image generation, image-to-image translation, style transfer
- Variational Autoencoders (VAEs)
  - Introduction to VAEs
    - Encoder-decoder architecture, latent space, variational inference
    - Loss functions and optimization
  - Applications of VAEs
    - Image generation, representation learning, anomaly detection
- Diffusion Models
  - Introduction to Diffusion Models
    - Forward and reverse diffusion processes, score matching
    - Training and sampling from diffusion models
  - Applications of Diffusion Models
    - Image generation, image editing, super-resolution

Module 4: Multi-modal Modeling

- Definition, motivation, and applications (e.g., image captioning, text-to-image generation, video generation with audio)
- Challenges and opportunities in multimodal generation
- Fusion Mechanisms for Generative Models
  - Concatenation: Combining encoded features
  - Attention Mechanisms: Dynamically focusing on relevant modalities
  - Gated Multimodal Units (GMUs): Selectively combining information
  - Transformer-based Fusion: Using transformers to fuse multimodal sequences
- Multimodal Generative Models
  - Generative Adversarial Networks (GANs) for multimodal generation (e.g., image-to-image translation with text prompts)
  - Variational Autoencoders (VAEs) for multimodal data generation and representation learning
  - Diffusion Models for generating high-quality multimodal content
- Applications and Case Studies
  - Image captioning and storytelling
  - Text-to-image synthesis (e.g., DALL-E 2)
  - Generating music with accompanying visuals
  - Multimodal machine translation

Module 5: Other Generative Models and Advanced Topics

- Autoregressive models (e.g., PixelCNN, WaveNet)
- Flow-based models (e.g., RealNVP, Glow)
- Generative models for text (e.g., RNNs, Transformers)
- Generative models for music and audio
- Ethical considerations in generative AI
- Trends and future directions in generative AI

Assessment:

- Assignments (practical implementation and analysis of generative models)
- Quizzes (testing understanding of key concepts)
- Final Project (developing a generative AI application)
- Class participation

Software and Tools:

- Python programming language
- Deep learning libraries (TensorFlow, PyTorch)
- Jupyter Notebook or Google Colab
- Cloud computing platforms (optional, e.g., AWS, Google Cloud)

Recommended Books:

- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.
- Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. (2017). The Elements of Statistical Learning. Springer.