Training Course on Multimodal Generative Models

Professional Development Training Course Outline

Category	Data Science	**Duration**	5 Days
Base Fee	$1,500 USD	**Delivery Mode**	Online / On-site Hybrid

Course Introduction

This intensive training course delves into the cutting-edge domain of **Multimodal Generative Models**, empowering participants to leverage the synergistic power of diverse data types for sophisticated content creation. As **Artificial Intelligence** rapidly evolves, the ability to seamlessly integrate and generate across text, image, audio, and other modalities is becoming paramount for **innovation** and competitive advantage in various industries. This program provides a comprehensive understanding of foundational concepts, advanced techniques, and practical applications in this exciting field.

Training Course on Multimodal Generative Models: Integrating Text, Image, and Other Data Types for Generation is designed to equip professionals with the **practical skills** and **theoretical knowledge** necessary to build, train, and deploy **state-of-the-art** multimodal generative AI systems. From understanding **Transformer architectures** and **Diffusion Models** to mastering **prompt engineering** for cross-modal generation, participants will gain hands-on experience with leading frameworks and tools. This course is essential for anyone looking to push the boundaries of **AI-driven creativity**, enhance **data understanding**, and unlock new possibilities in **human-computer interaction**

Programme Curriculum

Training Course on Multimodal Generative Models: Integrating Text, Image, and Other Data Types for Generation

Introduction

This intensive training course delves into the cutting-edge domain of **Multimodal Generative Models**, empowering participants to leverage the synergistic power of diverse data types for sophisticated content creation. As **Artificial Intelligence** rapidly evolves, the ability to seamlessly integrate and generate across text, image, audio, and other modalities is becoming paramount for **innovation** and competitive advantage in various industries. This program provides a comprehensive understanding of foundational concepts, advanced techniques, and practical

applications in this exciting field.

Training Course on Multimodal Generative Models: Integrating Text, Image, and Other Data Types for Generation is designed to equip professionals with the **practical skills** and **theoretical knowledge** necessary to build, train, and deploy **state-of-the-art** multimodal generative AI systems. From understanding **Transformer architectures** and **Diffusion Models** to mastering **prompt engineering** for cross-modal generation, participants will gain hands-on experience with leading frameworks and tools. This course is essential for anyone looking to push the boundaries of **AI-driven creativity**, enhance **data understanding**, and unlock new possibilities in **human-computer interaction**.

Course Duration

5 days

Course Objectives

1. Master the fundamental concepts of Multimodal AI and its distinct advantages over unimodal systems.
2. Understand and implement various data fusion techniques (early, late, intermediate) for integrating diverse data streams.
3. Explore and apply cutting-edge Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) for cross-modal content generation.
4. Gain proficiency in Diffusion Models and their role in high-quality image and audio synthesis.
5. Develop expertise in Transformer architectures and Attention Mechanisms for handling sequential and relational multimodal data.
6. Learn advanced Prompt Engineering strategies for effective and controlled multimodal content creation.
7. Implement and fine-tune pre-trained Large Multimodal Models (LMMs) for specific industry applications.
8. Evaluate the performance of multimodal generative models using relevant metrics like FID, Inception Score, and BLEU.
9. Address critical AI Ethics, bias mitigation, and trustworthy AI considerations in multimodal system design.
10. Design and build innovative multimodal AI applications for real-world problem-solving.
11. Understand the principles of Multimodal Retrieval-Augmented Generation (MM-RAG) for enhanced contextual understanding.
12. Explore the integration of Multimodal Generative AI into AI Agents and conversational systems.
13. Prepare for the future of AI by understanding emerging trends and research directions in multimodal generation.

Organizational Benefits

- Empower teams to rapidly develop and deploy novel AI solutions by leveraging multimodal data.
- Improve decision-making and insights through a more comprehensive interpretation of complex, multi-source data.
- Automate content creation, analysis, and interaction across various modalities, reducing manual effort.

- Stay ahead in the rapidly evolving AI landscape by adopting cutting-edge multimodal capabilities.
- Create more natural, intuitive, and engaging human-computer interactions through multimodal interfaces.
- Unlock opportunities for entirely new products and services driven by sophisticated multimodal AI.
- Develop more robust and resilient AI models by reducing reliance on single-modality data.

Target Audience

1. AI Engineers & Machine Learning Practitioners
2. Data Scientists
3. Researchers & Academics.
4. Software Developers
5. Product Managers.
6. Content Creators & Designers.
7. Researchers in NLP & Computer Vision.
8. Business Leaders & Strategists.

Course Outline

Module 1: Foundations of Multimodal Generative AI

- Introduction to Multimodal AI: Definition, evolution, and significance in the era of Generative AI.
- Understanding Different Modalities: Text, image, audio, video, and structured data types.
- Key Concepts: Data Fusion (early, late, hybrid), cross-modal embeddings, and joint representations.
- Challenges in Multimodal Learning: Data alignment, semantic gap, and computational complexity.
- **Case Study:** Analyzing how multimodal AI enhances medical diagnostics by integrating patient records, imaging scans, and clinical notes for more accurate disease prediction.

Module 2: Generative Models for Individual Modalities

- Revisiting Core Generative Models: GANs, VAEs, and their strengths and limitations for unimodal data.
- Deep Dive into Diffusion Models: Theory, architecture, and their role in high-fidelity image and audio generation.
- Transformer Architectures: Self-attention, encoders, and decoders for sequence modeling in text and other data.
- Pre-trained Models and Transfer Learning: Leveraging existing models for faster development.
- **Case Study:** Utilizing Stable Diffusion for generating photorealistic images from text descriptions for marketing campaigns.

Module 3: Integrating Text and Image for Generation

- Text-to-Image Synthesis: CLIP, DALL-E, Midjourney, and other state-of-the-art models.
- Image-to-Text Generation: Image captioning, visual question answering, and multimodal reasoning.
- Conditional Generation: Controlling output generation with specific textual or visual prompts.

- Attention Mechanisms in Multimodal Models: How models focus on relevant information across modalities.
- **Case Study:** Developing an AI system for e-commerce that generates product descriptions and marketing copy directly from product images and sparse tags.

Module 4: Audio and Video Integration in Generative Models

- Speech Synthesis (Text-to-Speech): Generating natural-sounding speech from text.
- Speech Recognition (Speech-to-Text): Transcribing spoken language into text for multimodal input.
- Video Generation and Manipulation: Creating realistic videos from text, images, or audio.
- Audio-Visual Co-learning: Training models to understand relationships between sound and visual elements.
- **Case Study:** Building an AI-powered content creation tool that generates short marketing videos with synchronized voiceovers from a text script and a few input images.

Module 5: Multimodal Fusion and Representation Learning

- Advanced Data Fusion Strategies: Exploring different fusion points and architectures.
- Joint Embedding Spaces: Learning shared representations for different modalities.
- Cross-Modal Alignment: Techniques for aligning disparate data types (e.g., temporal alignment in video and audio).
- Multimodal Representation Learning: Autoencoders, Siamese networks, and other approaches.
- **Case Study:** Creating a personalized recommendation system for movies that fuses user preferences (text), movie trailers (video), and genre tags (structured data) to suggest highly relevant content.

Module 6: Advanced Multimodal Generative Architectures

- Large Multimodal Models (LMMs): Understanding the scale and capabilities of models like GPT-4V, Gemini.
- Multimodal Transformers: Extending the Transformer architecture for diverse input types.
- Mixture of Experts (MoE) in Multimodal Contexts: Enhancing model efficiency and performance.
- Generative Flow Models and Normalizing Flows for multimodal data.
- **Case Study:** Leveraging a large multimodal model to answer complex scientific questions by interpreting research papers (text), diagrams (image), and experimental data (tables).

Module 7: Prompt Engineering and Control in Multimodal Generation

- Fundamentals of Prompt Engineering: Crafting effective prompts for desired outputs.
- Multimodal Prompting Techniques: Incorporating visual, auditory, and textual cues in prompts.
- Controllable Generation: Guiding the generative process for specific styles, attributes, or narratives.
- Iterative Prompt Refinement and Feedback Loops for optimizing output quality.
- **Case Study:** Using advanced prompt engineering to generate variations of a product design (image) based on customer feedback (text) and desired material properties (structured data).

Module 8: Ethical Considerations, Evaluation, and Deployment

- Ethical AI in Multimodal Generative Models: Bias, fairness, misinformation, and copyright issues.
- Trustworthy AI Principles: Transparency, explainability, and accountability in multimodal systems.
- Evaluation Metrics for Multimodal Generation: FID, Inception Score, Perplexity, CLIP Score, and human evaluation.
- Deployment Strategies: MLOps for multimodal models, cloud platforms, and API integration.
- **Case Study:** Implementing robust safety filters and bias detection mechanisms in a multimodal AI system designed for generating educational content to ensure responsible and equitable outputs.

Training Methodology

This training course employs a dynamic and interactive methodology to ensure maximum learning and practical skill development.

- **Hands-on Labs & Coding Sessions:** Extensive practical exercises using Python, TensorFlow, PyTorch, and popular generative AI frameworks (e.g., Hugging Face Transformers).
- **Interactive Lectures:** Engaging presentations covering theoretical foundations, model architectures, and key concepts.
- **Real-world Case Studies:** In-depth analysis of successful multimodal AI applications across various industries.
- **Live Demonstrations:** Showcasing the capabilities of state-of-the-art multimodal generative models.
- **Group Discussions & Collaborative Projects:** Fostering peer learning and problem-solving.
- **Expert-led Q&A Sessions:** Opportunity to clarify doubts and gain insights from experienced practitioners.
- **Continuous Assessment:** Quizzes, coding challenges, and a final project to reinforce learning and evaluate progress.

Register as a group from 3 participants for a Discount

Send us an email: info@fineskilltrainingcenter.org or call *+254769199797*

Certification

Upon successful completion of this training, participants will be issued with a globally- recognized certificate.

Tailor-Made Course

We also offer tailor-made courses based on your needs.

Key Notes

a. The participant must be conversant with English.

b. Upon completion of training the participant will be issued with an Authorized Training Certificate

c. Course duration is flexible and the contents can be modified to fit any number of days.

d. The course fee includes facilitation training materials, 2 coffee breaks, buffet lunch and A Certificate upon successful completion of Training.

e. One-year post-training support Consultation and Coaching provided after the course.

f. Payment should be done at least a week before commence of the training, to Fineskill Training Center account, as indicated in the invoice so as to enable us prepare better for you.

Upcoming Scheduled Sessions

Start Date	End Date	Location	Price
23 Nov 2026	27 Nov 2026	Online	$1,000
23 Nov 2026	27 Nov 2026	Physical	$1,500
23 Nov 2026	27 Nov 2026	Physical	$3,000
23 Nov 2026	27 Nov 2026	Physical	$3,000
23 Nov 2026	27 Nov 2026	Physical	$6,000
23 Nov 2026	27 Nov 2026	Physical	$4,500
23 Nov 2026	27 Nov 2026	Physical	$6,000
23 Nov 2026	27 Nov 2026	Physical	$6,000

Ready to enroll or request a customized program? Book online or contact us:

Register Online Now

Email: info@fineskilltrainingcenter.com | Website: www.fineskilltrainingcenter.com