Advanced Techniques in Large Language Model Development and Deployment

Duration: 03 days

Module 01: Data Preprocessing for LLM Training (Theory + Labs)

- **Tokenization**: Breaks text into smaller units like words or subwords for model input.
- **Text Embedding**: Maps words or phrases to continuous vectors for semantic understanding.
- **Image Embedding**: Represents images as vectors to incorporate visual information.

Module 02: Responsible AI and GenAI (Theory & Use Case)

- **Fairness**: Mitigates bias in models to ensure equitable outcomes.
- **Transparency**: Makes AI decision-making understandable for users.
- **Accountability**: Establishes frameworks for AI system responsibility.

Module 03: Security Perspectives in GenAI (Theory + lab)

- **Challenges**: Identifies unique security risks in AI systems.
- **Measures**: Implements strategies like robustness training and privacy protection.

Module 04: Model Optimization Techniques for GenAI models (Theory + Lab)

- **Pruning**: Removes unnecessary parameters to reduce model size.
- **Quantization**: Reduces numerical precision for faster inference.
- **Compression**: Shrinks model size through techniques like weight sharing.

Module 05: Model Monitoring After Deployment (Theory + Use Cases)

- **Monitoring Systems**: Continuously checks model performance post-deployment.
- **Drift Detection**: Identifies changes in data distribution over time.
- **Issue Resolution**: Diagnoses and resolves issues in deployed models.

Module 06: Deployment on Cloud Platforms (Theory + Use Cases)

- **Azure Deployment**: Deploys models on Microsoft Azure cloud infrastructure.
- Topics may include setting up infrastructure and managing scalability.