Multimodal Deep Learning:

Document, Image, and Video Analysis

Duration: 05 days

Chapter 01: Fundamentals of Python and OCR

- Introduction to Python for Data Science
- Data Structures in Python
- Control Flow and Functions
- NumPy and Pandas for Data Manipulation
- Optical Character Recognition (OCR) Fundamentals
- Python Libraries for OCR (e.g., Tesseract, OpenCV)

Chapter 02: Fundamentals of Machine Learning

- Introduction to Machine Learning
- Supervised, Unsupervised, and Reinforcement Learning
- Model Training and Evaluation

Chapter 03: Fundamentals of Deep Learning

- Neural Networks and Deep Learning
- Activation Functions and Layers
- Loss Functions and Optimization
- Deep Learning Frameworks (e.g., TensorFlow, PyTorch)

Chapter 04: Understanding Deepfake with Keras

- Introduction to Keras
- Understanding Deepfakes
- Generative Adversarial Networks (GANs)
- Deep Convolutional GANs (DCGANs)
- Project: Implementing DCGAN

Chapter 05: Mastering OCR using Deep Learning and OpenCV Python

- Advanced Optical Character Recognition (OCR) Techniques
- Different Image Pre-processing techniques used in OCR pipeline
- Different Text Detection techniques used in OCR pipeline such as EAST and CTPN
- Different Text Recognition techniques used in OCR pipeline such as CRNN (CNN+RNN+CTC)
- Implementing OCR on real-life examples

Chapter 06: Building a PDF Knowledge Bot using LLMs

- Introduction to Language Models (LLMs)
- Benefits of using open-source LLMs
- Extracting text snippets from PDFs

- Creating a vector store using an embedding model
- Querying the LLM
- Creating a knowledge Bot using Azure OpenAI Studio (Only demo from trainer)