1 Week 1: Edge AI Fundamentals & ONNX Runtime

1.1 Session 1: Introduction to Edge AI & ONNX Ecosystem

- What is Edge AI and why it matters
- Challenges of deploying AI at the edge (compute, memory, power)
- Overview of edge deployment approaches
- Introduction to ONNX as a deployment solution
- ONNX Runtime and its execution providers
- The CPUExecutionProvider and its advantages
- Project kickoff: Select individual project from sample options (see Sample Projects)
- Setting up GitHub repository for course project (see GitHub Repository Requirements)

PROJECT MILESTONE - Phase 1 - Edge Deployment Planning & GitHub Setup: Identify an edge AI application scenario for your individual project from the sample project options. Define resource constraints, user requirements, and performance targets. Create a project plan that includes model selection, optimization approach, and deployment strategy. Initialize a GitHub repository with appropriate README, structure, and documentation templates as specified in the GitHub requirements.

1.2 Session 2: Converting & Preparing Models for Edge Deployment

- Popular model frameworks for edge-deployable AI (PyTorch, TensorFlow, JAX)
- Converting models from various frameworks to ONNX
- Navigating common conversion challenges for modern models
- Understanding ONNX model structure and operations
- Model profiling and performance benchmarking
- Initial resource utilization assessment
- Deployment environment considerations
- GitHub: Model conversion workflow documentation