Jassey-Jabarr Fisayo

ITAI 3377: A.I. at the Edge & IIOT

Professor: Patricia McManus

Reflective Journal: Conceptual AI Model Development and Deployment

Setup Process

I kicked off this project by setting up my environment on my laptop. I installed Python 3.9 and TensorFlow, which went smoothly after I made sure the versions were compatible. I used Jupyter Notebook since I find it easier to test code step-by-step. I loaded the MNIST dataset—70,000 images of handwritten digits—and built a neural network to classify them into 10 digits. After training, I converted the model to TensorFlow Lite, shrinking it from 100MB to 25MB, and deployed it on a simulated edge device. The deployment part involved setting up an input pipeline and running real-time tests, which I wasn't sure I'd get right at first.

My Observations

Training the model was quicker than I thought once I got the hang of the neural network setup. The architecture—28x28 grayscale inputs, hidden layers for pattern recognition, and a 10-digit output—worked well, hitting decent accuracy after a few runs. The optimization step was tough, though. Shrinking the model to 25MB made it faster for edge devices, but I had to tweak things a lot to keep accuracy from dropping too much. During deployment, real-time testing on the edge device went okay, but I ran into a few errors with input processing that took time to fix. In the end, the model handled digit recognition very well, which felt like a win.

Reflections

This project showed me how tricky edge AI can be, but also how useful it is. I struggled at first with understanding the neural network layers—I didn't get why each part mattered until I saw the patterns it picked up. Balancing model size, speed, and accuracy was a real challenge; I kept worrying I'd mess up the performance. Figuring out data preprocessing and the TFLite conversion process made me realize how much prep work goes into making AI practical for edge devices.

The most valuable insight was learning how theoretical concepts translate into practical implementation decisions, especially in resource-constrained environments. This knowledge will be crucial for future projects involving edge AI deployment.