Reflective Journal: Deploying a Simple AI Model on a Simulated Edge Device

**Setup Process**

I started this project using Google Colab since it’s straightforward and skips the hassle of local setup. I didn’t need to install Python or TensorFlow—they’re already there in Colab. I just opened a new notebook and ran the code to load the MNIST dataset and train a simple CNN. I installed the Edge Impulse CLI on my laptop by downloading Node.js and running `npm install -g edge-impulse-cli` in the terminal.

After training, I converted the model to TFLite in Colab, downloaded the `model.tflite` file, and uploaded it to Edge Impulse using their web interface—no terminal commands needed, just to try a different experience .

**My observations**

The model trained faster than I expected. I used only 3 epochs and reduced the filters to 16 in the Conv2D layer to keep things light. The screenshot I took shows the training output: accuracy hit 98.79% on the training set and 97.99% on the validation set by the last epoch, with a validation loss of 0.0610. Those numbers felt solid for a basic model.

I uploaded the model to Edge Impulse and ran inference in their simulator with some MNIST test samples. It predicted digits correctly most of the time, with an accuracy around 97%, and the latency was quick—about 20 milliseconds per inference. The Edge Impulse dashboard was easy to use, and I took a screenshot of the inference results to include in my documentation.

**Reflections**

This project made edge computing feel a lot more practical than I thought. Seeing the model run on a simulated edge device showed me how AI can work in real-time scenarios—like in a factory or on a smart device—without needing a big server.

I liked how Colab simplified the coding part, but I struggled a bit with Edge Impulse at first since I’d never used it before. Once I got the hang of it, though, it was pretty cool to see my model in action. I think edge AI is the future for a lot of industries, especially IIoT, where quick decisions are critical. Next time, I’d probably try a more complex dataset to push the model further, but for now, I’m happy with how this turned out.

I am real grateful for this experience, I wish I could observe the use IIOT in real world application around me.