Assignment 7

In this assignment you will host an LLM in your raspberry pi/jetson nano and provide replies

1. Download gemma-2B in llama format. Llama format is a binary format that self contains all the necessary data for the Large language model to work
2. Move the file to the raspberry pi/jetson nano
3. Start a chat with the LLM where the instructions are: "You are a teacher" and ask in the chat "Implement a simple neural network in pytorch"
4. Host a web server in the raspberry pi/jetson nano and use the context "You are an user" and the prompt "How are you?" using curl from your computer