Lab 14. Building a ChatBot With Finetuned Model

To build the ChatBot, we use the open-source Chainlit Python package.

Change to the ChatBot Directory

```
$ cd <Local Path>/ChatBot
```

Install the chainlit package via

```
$ pip install chainlit
```

Step 2: Run app code

The file app_own.py contains the UI code based. Open the file and change GPT2_CONFIG_124 Dictionary in app_own.py file for correct context length (based on model setting you have used in Lab 12). Also change the path of model directory if needed.

This file loads and uses the GPT-2 weights we generated in Lab 12.

Excecute the following command from the terminal to start the UI server:

```
chainlit run app_own.py
```

```
def get_model_and_tokenizer():
    """
    Code to load a GPT-2 model with pretrained weights generated in part !
    This requires that you run the code in part 5 first, which generates
    """

    GPT_CONFIG_124M = {
        "vocab_size": 50257,    # Vocabulary size
        "context_length": 128,  # Shortened context length (orig: 1024)
        "emb_dim": 768,         # Embedding dimension
        "n_heads": 12,          # Number of attention heads
        "n_layers": 12,         # Number of layers
        "drop_rate": 0.1,       # Dropout rate
        "qkv_bias": False       # Query-key-value bias
    }

    tokenizer = tiktoken.get_encoding("gpt2")

    model_path = Path(".") / ".." / "gpt2" / "model.pth"
    if not model_path.exists():
        print(f"Could not find the {model_path} file. Please run the part
        sys.exit()
```

Running commands above should open a new browser tab where you can interact with the model. If the browser tab does not open automatically, inspect the terminal command and copy the local address into your browser address bar (usually, the address is http://localhost:8000).