**Milestone : 1**

**Fine Tuning using Open Source Meta Llama 2 7b**

**Environment Set up**

1. I have attached python notebook “.ipynb” file and also finetune.py file. Both are same and can be used to fine tune.
2. You have to install all the necessary packages that I have defined in the notebook.
3. I have used hugging face to download the llama model by using cli-login and providing auth token to it, after having gated access of meta llama model through registration at meta website.
4. Auth token can be generated by going to the hugging face account settings --> Access Tokens --> then click on the new token button, then give the token name and select TYPE = WRITE, then click on generate a token button.
5. For fine tuning of open source LLM (meta llama 2-7b model), it will require a GPU.

**User Guide**

1. I have provided “ python\_code\_finetuning\_xml.ipynb ” and “finetune.py” file. You can use any of these file to finetune. Both are same and can be used to finetune.
2. “ xml\_csv\_data\_final.csv” is the dataset on which I have fine tuned the llama model . I have created this dataset which contains prompts and xml data.
3. “ LLAMA27B\_XML\_FINETUNED\_MODEL “ is the fine tuned model folder. This is the model that has learned all the data of xmls and prompts present in the dataset.

**Running the code**

1. In the python code, first install all the required packages and frameworks

pip install -q accelerate==0.21.0 peft==0.4.0 bitsandbytes==0.40.2 transformers==4.31.0 trl==0.4.7

1. Import all the packages and frameworks provided in the python code.
2. “ model\_name ” : This is the base model which is “ meta-llama/Llama-2-7b-hf ” .
3. “new\_model” : This is the fine tuned model which is “LLAMA27B\_XML\_FINETUNED\_MODEL”
4. “huggingface-cli login” . This is the code to authenticate with hugging face to use the base model.