LLM's FROM SCRATCH

---------- Build your own LLM ----------

1. First thing is to install the PYTHON 3.X in your system. Check whether the python is linked to environment variables or not.
2. Open the command prompt(terminal) and navigate to the location where you want to start the work on creating the LLM.
3. Create the new directory by using "mkdir <your_directory_name>".
4. Create the virtual environment in the directory you have created in the previous step by using the command " python -m venv cuda " cuda here helps in accelerating the GPU's. This helps us when we work on the huge data or huge number of files/documents etc..
5. After creating the virtual env in your directory then activate it by using this command "cuda\Scripts\activate". If everything runs successfully then you can continue to the next step or if you got this below error message

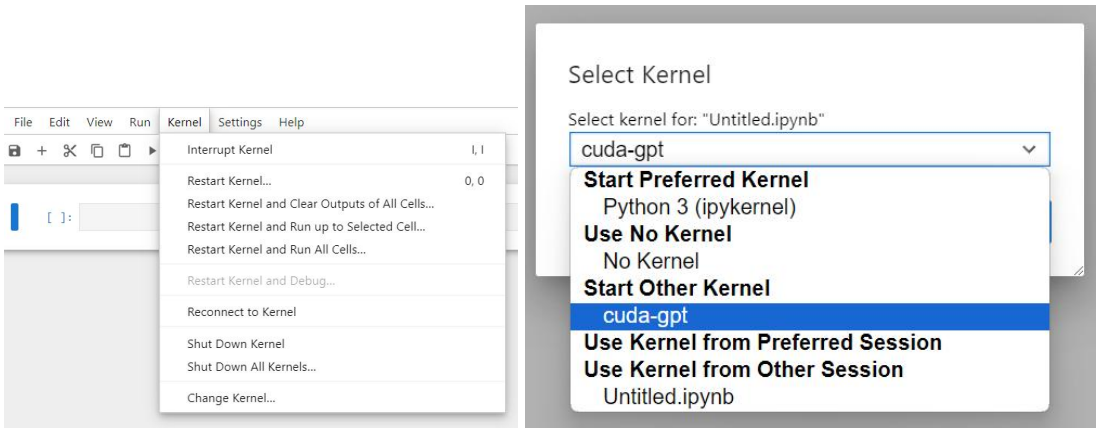
```
PS C:\Users\Nithin\Desktop> cuda\Scripts\activate
cuda\Scripts\activate : File C:\Users\Nithin\Desktop\cuda\Scripts\Activate.ps1 cannot be loaded because running
scripts is disabled on this system. For more information, see about_Execution_Policies at
https:/go.microsoft.com/fwlink/?LinkID=135170.
At line:1 char:1
+ cuda\Scripts\activate
+ ~~~~~~~~~~~~~~~~~~~~
    + CategoryInfo          : SecurityError: (:) [], PSSecurityException
    + FullyQualifiedErrorId : UnauthorizedAccess
```

6. Then you can use this command to eliminate this error "Set-ExecutionPolicy -ExecutionPolicy RemoteSigned -Scope Process". The above error rises due to you're trying to activate a Python virtual environment on Windows using PowerShell, but the execution of scripts is disabled on your system due to the default execution policy. This command will eliminate that error then you can activate your virtual env using the previous command which is "cuda\Scripts\activate".
7. Now let us install some packages here by using the command "pip install matplotlib numpy pylzma ipykernal jupyter".
8. You might get an error on installing **"pylzma".** It is because the visual studio C++ tools. You can download them from https://visualstudio.microsoft.com/downloads/?q=build+tools

9. After downloading the tools for free from the link above then execute the file and click on modify. Then from the popup select the Desktop development with C++ and .Net desktop build tools also Universal tools will be already selected then click on modify.
10. When it is done run the previous command again this time hopefully you won't get any error.
11. Now we have to install torch but here instead of using "pip install torch", We will be using "pip install torch --index-url https://download.pytorch.org/whl/cu118". This helps in installing torch along with cuda in it.
12. Now open the jupyter notebook from the same terminal using the command "jupyter notebook".
13. Make sure that everything is running correctly and then again come to terminal and click on cntrl+Q. This quits your notebook and type the following command for installing cuda in your notebook "python -m ipykernel install --user --name=cuda --display-name "cuda-gpt"". This will give the following output.

```
(cuda) (base) C:\Users\Nithin\Desktop\Nithin_LLM_DEV>python -m ipykernel install --user --name=cuda --display-name "cuda
-gpt"
0.00s - Debugger warning: It seems that frozen modules are being used, which may
0.00s - make the debugger miss breakpoints. Please pass -Xfrozen_modules=off
0.00s - to python to disable frozen modules.
0.00s - Note: Debugging will proceed. Set PYDEVD_DISABLE_FILE_VALIDATION=1 to disable this validation.
Installed kernelspec cuda in C:\Users\Nithin\AppData\Roaming\jupyter\kernels\cuda
```

14. Now again run the jupyter notebook from the same terminal. Then when you open/create a test .ipynb file and click on kernel and change kernel then you can see cuda-gpt there. Like below



15. Select the cuda-gpt and restart the kernel if needed.
16. This says that we are good to go!
17. Now you better rename your ipynb file as "BIGRAM" or whatever that make sense.

18. First we have to start with the normal dataset which you can find here https://www.gutenberg.org/cache/epub/22566/pg22566.txt
19. Just save this file with whatever name you would like to have.

    NOTE: Saving file in the same directory where we are doing all the work can be helpful.
20. After saving these file delete the license information and part after "The End [Illustration]". Delete Illustration also. So the starting of the text file would be "DOROTHY AND THE WIZARD IN OZ" and the ending will be "The End"
21. After this work you will be having the text file in your directory. Now we will be performing programming operation on this text file like opening, reading etc. Just the basic things.