

Building LLM Model using LLAMA3 integrating HUGGING FACE - By Sunitha Mekala

```
from google.colab import drive
drive.mount('/content/drive')
```

↗ Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

↕ New Section

!nvidia-smi

↗ Wed Feb 5 18:54:53 2025

+-----+-----+-----+-----+-----+									
NVIDIA-SMI 550.54.15			Driver Version: 550.54.15			CUDA Version: 12.4			
+-----+-----+-----+-----+-----+									
GPU	Name		Persistence-M	Bus-Id	Disp.A	Volatile	Uncorr.	ECC	
Fan	Temp	Perf	Pwr:Usage/Cap		Memory-Usage	GPU-Util	Compute M.		
							MIG M.		
=====									
0	Tesla T4		Off	00000000:00:04.0	Off			0	
N/A	40C	P8	9W / 70W		0MiB / 15360MiB	0%	Default		
							N/A		
+-----+-----+-----+-----+-----+									
+-----+-----+-----+-----+-----+									
Processes:									
GPU	GI	CI	PID	Type	Process name	GPU Memory			
	ID	ID				Usage			
=====									
No running processes found									
+-----+-----+-----+-----+-----+									

!pip install -r requirement.txt



```

Uninstalling ~
Successfully uninstalled nvidia-cuda-nvrtc-cu12-12.5.82
Attempting uninstall: nvidia-cuda-cupti-cu12
Found existing installation: nvidia-cuda-cupti-cu12 12.5.82
Uninstalling nvidia-cuda-cupti-cu12-12.5.82:
Successfully uninstalled nvidia-cuda-cupti-cu12-12.5.82
Attempting uninstall: nvidia-cublas-cu12
Found existing installation: nvidia-cublas-cu12 12.5.3.2
Uninstalling nvidia-cublas-cu12-12.5.3.2:
Successfully uninstalled nvidia-cublas-cu12-12.5.3.2
Attempting uninstall: nvidia-cusparse-cu12
Found existing installation: nvidia-cusparse-cu12 12.5.1.3
Uninstalling nvidia-cusparse-cu12-12.5.1.3:
Successfully uninstalled nvidia-cusparse-cu12-12.5.1.3
Attempting uninstall: nvidia-cudnn-cu12
Found existing installation: nvidia-cudnn-cu12 9.3.0.75
Uninstalling nvidia-cudnn-cu12-9.3.0.75:
Successfully uninstalled nvidia-cudnn-cu12-9.3.0.75
Attempting uninstall: tokenizers
Found existing installation: tokenizers 0.21.0
Uninstalling tokenizers-0.21.0:
Successfully uninstalled tokenizers-0.21.0
Attempting uninstall: nvidia-cusolver-cu12
Found existing installation: nvidia-cusolver-cu12 11.6.3.83
Uninstalling nvidia-cusolver-cu12-11.6.3.83:
Successfully uninstalled nvidia-cusolver-cu12-11.6.3.83
Attempting uninstall: transformers
Found existing installation: transformers 4.47.1
Uninstalling transformers-4.47.1:
Successfully uninstalled transformers-4.47.1
Attempting uninstall: accelerate
Found existing installation: accelerate 1.2.1
Uninstalling accelerate-1.2.1:
Successfully uninstalled accelerate-1.2.1

```

ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of the following dependency conflict:
sentence-transformers 3.3.1 requires transformers<5.0.0,>=4.41.0, but you have transformers 4.40.0 which is incompatible.

Successfully installed accelerate-0.29.3 bitsandbytes-0.43.1 nvidia-cublas-cu12-12.4.5.8 nvidia-cuda-cupti-cu12-12.4.127 nvidia-cuda-nvrtc-cu12-12.4.127 nvidia-cuda-runtime

```

import json
import torch
from transformers import (AutoTokenizer, AutoModelForCausalLM, BitsAndBytesConfig, pipeline)

```

```

config_data = json.load(open("config.json"))
HF_TOKEN = config_data["HF_TOKEN"]

```

```

model_name = "meta-llama/Meta-Llama-3-8B"

```

```

bnb_config = BitsAndBytesConfig(
    load_in_4bit=True,
    bnb_4bit_use_double_quant=True,
    bnb_4bit_quant_type="nf4",
    bnb_4bit_compute_dtype=torch.bfloat16
)

```

```
tokenizer = AutoTokenizer.from_pretrained(model_name,token=HF_TOKEN)
tokenizer.pad_token = tokenizer.eos_token
```

⚡ /usr/local/lib/python3.11/dist-packages/huggingface_hub/file_download.py:795: FutureWarning: `resume_download` is deprecated and will be removed in version 1.0.0. Downloads a warnings.warn(
Special tokens have been added in the vocabulary, make sure the associated word embeddings are fine-tuned or trained.

```
model = AutoModelForCausalLM.from_pretrained(  
    model_name,  
    token=HF_TOKEN,  
    quantization_config=bnb_config,  
    device_map='auto'  
)
```

⚡ Loading checkpoint shards: 100% 4/4 [01:28<00:00, 18.74s/it]

```
text_generator = pipeline(  
    task="text-generation",  
    model=model,  
    tokenizer=tokenizer,  
    max_new_tokens = 128,  
)
```

```
def get_response(prompt):  
    sequences = text_generator(prompt)  
    gen_text = sequences[0]['generated_text']  
    return gen_text
```

```
prompt = "Job description for Data Analyst"
```

```
llama3_response = get_response(prompt)
```

```
llama3_response
```

⚡ 'Job description for Data Analyst\nData Analyst is a professional who collects, analyzes, and interprets data to provide insights and inform decision-making. They work with a variety of data sources, including databases, spreadsheets, and statistical software, to identify trends and patterns in the data. They may also create visualizations, such as charts and graphs, to communicate their findings. Data Analysts may work in a variety of industries, including finance, healthcare, marketing, and technology. They may also specialize in a particular area, such as data visualization, predictive analytics, or data science.'

```
prompt = " What are the certifications related to Data Analyst "
```

```
llama3_response = get_response(prompt)
```

```
llama3_response
```

⚡ ' What are the certifications related to Data Analyst 2021? Data Analyst 2021 is certified by the following: Data Analyst 2021 is a course that provides you with the skills and knowledge you need to succeed in the field of data analysis. It covers the basics of data analysis, including how to collect, organize, and analyze data, as well as how to present and interpret data. The course also covers the latest trends and technologies in data analysis, including big data, data visualization, and predictive analytics.'

Start coding or [generate](#) with AI.