

## ✓ Fine-tuning a pre-trained model on a Dataset

I'll illustrate how to finetune the GPT-2 model from Hugging Face Transformers.

```
!pip install transformers datasets torch
```

```
from datasets import load_dataset
from huggingface_hub import login
```

### Retrieve the Hugging Face token

If using in Google Colab, using the updated method:

```
from google.colab import userdata
hf_token = userdata.get("YOUR_HUGGING_FACE_TOKEN") # Replace with your token

# Check if the token was retrieved correctly
if hf_token is None:
    print("Error: Hugging Face token not found.")
else:
    print(f"Token Retrieved: {\"YOUR_HUGGING_FACE_TOKEN\"[:5]}...") # Only print f

    # Log in using the Hugging Face API token
    login("YOUR_HUGGING_FACE_TOKEN")
```

**\*\* Load your Dataset\*\***

Here I'll be using the IMDB dataset, which contains movie reviews. The dataset is split into training and testing (validation) sets.

```
# Load the IMDB dataset
dataset = load_dataset("imdb")
```

```
# Print a sample from the dataset to verify
print(dataset["train"][7])
```

```
→ /usr/local/lib/python3.11/dist-packages/huggingface_hub/utils/_auth.py:94: Use
The secret `HF_TOKEN` does not exist in your Colab secrets.
To authenticate with the Hugging Face Hub, create a token in your settings tab
You will be able to reuse this secret in all of your notebooks.
Please note that authentication is recommended but still optional to access pu
warnings.warn(
```

```
README.md: 100% 7.81k/7.81k [00:00<00:00, 255kB/s]
train-00000-of- 21.0M/21.0M [00:00<00:00, 93.2MB/s]
00001.parquet: 100%
test-00000-of- 20.5M/20.5M [00:00<00:00, 74.1MB/s]
00001.parquet: 100%
unsupervised-00000-of- 42.0M/42.0M [00:00<00:00, 187MB/s]
00001.parquet: 100%
Generating train split: 100% 25000/25000 [00:00<00:00, 56148.58 examples/
s]
Generating test split: 100% 25000/25000 [00:00<00:00, 55661.36 examples/
s]
Generating unsupervised split: 100% 50000/50000 [00:00<00:00, 80859.48 examples/
s]
```

```
{'text': 'When I first saw a glimpse of this movie, I quickly noticed the actr
```

```
# Check the structure of the dataset
print(dataset)
```

## ✓ Finetune the pretrained GPT2 model

```
from transformers import GPT2Tokenizer, GPT2LMHeadModel, Trainer, TrainingArgument
```

```
# Filter out any empty entries
filtered_train = dataset['train'].filter(lambda x: x['text'].strip() != '')
filtered_validation = dataset['test'].filter(lambda x: x['text'].strip() != '')
```

```

# Load the GPT-2 model and tokenizer
model_name = "gpt2"
tokenizer = GPT2Tokenizer.from_pretrained(model_name)

# Set the pad token id to the eos token id (GPT-2 does not have a dedicated pad token)
tokenizer.pad_token = tokenizer.eos_token
model = GPT2LMHeadModel.from_pretrained(model_name)

# Tokenize the dataset and create labels (shifted input)
def tokenize_function(examples):
    # Tokenize input text
    encodings = tokenizer(examples['text'], truncation=True, padding="max_length")

    # Shift labels by one token for causal language modeling
    labels = encodings['input_ids'].copy()
    for i in range(len(labels)):
        labels[i] = labels[i][1:] + [tokenizer.pad_token_id] # Shift the input and add pad token

    # Add labels to encodings
    encodings['labels'] = labels
    return encodings

# Tokenize the training and validation datasets
tokenized_train = filtered_train.map(tokenize_function, batched=True)
tokenized_validation = filtered_validation.map(tokenize_function, batched=True)

# Set training arguments
training_args = TrainingArguments(
    output_dir="./results",
    learning_rate=2e-5,
    per_device_train_batch_size=2,
    num_train_epochs=3,
    logging_dir='./logs', # Directory for storing logs
    logging_steps=10,     # Log every 10 steps
    report_to="none",     # Disable WandB
    disable_tqdm=True     # Disable tqdm progress bar
)

# Create Trainer
trainer = Trainer(
    model=model,
    args=training_args,
    train_dataset=tokenized_train,
    eval_dataset=tokenized_validation,
)

# Fine-tune the model

```

```
trainer.train()
```



```
25000/25000 [01:06<00:00, 422.14 examples/
```

```
s]
```

```
25000/25000 [01:08<00:00, 397.68 examples/
```

```
s]
```

```
re` was set in the config but it is unrecognised.Using the default loss: `ForCa
11, 'grad_norm': 31.923748016357422, 'learning_rate': 1.99952e-05, 'epoch': 0.0
74, 'grad_norm': 18.054859161376953, 'learning_rate': 1.998986666666667e-05, 'e
3, 'grad_norm': 12.87031078338623, 'learning_rate': 1.9984533333333335e-05, 'ep
19, 'grad_norm': 4.184741497039795, 'learning_rate': 1.99792e-05, 'epoch': 0.00
13, 'grad_norm': 8.271995544433594, 'learning_rate': 1.997386666666667e-05, 'ep
27, 'grad_norm': 8.88296890258789, 'learning_rate': 1.9968533333333334e-05, 'ep
34, 'grad_norm': 5.019299507141113, 'learning_rate': 1.9963200000000002e-05, 'e
08, 'grad_norm': 5.493118762969971, 'learning_rate': 1.995786666666667e-05, 'e
37, 'grad_norm': 4.069597244262695, 'learning_rate': 1.9952533333333336e-05, 'e
02, 'grad_norm': 7.409961700439453, 'learning_rate': 1.99472e-05, 'epoch': 0.00
35, 'grad_norm': 3.5042500495910645, 'learning_rate': 1.994186666666667e-05, 'e
19, 'grad_norm': 4.437108516693115, 'learning_rate': 1.9936533333333335e-05, 'e
52, 'grad_norm': 6.663012504577637, 'learning_rate': 1.9931200000000003e-05, 'e
5, 'grad_norm': 7.043432712554932, 'learning_rate': 1.992586666666667e-05, 'epo
19, 'grad_norm': 5.719522953033447, 'learning_rate': 1.9920533333333334e-05, 'e
13, 'grad_norm': 7.606908798217773, 'learning_rate': 1.9915200000000002e-05, 'e
51, 'grad_norm': 6.5985517501831055, 'learning_rate': 1.990986666666667e-05, '
25, 'grad_norm': 3.3262441158294678, 'learning_rate': 1.9904533333333336e-05, '
35, 'grad_norm': 6.717874050140381, 'learning_rate': 1.98992e-05, 'epoch': 0.01
36, 'grad_norm': 5.851573467254639, 'learning_rate': 1.989386666666667e-05, 'ep
01, 'grad_norm': 5.2495574951171875, 'learning_rate': 1.9888533333333335e-05, '
15, 'grad_norm': 5.631509304046631, 'learning_rate': 1.9883200000000003e-05, 'e
03, 'grad_norm': 4.258436679840088, 'learning_rate': 1.987786666666667e-05, 'e
07, 'grad_norm': 6.341273307800293, 'learning_rate': 1.9872533333333333e-05, 'e
04, 'grad_norm': 5.05233097076416, 'learning_rate': 1.9867200000000002e-05, 'ep
18, 'grad_norm': 7.879485607147217, 'learning_rate': 1.986186666666667e-05, 'ep
07, 'grad_norm': 6.889856815338135, 'learning_rate': 1.9856533333333335e-05, 'e
35, 'grad_norm': 3.443033218383789, 'learning_rate': 1.98512e-05, 'epoch': 0.02
24, 'grad_norm': 3.7993249893188477, 'learning_rate': 1.984586666666667e-05, 'e
02, 'grad_norm': 6.788940906524658, 'learning_rate': 1.9840533333333334e-05, 'e
03, 'grad_norm': 5.81478214263916, 'learning_rate': 1.9835200000000003e-05, 'ep
11, 'grad_norm': 7.389142036437988, 'learning_rate': 1.982986666666667e-05, 'e
33, 'grad_norm': 6.091892242431641, 'learning_rate': 1.9824533333333336e-05, 'e
32, 'grad_norm': 6.3941473960876465, 'learning_rate': 1.98192e-05, 'epoch': 0.0
57, 'grad_norm': 5.9340410232543945, 'learning_rate': 1.981386666666667e-05, '
06, 'grad_norm': 29.516494750976562, 'learning_rate': 1.9808533333333335e-05, '
17, 'grad_norm': 4.09855318069458, 'learning_rate': 1.9803200000000004e-05, 'ep
38, 'grad_norm': 4.586108207702637, 'learning_rate': 1.979786666666667e-05, 'ep
18, 'grad_norm': 5.282919883728027, 'learning_rate': 1.9792533333333334e-05, 'e
58, 'grad_norm': 4.468644618988037, 'learning_rate': 1.9787200000000003e-05, 'e
3, 'grad_norm': 3.9015157222747803, 'learning_rate': 1.978186666666667e-05, 'e
12, 'grad_norm': 5.657168388366699, 'learning_rate': 1.9776533333333333e-05, 'e
14, 'grad_norm': 3.6321098804473877, 'learning_rate': 1.97712e-05, 'epoch': 0.0
```

14, 'grad\_norm': 3.032103880473877, 'learning\_rate': 1.97712e-05, 'epoch': 0.0  
19, 'grad\_norm': 4.708899974822998, 'learning\_rate': 1.976586666666667e-05, 'ep  
38, 'grad\_norm': 3.509389877319336, 'learning\_rate': 1.9760533333333335e-05, 'e  
55, 'grad\_norm': 5.379390716552734, 'learning\_rate': 1.97552e-05, 'epoch': 0.03  
1, 'grad\_norm': 2.57448148727417, 'learning\_rate': 1.974986666666667e-05, 'epoc  
51, 'grad\_norm': 5.679721355438232, 'learning\_rate': 1.9744533333333337e-05, 'e  
33, 'grad\_norm': 3.4319114685058594, 'learning\_rate': 1.97392e-05, 'epoch': 0.0  
78, 'grad\_norm': 4.249644756317139, 'learning\_rate': 1.9733866666666668e-05, 'e  
27, 'grad\_norm': 2.7283499240875244, 'learning\_rate': 1.9728533333333336e-05, '  
36, 'grad\_norm': 4.762662887573242, 'learning\_rate': 1.97232e-05, 'epoch': 0.04  
03, 'grad\_norm': 3.7060389518737793, 'learning\_rate': 1.971786666666666e-05, '  
9, 'grad\_norm': 2.8031599521636963, 'learning\_rate': 1.9712533333333335e-05, '  
34, 'grad\_norm': 4.6092023849487305, 'learning\_rate': 1.9707200000000003e-05, '  
39, 'grad\_norm': 3.709385633468628, 'learning\_rate': 1.970186666666667e-05, 'ep  
53, 'grad\_norm': 6.565674781799316, 'learning\_rate': 1.9696533333333334e-05, 'e  
51, 'grad\_norm': 3.671675682067871, 'learning\_rate': 1.9691200000000002e-05, 'e  
36, 'grad\_norm': 4.1044087409973145, 'learning\_rate': 1.968586666666667e-05, '  
21, 'grad\_norm': 5.255737781524658, 'learning\_rate': 1.9680533333333336e-05, 'e  
1, 'grad\_norm': 3.5935940742492676, 'learning\_rate': 1.96752e-05, 'epoch': 0.04  
31, 'grad\_norm': 2.446730852127075, 'learning\_rate': 1.966986666666667e-05, 'ep  
56, 'grad\_norm': 3.0968916416168213, 'learning\_rate': 1.9664533333333335e-05, '  
54, 'grad\_norm': 4.4465765953063965, 'learning\_rate': 1.96592e-05, 'epoch': 0.0  
39, 'grad\_norm': 4.201115608215332, 'learning\_rate': 1.9653866666666668e-05, 'e  
59, 'grad\_norm': 3.3887112140655518, 'learning\_rate': 1.9648533333333333e-05, '  
36, 'grad\_norm': 2.83296799659729, 'learning\_rate': 1.9643200000000002e-05, 'ep  
31, 'grad\_norm': 4.292440414428711, 'learning\_rate': 1.9637866666666667e-05, 'e  
56, 'grad\_norm': 3.439561128616333, 'learning\_rate': 1.9632533333333336e-05, 'e  
52, 'grad\_norm': 2.8038508892059326, 'learning\_rate': 1.96272e-05, 'epoch': 0.0  
15, 'grad\_norm': 4.299710750579834, 'learning\_rate': 1.962186666666667e-05, 'ep  
37, 'grad\_norm': 3.387451648712158, 'learning\_rate': 1.9616533333333334e-05, 'e  
08, 'grad\_norm': 4.337559223175049, 'learning\_rate': 1.9611200000000003e-05, 'e  
35, 'grad\_norm': 3.6200218200683594, 'learning\_rate': 1.9605866666666668e-05, '  
15, 'grad\_norm': 2.68186616897583, 'learning\_rate': 1.9600533333333337e-05, 'ep  
04, 'grad\_norm': 2.5026016235351562, 'learning\_rate': 1.9595200000000002e-05, '  
05, 'grad\_norm': 3.036390781402588, 'learning\_rate': 1.9589866666666667e-05, 'e  
14, 'grad\_norm': 3.3040730953216553, 'learning\_rate': 1.9584533333333335e-05, '  
53, 'grad\_norm': 5.250816822052002, 'learning\_rate': 1.9579200000000004e-05, 'e  
01, 'grad\_norm': 3.3328020572662354, 'learning\_rate': 1.9573866666666667e-05, 'e  
07, 'grad\_norm': 3.691269636154175, 'learning\_rate': 1.9568533333333334e-05, 'e  
12, 'grad\_norm': 3.73764967918396, 'learning\_rate': 1.9563200000000003e-05, 'ep  
01, 'grad\_norm': 2.653515577316284, 'learning\_rate': 1.9557866666666668e-05, 'e  
03, 'grad\_norm': 2.5527327060699463, 'learning\_rate': 1.9552533333333333e-05, '  
33, 'grad\_norm': 3.145449161529541, 'learning\_rate': 1.95472e-05, 'epoch': 0.06  
31, 'grad\_norm': 2.0324044227600098, 'learning\_rate': 1.9541866666666667e-05, 'e  
53, 'grad\_norm': 3.1050870418548584, 'learning\_rate': 1.9536533333333335e-05, '  
16, 'grad\_norm': 3.386395215988159, 'learning\_rate': 1.95312e-05, 'epoch': 0.07  
29, 'grad\_norm': 3.534799098968506, 'learning\_rate': 1.9525866666666667e-05, 'ep  
12, 'grad\_norm': 3.704958915710449, 'learning\_rate': 1.9520533333333337e-05, 'e  
34, 'grad\_norm': 3.327565908432007, 'learning\_rate': 1.9515200000000002e-05, 'e  
28, 'grad\_norm': 2.986471176147461, 'learning\_rate': 1.9509866666666668e-05, 'e  
19, 'grad\_norm': 3.717557907104492, 'learning\_rate': 1.9504533333333336e-05, 'e  
31, 'grad\_norm': 3.462099075317383, 'learning\_rate': 1.94992e-05, 'epoch': 0.07  
77, 'grad\_norm': 3.9602081775665283, 'learning\_rate': 1.949386666666666e-05, '  
53, 'grad\_norm': 3.3070006370544434, 'learning\_rate': 1.9488533333333335e-05, '  
55, 'grad\_norm': 2.8105702820718804, 'learning\_rate': 1.9483200000000003e-05, '

50, 'grad\_norm': 2.9105705830718994, 'learning\_rate': 1.9483200000000003e-05, 'epoch': 0.07  
51, 'grad\_norm': 4.235452175140381, 'learning\_rate': 1.9477866666666667e-05, 'epoch': 0.07  
51, 'grad\_norm': 2.965022563934326, 'learning\_rate': 1.9472533333333334e-05, 'epoch': 0.07  
53, 'grad\_norm': 3.3100271224975586, 'learning\_rate': 1.9467200000000002e-05, 'epoch': 0.07  
54, 'grad\_norm': 2.7340214252471924, 'learning\_rate': 1.9461866666666667e-05, 'epoch': 0.07  
56, 'grad\_norm': 2.6768863201141357, 'learning\_rate': 1.9456533333333333e-05, 'epoch': 0.07  
56, 'grad\_norm': 3.5833137035369873, 'learning\_rate': 1.94512e-05, 'epoch': 0.07  
51, 'grad\_norm': 3.085289478302002, 'learning\_rate': 1.9445866666666667e-05, 'epoch': 0.07  
53, 'grad\_norm': 2.799130439758301, 'learning\_rate': 1.9440533333333335e-05, 'epoch': 0.07  
52, 'grad\_norm': 3.284471035003662, 'learning\_rate': 1.94352e-05, 'epoch': 0.08  
55, 'grad\_norm': 3.8868002891540527, 'learning\_rate': 1.9429866666666667e-05, 'epoch': 0.08  
59, 'grad\_norm': 3.749011278152466, 'learning\_rate': 1.9424533333333337e-05, 'epoch': 0.08  
58, 'grad\_norm': 2.5727362632751465, 'learning\_rate': 1.9419200000000002e-05, 'epoch': 0.08  
57, 'grad\_norm': 3.254420280456543, 'learning\_rate': 1.9413866666666667e-05, 'epoch': 0.08  
52, 'grad\_norm': 3.2878849506378174, 'learning\_rate': 1.9408533333333336e-05, 'epoch': 0.08  
55, 'grad\_norm': 2.077953338623047, 'learning\_rate': 1.94032e-05, 'epoch': 0.08  
57, 'grad\_norm': 1.6306415796279907, 'learning\_rate': 1.9397866666666667e-05, 'epoch': 0.08  
52, 'grad\_norm': 2.54595685005188, 'learning\_rate': 1.9392533333333335e-05, 'epoch': 0.08  
59, 'grad\_norm': 3.3273706436157227, 'learning\_rate': 1.9387200000000003e-05, 'epoch': 0.08  
58, 'grad\_norm': 3.056511402130127, 'learning\_rate': 1.9381866666666668e-05, 'epoch': 0.08  
52, 'grad\_norm': 3.396552801132202, 'learning\_rate': 1.9376533333333333e-05, 'epoch': 0.08  
53, 'grad\_norm': 2.9605705738067627, 'learning\_rate': 1.9371200000000002e-05, 'epoch': 0.08  
56, 'grad\_norm': 2.2199580669403076, 'learning\_rate': 1.9365866666666667e-05, 'epoch': 0.08  
52, 'grad\_norm': 2.5890350341796875, 'learning\_rate': 1.9360533333333335e-05, 'epoch': 0.08  
53, 'grad\_norm': 2.9829208850860596, 'learning\_rate': 1.93552e-05, 'epoch': 0.09  
57, 'grad\_norm': 2.8248190879821777, 'learning\_rate': 1.9349866666666667e-05, 'epoch': 0.09  
54, 'grad\_norm': 3.819504737854004, 'learning\_rate': 1.9344533333333334e-05, 'epoch': 0.09  
56, 'grad\_norm': 2.0617995262145996, 'learning\_rate': 1.9339200000000003e-05, 'epoch': 0.09  
53, 'grad\_norm': 3.5061588287353516, 'learning\_rate': 1.9333866666666668e-05, 'epoch': 0.09  
52, 'grad\_norm': 3.1763722896575928, 'learning\_rate': 1.9328533333333333e-05, 'epoch': 0.09  
51, 'grad\_norm': 2.7464303970336914, 'learning\_rate': 1.93232e-05, 'epoch': 0.10  
55, 'grad\_norm': 2.2818691730499268, 'learning\_rate': 1.9317866666666667e-05, 'epoch': 0.10  
52, 'grad\_norm': 2.201746940612793, 'learning\_rate': 1.9312533333333335e-05, 'epoch': 0.10  
51, 'grad\_norm': 2.1305277347564697, 'learning\_rate': 1.93072e-05, 'epoch': 0.10  
51, 'grad\_norm': 1.8890564441680908, 'learning\_rate': 1.9301866666666667e-05, 'epoch': 0.10  
57, 'grad\_norm': 2.6733577251434326, 'learning\_rate': 1.9296533333333337e-05, 'epoch': 0.10  
55, 'grad\_norm': 1.9467169046401978, 'learning\_rate': 1.9291200000000003e-05, 'epoch': 0.10  
54, 'grad\_norm': 2.4373247623443604, 'learning\_rate': 1.9285866666666668e-05, 'epoch': 0.10  
58, 'grad\_norm': 2.859156608581543, 'learning\_rate': 1.9280533333333336e-05, 'epoch': 0.10  
57, 'grad\_norm': 2.1992812156677246, 'learning\_rate': 1.92752e-05, 'epoch': 0.11  
55, 'grad\_norm': 2.7341206073760986, 'learning\_rate': 1.9269866666666667e-05, 'epoch': 0.11  
51, 'grad\_norm': 2.5188515186309814, 'learning\_rate': 1.9264533333333335e-05, 'epoch': 0.11  
56, 'grad\_norm': 2.3780319690704346, 'learning\_rate': 1.9259200000000004e-05, 'epoch': 0.11  
54, 'grad\_norm': 2.4923925399780273, 'learning\_rate': 1.9253866666666667e-05, 'epoch': 0.11  
53, 'grad\_norm': 3.072009325027466, 'learning\_rate': 1.9248533333333334e-05, 'epoch': 0.11  
55, 'grad\_norm': 3.0443921089172363, 'learning\_rate': 1.9243200000000002e-05, 'epoch': 0.11  
57, 'grad\_norm': 2.85638165473938, 'learning\_rate': 1.9237866666666668e-05, 'epoch': 0.11  
51, 'grad\_norm': 2.4271647930145264, 'learning\_rate': 1.9232533333333333e-05, 'epoch': 0.11  
58, 'grad\_norm': 1.9803913831710815, 'learning\_rate': 1.92272e-05, 'epoch': 0.12  
53, 'grad\_norm': 2.896326780319214, 'learning\_rate': 1.9221866666666667e-05, 'epoch': 0.12  
57, 'grad\_norm': 2.7296016216278076, 'learning\_rate': 1.9216533333333335e-05, 'epoch': 0.12  
51, 'grad\_norm': 2.264285087585449, 'learning\_rate': 1.92112e-05, 'epoch': 0.12  
58, 'grad\_norm': 2.326590061187744, 'learning\_rate': 1.9205866666666667e-05, 'epoch': 0.12  
58, 'grad\_norm': 2.6985809803009033, 'learning\_rate': 1.9200533333333337e-05, 'epoch': 0.12  
52, 'grad\_norm': 1.6002546714700715, 'learning\_rate': 1.9195200000000003e-05, 'epoch': 0.12

2, 'grad\_norm': 1.6003546714782715, 'learning\_rate': 1.9195200000000002e-05, 'epoch': 0.1  
36, 'grad\_norm': 2.0030534267425537, 'learning\_rate': 1.9189866666666667e-05, 'epoch': 0.1  
8, 'grad\_norm': 2.5907721519470215, 'learning\_rate': 1.9184533333333336e-05, 'epoch': 0.1  
4, 'grad\_norm': 2.6764469146728516, 'learning\_rate': 1.91792e-05, 'epoch': 0.1  
2, 'grad\_norm': 1.7511640787124634, 'learning\_rate': 1.9173866666666666e-05, 'epoch': 0.1  
9, 'grad\_norm': 2.8010876178741455, 'learning\_rate': 1.9168533333333335e-05, 'epoch': 0.1  
9, 'grad\_norm': 2.2684836387634277, 'learning\_rate': 1.9163200000000003e-05, 'epoch': 0.1  
2, 'grad\_norm': 2.348691940307617, 'learning\_rate': 1.9157866666666668e-05, 'epoch': 0.1  
9, 'grad\_norm': 3.012720823287964, 'learning\_rate': 1.9152533333333333e-05, 'epoch': 0.1  
6, 'grad\_norm': 2.202831268310547, 'learning\_rate': 1.9147200000000002e-05, 'epoch': 0.1  
4, 'grad\_norm': 2.885568857192993, 'learning\_rate': 1.9141866666666667e-05, 'epoch': 0.1  
6, 'grad\_norm': 2.66544246673584, 'learning\_rate': 1.9136533333333336e-05, 'epoch': 0.1  
9, 'grad\_norm': 2.6761248111724854, 'learning\_rate': 1.91312e-05, 'epoch': 0.1  
6, 'grad\_norm': 2.7393078804016113, 'learning\_rate': 1.9125866666666667e-05, 'epoch': 0.1  
5, 'grad\_norm': 2.7302680015563965, 'learning\_rate': 1.9120533333333334e-05, 'epoch': 0.1  
7, 'grad\_norm': 2.929361581802368, 'learning\_rate': 1.9115200000000003e-05, 'epoch': 0.1  
4, 'grad\_norm': 1.7791801691055298, 'learning\_rate': 1.9109866666666668e-05, 'epoch': 0.1  
8, 'grad\_norm': 2.1272494792938232, 'learning\_rate': 1.9104533333333337e-05, 'epoch': 0.1  
8, 'grad\_norm': 3.1863038539886475, 'learning\_rate': 1.9099200000000002e-05, 'epoch': 0.1  
7, 'grad\_norm': 2.4651927947998047, 'learning\_rate': 1.9093866666666667e-05, 'epoch': 0.1  
2, 'grad\_norm': 2.4736461639404297, 'learning\_rate': 1.9088533333333335e-05, 'epoch': 0.1  
2, 'grad\_norm': 2.8782835006713867, 'learning\_rate': 1.90832e-05, 'epoch': 0.13  
8, 'grad\_norm': 2.6151857376098633, 'learning\_rate': 1.9077866666666667e-05, 'epoch': 0.1  
3, 'grad\_norm': 2.6613547801971436, 'learning\_rate': 1.9072533333333334e-05, 'epoch': 0.1  
5, 'grad\_norm': 2.5885121822357178, 'learning\_rate': 1.9067200000000003e-05, 'epoch': 0.1  
9, 'grad\_norm': 2.036398410797119, 'learning\_rate': 1.9061866666666668e-05, 'epoch': 0.1  
5, 'grad\_norm': 1.8508795499801636, 'learning\_rate': 1.9056533333333336e-05, 'epoch': 0.1  
5, 'grad\_norm': 2.0405731201171875, 'learning\_rate': 1.90512e-05, 'epoch': 0.1  
8, 'grad\_norm': 3.238032341003418, 'learning\_rate': 1.9045866666666667e-05, 'epoch': 0.1  
6, 'grad\_norm': 2.5200254917144775, 'learning\_rate': 1.9040533333333335e-05, 'epoch': 0.1  
3, 'grad\_norm': 2.0291454792022705, 'learning\_rate': 1.9035200000000004e-05, 'epoch': 0.1  
5, 'grad\_norm': 2.660695791244507, 'learning\_rate': 1.9029866666666667e-05, 'epoch': 0.1  
1, 'grad\_norm': 2.5855159759521484, 'learning\_rate': 1.9024533333333334e-05, 'epoch': 0.1  
8, 'grad\_norm': 2.350635290145874, 'learning\_rate': 1.9019200000000002e-05, 'epoch': 0.1  
2, 'grad\_norm': 2.0199406147003174, 'learning\_rate': 1.9013866666666668e-05, 'epoch': 0.1  
6, 'grad\_norm': 1.849249243736267, 'learning\_rate': 1.9008533333333333e-05, 'epoch': 0.1  
7, 'grad\_norm': 1.5212446451187134, 'learning\_rate': 1.90032e-05, 'epoch': 0.1  
4, 'grad\_norm': 1.846662163734436, 'learning\_rate': 1.8997866666666667e-05, 'epoch': 0.1  
9, 'grad\_norm': 2.652993679046631, 'learning\_rate': 1.8992533333333335e-05, 'epoch': 0.1  
6, 'grad\_norm': 1.7849584817886353, 'learning\_rate': 1.89872e-05, 'epoch': 0.1  
8, 'grad\_norm': 2.187591552734375, 'learning\_rate': 1.8981866666666667e-05, 'epoch': 0.1  
2, 'grad\_norm': 2.701629400253296, 'learning\_rate': 1.8976533333333337e-05, 'epoch': 0.1  
1, 'grad\_norm': 2.6344411373138428, 'learning\_rate': 1.8971200000000002e-05, 'epoch': 0.1  
8, 'grad\_norm': 1.3634669780731201, 'learning\_rate': 1.8965866666666667e-05, 'epoch': 0.1  
3, 'grad\_norm': 2.0110208988189697, 'learning\_rate': 1.8960533333333336e-05, 'epoch': 0.1  
3, 'grad\_norm': 2.1830334663391113, 'learning\_rate': 1.89552e-05, 'epoch': 0.1  
4, 'grad\_norm': 2.9354734420776367, 'learning\_rate': 1.8949866666666666e-05, 'epoch': 0.1  
4, 'grad\_norm': 2.277343511581421, 'learning\_rate': 1.8944533333333335e-05, 'epoch': 0.1  
1, 'grad\_norm': 3.1606900691986084, 'learning\_rate': 1.8939200000000003e-05, 'epoch': 0.1  
7, 'grad\_norm': 2.555800676345825, 'learning\_rate': 1.8933866666666667e-05, 'epoch': 0.1  
3, 'grad\_norm': 1.5999627113342285, 'learning\_rate': 1.8928533333333334e-05, 'epoch': 0.1  
7, 'grad\_norm': 1.8414336442947388, 'learning\_rate': 1.8923200000000002e-05, 'epoch': 0.1  
2, 'grad\_norm': 1.877819299697876, 'learning\_rate': 1.8917866666666667e-05, 'epoch': 0.1  
3, 'grad\_norm': 2.826016664505005, 'learning\_rate': 1.8912533333333332e-05, 'epoch': 0.1

2, 'grad\_norm': 2.032562017440796, 'learning\_rate': 1.89072e-05, 'epoch': 0.16  
1, 'grad\_norm': 2.4123737812042236, 'learning\_rate': 1.890186666666667e-05, 'e  
36, 'grad\_norm': 2.0473906993865967, 'learning\_rate': 1.8896533333333335e-05, '  
36, 'grad\_norm': 3.188951253890991, 'learning\_rate': 1.88912e-05, 'epoch': 0.16  
19, 'grad\_norm': 2.0698091983795166, 'learning\_rate': 1.8885866666666668e-05, '  
14, 'grad\_norm': 3.2507898807525635, 'learning\_rate': 1.8880533333333337e-05, '  
58, 'grad\_norm': 3.401780366897583, 'learning\_rate': 1.8875200000000002e-05, 'e  
02, 'grad\_norm': 2.5277698040008545, 'learning\_rate': 1.8869866666666667e-05, '  
14, 'grad\_norm': 2.0640721321105957, 'learning\_rate': 1.8864533333333335e-05, '  
15, 'grad\_norm': 2.2594547271728516, 'learning\_rate': 1.88592e-05, 'epoch': 0.1  
03, 'grad\_norm': 2.7406258583068848, 'learning\_rate': 1.885386666666667e-05, 'e  
04, 'grad\_norm': 1.9385658502578735, 'learning\_rate': 1.8848533333333334e-05, '  
79, 'grad\_norm': 1.7795207500457764, 'learning\_rate': 1.8843200000000003e-05, '  
31, 'grad\_norm': 1.7314462661743164, 'learning\_rate': 1.8837866666666668e-05, '  
03, 'grad\_norm': 3.4664459228515625, 'learning\_rate': 1.8832533333333333e-05, '  
2, 'grad\_norm': 2.251551389694214, 'learning\_rate': 1.88272e-05, 'epoch': 0.176  
13, 'grad\_norm': 2.238996744155884, 'learning\_rate': 1.8821866666666667e-05, 'e  
02, 'grad\_norm': 2.6877071857452393, 'learning\_rate': 1.8816533333333335e-05, '  
51, 'grad\_norm': 1.3466393947601318, 'learning\_rate': 1.88112e-05, 'epoch': 0.1  
71, 'grad\_norm': 2.5820021629333496, 'learning\_rate': 1.880586666666667e-05, 'e  
72, 'grad\_norm': 2.0020759105682373, 'learning\_rate': 1.8800533333333334e-05, '  
52, 'grad\_norm': 2.073531150817871, 'learning\_rate': 1.8795200000000003e-05, 'e  
09, 'grad\_norm': 1.9323827028274536, 'learning\_rate': 1.8789866666666668e-05, '  
11, 'grad\_norm': 1.7184467315673828, 'learning\_rate': 1.8784533333333336e-05, '  
22, 'grad\_norm': 2.1026384830474854, 'learning\_rate': 1.87792e-05, 'epoch': 0.1  
32, 'grad\_norm': 2.885157823562622, 'learning\_rate': 1.877386666666667e-05, 'ep  
51, 'grad\_norm': 2.6526312828063965, 'learning\_rate': 1.8768533333333335e-05, '  
27, 'grad\_norm': 1.9212231636047363, 'learning\_rate': 1.87632e-05, 'epoch': 0.1  
26, 'grad\_norm': 1.8599188327789307, 'learning\_rate': 1.875786666666667e-05, 'e  
51, 'grad\_norm': 2.6501588821411133, 'learning\_rate': 1.8752533333333337e-05, '  
27, 'grad\_norm': 3.067580461502075, 'learning\_rate': 1.8747200000000002e-05, 'e  
02, 'grad\_norm': 1.7258630990982056, 'learning\_rate': 1.8741866666666668e-05, '  
52, 'grad\_norm': 2.279006242752075, 'learning\_rate': 1.8736533333333336e-05, 'e  
54, 'grad\_norm': 2.1680591106414795, 'learning\_rate': 1.87312e-05, 'epoch': 0.1  
57, 'grad\_norm': 3.334845781326294, 'learning\_rate': 1.8725866666666666e-05, 'e  
3, 'grad\_norm': 2.006725788116455, 'learning\_rate': 1.8720533333333335e-05, 'ep  
05, 'grad\_norm': 1.4907320737838745, 'learning\_rate': 1.8715200000000003e-05, '  
54, 'grad\_norm': 2.3607771396636963, 'learning\_rate': 1.870986666666667e-05, 'e  
59, 'grad\_norm': 2.5421347618103027, 'learning\_rate': 1.8704533333333334e-05, '  
55, 'grad\_norm': 1.8589134216308594, 'learning\_rate': 1.8699200000000002e-05, '  
26, 'grad\_norm': 2.29250431060791, 'learning\_rate': 1.869386666666667e-05, 'epo  
56, 'grad\_norm': 2.438027858734131, 'learning\_rate': 1.8688533333333332e-05, 'e  
72, 'grad\_norm': 2.1451244354248047, 'learning\_rate': 1.86832e-05, 'epoch': 0.1  
23, 'grad\_norm': 2.411574602127075, 'learning\_rate': 1.867786666666667e-05, 'ep  
17, 'grad\_norm': 2.503014326095581, 'learning\_rate': 1.8672533333333335e-05, 'e  
38, 'grad\_norm': 1.7427167892456055, 'learning\_rate': 1.86672e-05, 'epoch': 0.2  
33, 'grad\_norm': 1.7571697235107422, 'learning\_rate': 1.8661866666666668e-05, '  
08, 'grad\_norm': 2.414395570755005, 'learning\_rate': 1.8656533333333337e-05, 'e  
09, 'grad\_norm': 1.8756753206253052, 'learning\_rate': 1.8651200000000002e-05, '  
15, 'grad\_norm': 2.8002967834472656, 'learning\_rate': 1.8645866666666667e-05, '  
09, 'grad\_norm': 1.928343653678894, 'learning\_rate': 1.8640533333333336e-05, 'e  
55, 'grad\_norm': 2.327413558959961, 'learning\_rate': 1.86352e-05, 'epoch': 0.20  
19, 'grad\_norm': 2.609231948852539, 'learning\_rate': 1.8629866666666666e-05, 'e  
59, 'grad\_norm': 2.136284112930298, 'learning\_rate': 1.8624533333333334e-05, 'e



29, 'grad\_norm': 1.6790283918380737, 'learning\_rate': 1.8619200000000003e-05, '  
38, 'grad\_norm': 1.8968563079833984, 'learning\_rate': 1.8613866666666668e-05, '  
5, 'grad\_norm': 2.6525118350982666, 'learning\_rate': 1.8608533333333333e-05, 'e  
38, 'grad\_norm': 1.8964229822158813, 'learning\_rate': 1.8603200000000002e-05, '  
56, 'grad\_norm': 1.5852984189987183, 'learning\_rate': 1.8597866666666667e-05, '  
97, 'grad\_norm': 2.0912671089172363, 'learning\_rate': 1.8592533333333335e-05, '  
07, 'grad\_norm': 2.357309103012085, 'learning\_rate': 1.85872e-05, 'epoch': 0.21  
15, 'grad\_norm': 2.4163811206817627, 'learning\_rate': 1.8581866666666667e-05, 'e  
91, 'grad\_norm': 2.6044740676879883, 'learning\_rate': 1.8576533333333334e-05, '  
51, 'grad\_norm': 1.850614070892334, 'learning\_rate': 1.8571200000000003e-05, 'e  
24, 'grad\_norm': 1.9111969470977783, 'learning\_rate': 1.8565866666666668e-05, '  
23, 'grad\_norm': 2.736480712890625, 'learning\_rate': 1.8560533333333336e-05, 'e  
19, 'grad\_norm': 2.370161294937134, 'learning\_rate': 1.85552e-05, 'epoch': 0.21  
18, 'grad\_norm': 2.1418237686157227, 'learning\_rate': 1.8549866666666667e-05, '  
32, 'grad\_norm': 2.217397928237915, 'learning\_rate': 1.8544533333333335e-05, 'e  
22, 'grad\_norm': 1.4949636459350586, 'learning\_rate': 1.85392e-05, 'epoch': 0.2  
34, 'grad\_norm': 1.6078481674194336, 'learning\_rate': 1.8533866666666667e-05, 'e  
24, 'grad\_norm': 1.889177680015564, 'learning\_rate': 1.8528533333333334e-05, 'e  
57, 'grad\_norm': 2.352154016494751, 'learning\_rate': 1.8523200000000002e-05, 'e  
32, 'grad\_norm': 2.8158345222473145, 'learning\_rate': 1.8517866666666668e-05, '  
36, 'grad\_norm': 2.045478105545044, 'learning\_rate': 1.8512533333333336e-05, 'e  
15, 'grad\_norm': 2.6383323669433594, 'learning\_rate': 1.85072e-05, 'epoch': 0.2  
77, 'grad\_norm': 1.99264395236969, 'learning\_rate': 1.8501866666666666e-05, 'ep  
36, 'grad\_norm': 1.5785027742385864, 'learning\_rate': 1.8496533333333335e-05, '  
16, 'grad\_norm': 2.7614388465881348, 'learning\_rate': 1.8491200000000003e-05, '  
37, 'grad\_norm': 2.138429880142212, 'learning\_rate': 1.8485866666666667e-05, 'ep  
23, 'grad\_norm': 1.7542645931243896, 'learning\_rate': 1.8480533333333334e-05, '  
91, 'grad\_norm': 2.5335211753845215, 'learning\_rate': 1.8475200000000002e-05, '  
1, 'grad\_norm': 1.9256311655044556, 'learning\_rate': 1.8469866666666667e-05, 'ep  
58, 'grad\_norm': 1.9427850246429443, 'learning\_rate': 1.8464533333333336e-05, '  
35, 'grad\_norm': 1.3095953464508057, 'learning\_rate': 1.84592e-05, 'epoch': 0.2  
98, 'grad\_norm': 2.25569486618042, 'learning\_rate': 1.8453866666666667e-05, 'epo  
3, 'grad\_norm': 2.624191999435425, 'learning\_rate': 1.8448533333333335e-05, 'ep  
, 'grad\_norm': 1.8466638326644897, 'learning\_rate': 1.84432e-05, 'epoch': 0.233  
14, 'grad\_norm': 1.7888233661651611, 'learning\_rate': 1.8437866666666667e-05, 'e  
15, 'grad\_norm': 1.482723355293274, 'learning\_rate': 1.8432533333333337e-05, 'e  
59, 'grad\_norm': 2.0245442390441895, 'learning\_rate': 1.8427200000000002e-05, '  
96, 'grad\_norm': 1.7121461629867554, 'learning\_rate': 1.8421866666666667e-05, '  
57, 'grad\_norm': 1.3113757371902466, 'learning\_rate': 1.8416533333333336e-05, '  
52, 'grad\_norm': 2.0804638862609863, 'learning\_rate': 1.84112e-05, 'epoch': 0.2  
78, 'grad\_norm': 1.8422520160675049, 'learning\_rate': 1.8405866666666666e-05, '  
27, 'grad\_norm': 2.6599180698394775, 'learning\_rate': 1.8400533333333335e-05, '  
39, 'grad\_norm': 2.5066521167755127, 'learning\_rate': 1.8395200000000003e-05, '  
37, 'grad\_norm': 2.2836356163024902, 'learning\_rate': 1.8389866666666668e-05, '  
, 'grad\_norm': 2.3025505542755127, 'learning\_rate': 1.8384533333333333e-05, 'epo  
37, 'grad\_norm': 1.5027902126312256, 'learning\_rate': 1.8379200000000002e-05, '  
7, 'grad\_norm': 2.3068037033081055, 'learning\_rate': 1.8373866666666667e-05, 'ep  
72, 'grad\_norm': 1.722974181175232, 'learning\_rate': 1.8368533333333332e-05, 'e  
36, 'grad\_norm': 1.659519910812378, 'learning\_rate': 1.83632e-05, 'epoch': 0.24  
38, 'grad\_norm': 2.618854522705078, 'learning\_rate': 1.8357866666666667e-05, 'ep  
24, 'grad\_norm': 1.8366321325302124, 'learning\_rate': 1.8352533333333334e-05, '  
59, 'grad\_norm': 2.4209890365600586, 'learning\_rate': 1.83472e-05, 'epoch': 0.2  
9, 'grad\_norm': 1.605709195137024, 'learning\_rate': 1.8341866666666668e-05, 'ep  
, 'grad\_norm': 1.7040680646896362, 'learning\_rate': 1.8336533333333333e-05, 'ep

37 -> 'word\_embeddings': 1, 6722222808544212 -> 'learning\_rate': 1, 822126, 05 -> 'length': 0, 2

## Save the finetuned model

```
model.save_pretrained('./fine_tuned_model')
tokenizer.save_pretrained('./fine_tuned_model')
```

## Generate Text using the fine-tuned model:

```

import torch
from transformers import GPT2LMHeadModel, GPT2Tokenizer

# Path to your fine-tuned model directory
fine_tuned_model_path = './fine_tuned_model' # Adjust if necessary

# Initialize the model and tokenizer with your fine-tuned model
model = GPT2LMHeadModel.from_pretrained(fine_tuned_model_path)
tokenizer = GPT2Tokenizer.from_pretrained(fine_tuned_model_path)

# Function to generate text
def generate_text(prompt, max_length=50):
    # Encode the input prompt
    input_ids = tokenizer.encode(prompt, return_tensors='pt')

    # Create attention mask - to indicate which tokens should be attended to
    attention_mask = torch.ones(input_ids.shape, dtype=torch.long)

    # Generate text
    with torch.no_grad():
        output = model.generate(
            input_ids,
            attention_mask=attention_mask,
            max_length=max_length,
            num_return_sequences=1,
            do_sample=True, # Use sampling to generate more diverse outputs
            top_k=50,       # Limit the sampling pool to the top k tokens
            top_p=0.95,     # Use nucleus sampling
            temperature=0.7 # Control randomness in generation
        )

    # Decode the generated text
    generated_text = tokenizer.decode(output[0], skip_special_tokens=True)
    return generated_text

# Example usage
prompt = "The movie was"
generated = generate_text(prompt)
print(generated)

```

## Conclusion

In this notebook, I illustrated the process of fine-tuning a pre-trained model from Hugging Face using a specific dataset.

Note: To run the code, you will need a Hugging Face API key. Additionally, please be aware that the training process can be time-consuming, even when utilizing a GPU. For your convenience, this notebook has been saved as a PDF, allowing for easy sharing and reference.