

```
!pip install transformers datasets
!pip install accelerate
```

```
⇒ Requirement already satisfied: transformers in /usr/local/lib/python3.11/dist-pa
Requirement already satisfied: datasets in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: huggingface-hub<1.0,>=0.30.0 in /usr/local/lib/py
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-pac
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.11/di
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: tokenizers<0.22,>=0.21 in /usr/local/lib/python3.
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.11/d
Requirement already satisfied: tqdm>=4.27 in /usr/local/lib/python3.11/dist-pack
Requirement already satisfied: pyarrow>=8.0.0 in /usr/local/lib/python3.11/dist-
Requirement already satisfied: dill<0.3.8,>=0.3.0 in /usr/local/lib/python3.11/d
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages
Requirement already satisfied: xxhash in /usr/local/lib/python3.11/dist-packages
Requirement already satisfied: multiprocessing in /usr/local/lib/python3.11/dist-pa
Requirement already satisfied: fsspec>=2021.11.1 in /usr/local/lib/python3.11/di
Requirement already satisfied: aiohttp in /usr/local/lib/python3.11/dist-package
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/python3
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.11/dis
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-p
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.11/di
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/dis
Requirement already satisfied: yarll<2.0,>=1.17.0 in /usr/local/lib/python3.11/di
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/pyth
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-pa
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/d
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/d
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-pa
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: accelerate in /usr/local/lib/python3.11/dist-pack
Requirement already satisfied: numpy<3.0.0,>=1.17 in /usr/local/lib/python3.11/d
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist
Requirement already satisfied: psutil in /usr/local/lib/python3.11/dist-packages
Requirement already satisfied: pyyaml in /usr/local/lib/python3.11/dist-packages
Requirement already satisfied: torch>=2.0.0 in /usr/local/lib/python3.11/dist-pa
Requirement already satisfied: huggingface-hub>=0.21.0 in /usr/local/lib/python3
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.11/d
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: fsspec>=2023.5.0 in /usr/local/lib/python3.11/dis
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: tqdm>=4.42.1 in /usr/local/lib/python3.11/dist-pa
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/pyth
Requirement already satisfied: networkx in /usr/local/lib/python3.11/dist-packag
Requirement already satisfied: jinja2 in /usr/local/lib/python3.11/dist-packages
Collecting nvidia-cuda-nvrtc-cu12==12.4.127 (from torch>=2.0.0->accelerate)
  Downloading nvidia_cuda_nvrtc_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.1
```

```
Collecting nvidia-cuda-runtime-cu12==12.4.127 (from torch>=2.0.0->accelerate)
  Downloading nvidia_cuda_runtime_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl
Collecting nvidia-cuda-cupti-cu12==12.4.127 (from torch>=2.0.0->accelerate)
  Downloading nvidia_cuda_cupti_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl
Collecting nvidia-cudnn-cu12==9.1.0.70 (from torch>=2.0.0->accelerate)
  Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-manylinux2014_x86_64.whl.metadata
```

```
from transformers import (
    GPT2Tokenizer,
    GPT2LMHeadModel,
    TextDataset,
    DataCollatorForLanguageModeling,
    Trainer,
    TrainingArguments,
)
DataCollatorForLanguageModeling,
Trainer,
TrainingArguments
```

```
 transformers.training_args.TrainingArguments
```

```
model_name = "gpt2"
```

```
tokenizer = GPT2Tokenizer.from_pretrained(model_name)
model = GPT2LMHeadModel.from_pretrained(model_name)
```

```
tokenizer.pad_token = tokenizer.eos_token
model.resize_token_embeddings(len(tokenizer))
```

```

```

```
from google.colab import files
uploaded = files.upload()
```



```
def load_dataset(file_path, tokenizer):
    return TextDataset(
        tokenizer=tokenizer,
        file_path=file_path,
        block_size=128
    )

train_dataset = load_dataset("custom_data.txt", tokenizer)
data_collator = DataCollatorForLanguageModeling(
    tokenizer=tokenizer, mlm=False,
)

 /usr/local/lib/python3.11/dist-packages/transformers/data/datasets/language_modeling.py:100:
warnings.warn(

# Check if the file exists and has content
import os
if not os.path.exists("custom_data.txt") or os.stat("custom_data.txt").st_size == 0:
    print("Error: custom_data.txt not found or is empty.")
    # You might want to add code here to handle the missing/empty file,
    # like prompting the user to upload the file or creating a dummy file.
else:
    def load_dataset(file_path, tokenizer):
        return TextDataset(
            tokenizer=tokenizer,
            file_path=file_path,
            block_size=128
        )

    train_dataset = load_dataset("custom_data.txt", tokenizer)

# Add a check for the dataset size
if len(train_dataset) == 0:
    print("Error: The loaded dataset is empty. Please check custom_data.txt content.")
else:
    data_collator = DataCollatorForLanguageModeling(
        tokenizer=tokenizer, mlm=False,
    )

    training_args = TrainingArguments(
        output_dir="./gpt2-finetuned",
        overwrite_output_dir=True,
        num_train_epochs=3,
        per_device_train_batch_size=2,
        save_steps=500,
        save_total_limit=2,
        logging_dir='./logs',
    )

    trainer = Trainer(
        model=model,
        args=training_args,
        data_collator=data_collator,
```

```

        train_dataset=train_dataset,
    )

    trainer.train()

```

⇒ Error: The loaded dataset is empty. Please check custom_data.txt content.
 /usr/local/lib/python3.11/dist-packages/transformers/data/datasets/language_modeling.py:100: UserWarning: The loaded dataset is empty. Please check custom_data.txt content.
 warnings.warn(

```

trainer.save_model("./gpt2-finetuned")
tokenizer.save_pretrained("./gpt2-finetuned")

```

⇒ ('./gpt2-finetuned/tokenizer_config.json',
 './gpt2-finetuned/special_tokens_map.json',
 './gpt2-finetuned/vocab.json',
 './gpt2-finetuned/merges.txt',
 './gpt2-finetuned/added_tokens.json')

```

from transformers import pipeline

```

```

generator = pipeline('text-generation', model="./gpt2-finetuned", tokenizer="./gpt2-finetuned")
prompt = "Once upon a time"
results = generator(prompt, max_length=100, num_return_sequences=1)

print(results[0]['generated_text'])

```

⇒ Device set to use cpu
 Truncation was not explicitly activated but `max_length` is provided a specific value of 100. It is recommended to explicitly activate it with `truncation=True`.
 Once upon a time, what seems to you that the lightest person could make a living

A little later, with the arrival of the angel of death by his side, it was brought

As a member of the Quidditch team, the one who brought

