```
!pip  install  evaluate
```

```
Collecting evaluate
  Downloading evaluate-0.4.6-py3-none-any.whl.metadata (9.5 kB)
Requirement already satisfied: datasets>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from evaluate) (3.6.0)
Requirement already satisfied: numpy>=1.17 in /usr/local/lib/python3.11/dist-packages (from evaluate) (1.26.4)
Requirement already satisfied: dill in /usr/local/lib/python3.11/dist-packages (from evaluate) (0.3.8)
Requirement already satisfied: pandas in /usr/local/lib/python3.11/dist-packages (from evaluate) (2.2.3)
Requirement already satisfied: requests>=2.19.0 in /usr/local/lib/python3.11/dist-packages (from evaluate) (2.32.4)
Requirement already satisfied: tqdm>=4.62.1 in /usr/local/lib/python3.11/dist-packages (from evaluate) (4.67.1)
Requirement already satisfied: xxhash in /usr/local/lib/python3.11/dist-packages (from evaluate) (3.5.0)
Requirement already satisfied: multiprocess in /usr/local/lib/python3.11/dist-packages (from evaluate) (0.70.16)
Requirement already satisfied: fsspec>=2021.05.0 in /usr/local/lib/python3.11/dist-packages (from fsspec[http]>=2021.05.0->evaluate) (2025.
Requirement already satisfied: huggingface-hub>=0.7.0 in /usr/local/lib/python3.11/dist-packages (from evaluate) (0.33.1)
Requirement already satisfied: packaging in /usr/local/lib/python3.11/dist-packages (from evaluate) (25.0)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from datasets>=2.0.0->evaluate) (3.18.0)
Requirement already satisfied: pyarrow>=15.0.0 in /usr/local/lib/python3.11/dist-packages (from datasets>=2.0.0->evaluate) (19.0.1)
Collecting fsspec>=2021.05.0 (from fsspec[http]>=2021.05.0->evaluate)
  Downloading fsspec-2025.3.0-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.11/dist-packages (from datasets>=2.0.0->evaluate) (6.0.2)
Requirement already satisfied: aiohttp!=4.0.0a0,!=4.0.0a1 in /usr/local/lib/python3.11/dist-packages (from fsspec[http]>=2021.05.0->evaluat
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.7.0->evaluate
Requirement already satisfied: hf-xet<2.0.0,>=1.1.2 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub>=0.7.0->evaluate) (1.1
Requirement already satisfied: mkl_fft in /usr/local/lib/python3.11/dist-packages (from numpy>=1.17->evaluate) (1.3.8)
Requirement already satisfied: mkl_random in /usr/local/lib/python3.11/dist-packages (from numpy>=1.17->evaluate) (1.2.4)
Requirement already satisfied: mkl_umath in /usr/local/lib/python3.11/dist-packages (from numpy>=1.17->evaluate) (0.1.1)
Requirement already satisfied: mkl in /usr/local/lib/python3.11/dist-packages (from numpy>=1.17->evaluate) (2025.2.0)
Requirement already satisfied: tbb4py in /usr/local/lib/python3.11/dist-packages (from numpy>=1.17->evaluate) (2022.2.0)
Requirement already satisfied: mkl-service in /usr/local/lib/python3.11/dist-packages (from numpy>=1.17->evaluate) (2.4.1)
Requirement already satisfied: charset_normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests>=2.19.0->evaluate) (3.4.2
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests>=2.19.0->evaluate) (3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests>=2.19.0->evaluate) (2.5.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests>=2.19.0->evaluate) (2025.6.15)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.11/dist-packages (from pandas->evaluate) (2.9.0.post0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas->evaluate) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas->evaluate) (2025.2)
Requirement already satisfied: aiohappyeyeballs>=2.5.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[
Requirement already satisfied: aiosignal>=1.1.2 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http]>=
Requirement already satisfied: attrs>=17.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http]>=202
Requirement already satisfied: frozenlist>=1.1.1 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http]>
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http]>=
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fsspec[http]>
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.8.2->pandas->evaluate) (1.17.0)
Requirement already satisfied: intel-openmp<2026,>=2024 in /usr/local/lib/python3.11/dist-packages (from mkl->numpy>=1.17->evaluate) (2024.
Requirement already satisfied: tbb==2022.* in /usr/local/lib/python3.11/dist-packages (from mkl->numpy>=1.17->evaluate) (2022.2.0)
Requirement already satisfied: tcmlib==1.* in /usr/local/lib/python3.11/dist-packages (from tbb==2022.*->mkl->numpy>=1.17->evaluate) (1.4.0
Requirement already satisfied: intel-cmplr-lib-rt in /usr/local/lib/python3.11/dist-packages (from mkl_umath->numpy>=1.17->evaluate) (2024.
Requirement already satisfied: intel-cmplr-lib-ur==2024.2.0 in /usr/local/lib/python3.11/dist-packages (from intel-openmp<2026,>=2024->mkl-
Downloading evaluate-0.4.6-py3-none-any.whl (84 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 84.1/84.1 kB 2.7 MB/s eta 0:00:00
Downloading fsspec-2025.3.0-py3-none-any.whl (193 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 193.6/193.6 kB 7.2 MB/s eta 0:00:00
Installing collected packages: fsspec, evaluate
  Attempting uninstall: fsspec
    Found existing installation: fsspec 2025.5.1
    Uninstalling fsspec-2025.5.1:
      Successfully uninstalled fsspec-2025.5.1
ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is the source of
bigframes 2.8.0 requires google-cloud-bigquery-storage<3.0.0,>=2.30.0, which is not installed
```

```
import  torch
from  transformers  import  (
        AutoTokenizer,  AutoModelForCausalLM,
        DataCollatorForLanguageModeling,
        TrainingArguments,  Trainer,  pipeline,
        EarlyStoppingCallback
)
from  ray  import  tune
from  datasets  import  load_dataset
import  pandas  as  pd
import  matplotlib.pyplot  as  plt

device  =  "cuda"  if  torch.cuda.is_available()  else  "cpu"
```

```
2025-09-24 10:19:39.109533: E external/local_xla/xla/stream_executor/cuda/cuda_fft.cc:477] Unable to register cuFFT factory: Attempting to re
WARNING: All log messages before absl::InitializeLog() is called are written to STDERR
E0000 00:00:1758709179.444096      36 cuda_dnn.cc:8310] Unable to register cuDNN factory: Attempting to register factory for plugin cuDNN whe
E0000 00:00:1758709179.552106      36 cuda_blas.cc:1418] Unable to register cuBLAS factory: Attempting to register factory for plugin cuBLAS
```

```
#  Data  &  Model
model_id  =  "HuggingFaceTB/SmolLM2-135M"
tokenizer  =  AutoTokenizer.from_pretrained(model_id)
tokenizer.pad_token  =  tokenizer.eos_token
```

```
raw_datasets = load_dataset("ag_news")
train_dataset = raw_datasets["train"]
eval_dataset = raw_datasets["test"]

def tokenize_function(batch):
    return tokenizer(batch["text"], truncation=True)

tokenized_train = train_dataset.map(tokenize_function, batched=True, remove_columns=["text"])
tokenized_eval = eval_dataset.map(tokenize_function, batched=True, remove_columns=["text"])

data_collator = DataCollatorForLanguageModeling(tokenizer=tokenizer, mlm=False)
```

```
tokenizer_config.json: 0.00B [00:00, ?B/s]
vocab.json: 0.00B [00:00, ?B/s]
merges.txt: 0.00B [00:00, ?B/s]
tokenizer.json: 0.00B [00:00, ?B/s]
special_tokens_map.json:  0%|          | 0.00/831 [00:00<?, ?B/s]
README.md: 0.00B [00:00, ?B/s]
train-00000-of-00001.parquet:  0%|          | 0.00/18.6M [00:00<?, ?B/s]
test-00000-of-00001.parquet:  0%|          | 0.00/1.23M [00:00<?, ?B/s]
Generating train split:  0%|          | 0/120000 [00:00<?, ? examples/s]
```

```
# draw loss curve
def plot_loss(trainer, title):
    logs = trainer.state.log_history
    df = pd.DataFrame(logs)
    train_loss = df.dropna(subset=["loss"])
    eval_loss = df.dropna(subset=["eval_loss"])
    plt.figure(figsize=(7,5))
    plt.plot(train_loss["step"], train_loss["loss"], label="Training Loss")
    plt.plot(eval_loss["step"], eval_loss["eval_loss"], label="Validation Loss")
    plt.xlabel("Step")
    plt.ylabel("Loss")
    plt.title(title)
    plt.legend()
    plt.show()
```

```
# text generation function
def show_output(trainer, title):
    gen_pipe = pipeline("text-generation", model=trainer.model, tokenizer=tokenizer,
                        device=0 if device=="cuda" else -1)
    prompt = "Wall"
    print(f"\n=== {title} ===")
    print(gen_pipe(prompt, max_new_tokens=30)[0]["generated_text"])
```

```
# text generation function
def show_output2(trainer, title, prompts, max_new_tokens=40):
    gen_pipe = pipeline(
        "text-generation",
        model=trainer.model,
        tokenizer=tokenizer,
        device=0 if device=="cuda" else -1
    )
    print(f"\n=== {title} ===")
    for prompt in prompts:
        generated = gen_pipe(prompt, max_new_tokens=max_new_tokens)[0]["generated_text"]
        print(f"Prompt: {prompt}\nGenerated: {generated}\n{'-'*40}")
```

```
# 使用最优超参数在全量数据训练最终 Good Run，并加早停
final_args = TrainingArguments(
    output_dir="good-run-final",
    per_device_train_batch_size=8,
    learning_rate= 2e-4,
    num_train_epochs= 2,
    weight_decay= 0.01,
    lr_scheduler_type="cosine",

    eval_strategy="steps",          # 每轮评估一次
    eval_steps=1000,
    save_strategy="steps",                      # 每轮保存 checkpoint
    save_steps=1000,
    load_best_model_at_end=True,          # 训练结束加载验证集最优模型
    metric_for_best_model="eval_loss",    # 指标：最小化 eval_loss

    logging_steps=1000,
    save_total_limit=2,                           # 最多保留 2 个 checkpoint，节省空间
    report_to="none"
)

early_stop_callback = EarlyStoppingCallback(
```

```
        early_stopping_patience=1,
        early_stopping_threshold=0.0
)

final_good_trainer  =  Trainer(
        model=AutoModelForCausalLM.from_pretrained(model_id).to(device),
        args=final_args,
        train_dataset=tokenized_train,              #  全量训练
        eval_dataset=tokenized_eval,
        processing_class=tokenizer,
        data_collator=data_collator,
        callbacks=[early_stop_callback]
)

final_good_trainer.train()
plot_loss(final_good_trainer,  "Good  Run:  Healthy  Training  (Final)")
show_output(final_good_trainer,  "Good  Run  Output  (Final)")
```