```
total_frozen_params=210236928
trainable parameters: 28366848
/tmp/ipython-input-32-347286690.py:16: FutureWarning: `torch.cuda.amp.GradScaler(args...)` is deprecated. Please use `torch.amp.GradScaler('cuda', args...)` instead.
  self.scaler = GradScaler()
/tmp/ipython-input-32-347286690.py:38: UserWarning: Argument(s) 'always_apply' are not valid for transform Normalize
  A.Normalize(mean=[0.5,0.5,0.5],std=[0.5,0.5,0.5],always_apply=True),
DistillationTrainer initialized. Ready to create a distilled dataset.
Loading pre-trained model for distillation...
Creating distilled dataset with top 5.0% hardest examples...
Calculating losses for distillation: 100%                                    782/782 [06:35<00:00,  2.40it/s]
Original dataset size: 100000
Distilled dataset size: 5000

--- Training on DISTILLED dataset from scratch ---
total_frozen_params=210236928
trainable parameters: 28366848
/tmp/ipython-input-32-347286690.py:16: FutureWarning: `torch.cuda.amp.GradScaler(args...)` is deprecated. Please use `torch.amp.GradScaler('cuda', args...)` instead.
  self.scaler = GradScaler()
/tmp/ipython-input-32-347286690.py:38: UserWarning: Argument(s) 'always_apply' are not valid for transform Normalize
  A.Normalize(mean=[0.5,0.5,0.5],std=[0.5,0.5,0.5],always_apply=True),
validating: 100%                                    20/20 [25:00<00:00, 62.77s/it]

train loss: 8.889: 100%                                    40/40 [00:25<00:00,  2.44it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 8.851: 100%                                    79/79 [00:30<00:00,  3.93it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
   train_loss train_perplexity  val_loss val_perplexity
0    8.762093      6387.468973  8.731712    6196.328365
saving best model...

train loss: 7.975: 100%                                    40/40 [00:46<00:00,  3.06it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 7.972: 100%                                    79/79 [00:30<00:00,  3.78it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
   train_loss train_perplexity  val_loss val_perplexity
1    8.508478      4956.611654  7.826664    2506.553611
saving best model...
unfreezing GPT2 entirely...
total_frozen_params=0

train loss: 6.058: 100%                                    40/40 [00:36<00:00,  2.90it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 6.007: 100%                                    79/79 [00:30<00:00,  3.84it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
   train_loss train_perplexity  val_loss val_perplexity
2    7.08571      1194.771608  5.698996     298.567547
saving best model...

train loss: 5.890: 100%                                    40/40 [00:46<00:00,  2.97it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 5.620: 100%                                    79/79 [00:30<00:00,  3.81it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
   train_loss train_perplexity  val_loss val_perplexity
3    6.293235       540.900246  5.111638     165.941988
saving best model...

train loss: 5.837: 100%                                    40/40 [00:46<00:00,  2.90it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 5.123: 100%                                    79/79 [00:30<00:00,  3.84it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
   train_loss train_perplexity val_loss val_perplexity
4    5.759018       317.036972  4.63689     103.222792
saving best model...

train loss: 4.972: 100%                                    40/40 [00:47<00:00,  2.96it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 4.783: 100%                                    79/79 [00:30<00:00,  3.80it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
   train_loss train_perplexity  val_loss val_perplexity
5    5.305374       201.416249  4.341612      76.831289
saving best model...

train loss: 4.766: 100%                                    40/40 [00:47<00:00,  2.86it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 4.361: 100%                                    79/79 [00:30<00:00,  3.73it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
   train_loss train_perplexity  val_loss val_perplexity
6    4.921994       137.276027  3.798373      44.628513
saving best model...

train loss: 4.449: 100%                                    40/40 [00:47<00:00,  2.80it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
```

```
valid loss: 4.028: 100%                                    79/79 [00:30<00:00,  3.90it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
   train_loss train_perplexity  val_loss val_perplexity
7   4.594502         98.938899  3.470978      32.168183
saving best model...

train loss: 4.089: 100%                                    40/40 [00:46<00:00,  2.93it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 3.905: 100%                                    79/79 [00:30<00:00,  3.78it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
   train_loss train_perplexity  val_loss val_perplexity
8   4.337503         76.516227  3.341515      28.261916
saving best model...

train loss: 3.972: 100%                                    40/40 [00:46<00:00,  2.99it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 3.786: 100%                                    79/79 [00:30<00:00,  3.71it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
   train_loss train_perplexity  val_loss val_perplexity
9   4.138424         62.703926  3.240253      25.540194
saving best model...

train loss: 4.047: 100%                                    40/40 [00:47<00:00,  2.82it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 3.639: 100%                                    79/79 [00:30<00:00,  3.91it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
    train_loss train_perplexity  val_loss val_perplexity
10   3.980475         53.542471   3.20237      24.590746
saving best model...

train loss: 3.618: 100%                                    40/40 [00:46<00:00,  2.95it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 3.646: 100%                                    79/79 [00:30<00:00,  3.76it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
    train_loss train_perplexity  val_loss val_perplexity
11   3.834485         46.269586  3.165132      23.691862
saving best model...

train loss: 4.051: 100%                                    40/40 [00:46<00:00,  2.88it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 3.583: 100%                                    79/79 [00:30<00:00,  4.00it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
    train_loss train_perplexity  val_loss val_perplexity
12   3.731618          41.74658  3.144382      23.205325
saving best model...

train loss: 3.975: 100%                                    40/40 [00:47<00:00,  2.91it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 3.533: 100%                                    79/79 [00:30<00:00,  3.88it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
    train_loss train_perplexity  val_loss val_perplexity
13   3.636651         37.964491  3.122258      22.697562
saving best model...

train loss: 3.456: 100%                                    40/40 [00:46<00:00,  2.92it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 3.512: 100%                                    79/79 [00:30<00:00,  3.88it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
    train_loss train_perplexity  val_loss val_perplexity
14   3.538034         34.399213  3.151534        23.3719

train loss: 3.604: 100%                                    40/40 [00:25<00:00,  3.02it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 3.512: 100%                                    79/79 [00:30<00:00,  3.82it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
    train_loss train_perplexity  val_loss val_perplexity
15   3.472888         32.229687  3.116772      22.573402
saving best model...

train loss: 3.189: 100%                                    40/40 [00:46<00:00,  2.88it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():

valid loss: 3.512: 100%                                    79/79 [00:30<00:00,  3.81it/s]

/tmp/ipython-input-32-347286690.py:88: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
    train_loss train_perplexity  val_loss val_perplexity
16   3.413554         30.372987  3.127696       22.82135

train loss: 3.288: 100%                                    40/40 [00:25<00:00,  2.80it/s]

/tmp/ipython-input-32-347286690.py:58: FutureWarning: `torch.cuda.amp.autocast(args...)` is deprecated. Please use `torch.amp.autocast('cuda', args...)` instead.
  with autocast():
```

```
    train_loss train_perplexity  val_loss val_perplexity
17   3.385379        29.529189  3.128729      22.844932
```
```
    train_loss train_perplexity  val_loss val_perplexity
18   3.373603        29.183474  3.134365       22.97405
```
```
    train_loss train_perplexity  val_loss val_perplexity
19   3.348445        28.458456  3.132233      22.925104
Metrics for the model trained on the distilled dataset:
    train_loss train_perplexity  val_loss val_perplexity
0    8.762093      6387.468973  8.731712    6196.328365
1    8.508478      4956.611654  7.826664    2506.553611
2     7.08571      1194.771608  5.698996     298.567547
3    6.293235       540.900246  5.111638     165.941988
4    5.759018       317.036972   4.63689     103.222792
5    5.305374       201.416249  4.341612      76.831289
6    4.921994       137.276027  3.798373      44.628513
7    4.594502        98.938899  3.470978      32.168183
8    4.337503        76.516227  3.341515      28.261916
9    4.138424        62.703926  3.240253      25.540194
10   3.980475        53.542471   3.20237      24.590746
11   3.834485        46.269586  3.165132      23.691862
12   3.731618         41.74658  3.144382      23.205325
13   3.636651        37.964491  3.122258      22.697562
14   3.538034        34.399213  3.151534        23.3719
15   3.472888        32.229687  3.116772      22.573402
16   3.413554        30.372987  3.127696       22.82135
17   3.385379        29.529189  3.128729      22.844932
18   3.373603        29.183474  3.134365       22.97405
19   3.348445        28.458456  3.132233      22.925104


--- COMPARISON ---
Full Dataset Training Metrics:
Empty DataFrame
Columns: [train_loss, train_perplexity, val_loss, val_perplexity]
Index: []

Distilled Dataset Training Metrics:
    train_loss train_perplexity  val_loss val_perplexity
0    8.762093      6387.468973  8.731712    6196.328365
1    8.508478      4956.611654  7.826664    2506.553611
2     7.08571      1194.771608  5.698996     298.567547
3    6.293235       540.900246  5.111638     165.941988
4    5.759018       317.036972   4.63689     103.222792
5    5.305374       201.416249  4.341612      76.831289
6    4.921994       137.276027  3.798373      44.628513
7    4.594502        98.938899  3.470978      32.168183
8    4.337503        76.516227  3.341515      28.261916
9    4.138424        62.703926  3.240253      25.540194
10   3.980475        53.542471   3.20237      24.590746
11   3.834485        46.269586  3.165132      23.691862
12   3.731618         41.74658  3.144382      23.205325
13   3.636651        37.964491  3.122258      22.697562
14   3.538034        34.399213  3.151534        23.3719
15   3.472888        32.229687  3.116772      22.573402
16   3.413554        30.372987  3.127696       22.82135
17   3.385379        29.529189  3.128729      22.844932
18   3.373603        29.183474  3.134365       22.97405
19   3.348445        28.458456  3.132233      22.925104
```