```
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 30.0% hardest examples...
```

Calculating losses for distillation: 100% 782/782 [28:38<00:00, 1.84s/it]

```
Original dataset size: 100000
Distilled dataset size: 30000

--- 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% 10/10 [31:11<00:00, 170.77s/it]

train loss: 6.843: 100% 235/235 [02:01<00:00, 2.10it/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: 6.375: 100% 79/79 [03:05<00:00, 6.19s/it]

```
/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   7.965839      2880.845148  6.222074     503.747119
saving best model...
```

train loss: 6.419: 100% 235/235 [02:05<00:00, 2.12it/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: 5.981: 100% 79/79 [00:29<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
1   6.468854       644.744214  5.752325     314.922062
saving best model...
unfreezing GPT2 entirely...
total_frozen_params=0
```

train loss: 3.838: 100% 235/235 [02:14<00:00, 2.74it/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: 3.470: 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
2   4.749463       115.522185  3.003615      20.158285
saving best model...
```

train loss: 3.662: 100% 235/235 [02:21<00:00, 2.40it/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: 3.119: 100% 79/79 [00:29<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
3   3.576566        35.750552  2.718361      15.155464
saving best model...
```

train loss: 3.373: 100% 235/235 [02:24<00:00, 2.81it/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: 2.994: 100% 79/79 [00:29<00:00, 3.94it/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   3.261669        26.093044  2.627578      13.840204
saving best model...
```

train loss: 3.170: 100% 235/235 [02:24<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: 2.904: 100% 79/79 [00:29<00:00, 4.10it/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   3.067882        21.496323  2.543022      12.718051
saving best model...
```

train loss: 3.185: 100% 235/235 [02:24<00:00, 2.78it/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: 2.883: 100% 79/79 [00:29<00:00, 3.92it/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   2.922528        18.588212  2.523882      12.476933
saving best model...
```

train loss: 2.922: 100% 235/235 [02:24<00:00, 2.78it/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: 2.891: 100% 79/79 [00:29<00:00, 3.92it/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   2.803505        16.502381  2.507999      12.280331
saving best model...
```

train loss: 2.732: 100% 235/235 [02:26<00:00, 2.53it/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: 2.905: 100% 79/79 [00:29<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
8   2.722177        15.213406   2.5098      12.302469
```

train loss: 2.467: 100% 235/235 [02:02<00:00, 2.50it/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: 2.905: 100% 79/79 [00:29<00:00, 3.96it/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   2.680049        14.585812  2.515511      12.372926
Metrics for the model trained on the distilled dataset:
  train_loss train_perplexity  val_loss val_perplexity
0   7.965839      2880.845148  6.222074     503.747119
1   6.468854       644.744214  5.752325     314.922062
2   4.749463       115.522185  3.003615      20.158285
3   3.576566        35.750552  2.718361      15.155464
4   3.261669        26.093044  2.627578      13.840204
5   3.067882        21.496323  2.543022      12.718051
6   2.922528        18.588212  2.523882      12.476933
7   2.803505        16.502381  2.507999      12.280331
8   2.722177        15.213406    2.5098      12.302469
9   2.680049        14.585812  2.515511      12.372926


--- 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   7.965839      2880.845148  6.222074     503.747119
1   6.468854       644.744214  5.752325     314.922062
2   4.749463       115.522185  3.003615      20.158285
3   3.576566        35.750552  2.718361      15.155464
4   3.261669        26.093044  2.627578      13.840204
5   3.067882        21.496323  2.543022      12.718051
6   2.922528        18.588212  2.523882      12.476933
7   2.803505        16.502381  2.507999      12.280331
8   2.722177        15.213406    2.5098      12.302469
9   2.680049        14.585812  2.515511      12.372926
```