Here I have shown the performance of my two approaches: BART-A and BART-B, on train and valid datasets.

**BART-A: Finetuning with Augmentation**

```
Num examples = 14364
  Num Epochs = 3
  Instantaneous batch size per device = 16
  Total train batch size (w. parallel, distributed & accumulation) = 16
  Gradient Accumulation steps = 1
  Total optimization steps = 2694
```

[2694/2694 38:07, Epoch 3/3]

| Step | Training Loss | Validation Loss |
|---|---|---|
| 100 | 5.360100 | 4.292557 |
| 200 | 2.411200 | 1.818061 |
| 300 | 0.346900 | 0.329011 |
| 400 | 0.343800 | 0.349541 |
| 500 | 0.324200 | 0.294574 |
| 600 | 0.292600 | 0.356346 |
| 700 | 0.264700 | 0.252039 |
| 800 | 0.264000 | 0.244744 |
| 900 | 0.215900 | 0.234657 |
| 1000 | 0.197200 | 0.285489 |
| 1100 | 0.163700 | 0.234189 |
| 1200 | 0.160200 | 0.243582 |
| 1300 | 0.145900 | 0.236006 |
| 1400 | 0.132800 | 0.237120 |
| 1500 | 0.140000 | 0.229246 |
| 1600 | 0.139800 | 0.229793 |
| 1700 | 0.131200 | 0.239360 |

| Step | Training Loss | Validation Loss |
|---|---|---|
| 1800 | 0.083600 | 0.217151 |
| 1900 | 0.104500 | 0.237082 |
| 2000 | 0.089100 | 0.217283 |
| 2100 | 0.093200 | 0.227747 |
| 2200 | 0.083600 | 0.219020 |
| 2300 | 0.124800 | 0.212462 |
| 2400 | 0.067300 | 0.228526 |
| 2500 | 0.084300 | 0.216207 |
| 2600 | 0.070100 | 0.219425 |

**BART-B: Finetuning without Augmentation**

```
Num examples = 7182
  Num Epochs = 3
  Instantaneous batch size per device = 16
  Total train batch size (w. parallel, distributed & accumulation) = 16
  Gradient Accumulation steps = 1
  Total optimization steps = 1347
```

[1347/1347 17:16, Epoch 3/3]

| Step | Training Loss | Validation Loss |
|---|---|---|
| 100 | 5.066300 | 4.120363 |
| 200 | 2.238700 | 1.731243 |
| 300 | 0.174500 | 0.178013 |
| 400 | 0.149500 | 0.168917 |
| 500 | 0.137100 | 0.199508 |
| 600 | 0.136900 | 0.173199 |
| 700 | 0.103000 | 0.152560 |
| 800 | 0.118100 | 0.143438 |

| Step | Training Loss | Validation Loss |
|---|---|---|
| 900 | 0.090200 | 0.142074 |
| 1000 | 0.059500 | 0.152700 |
| 1100 | 0.059300 | 0.139933 |
| 1200 | 0.058800 | 0.140943 |
| 1300 | 0.074100 | 0.130499 |