7b pretrain acc: **0.4621**


```
Probabilities for choices: {'0': 0.45703125, '1': 0.54296875}
Predicted Label: 1, True Label: 0
Evaluating:  97%|                                          | 269/277 [02:47<00:05,  1.44it/s]
Probabilities for choices: {'0': 0.4072265625, '1': 0.5927734375}
Predicted Label: 1, True Label: 1
Evaluating:  97%|                                          | 270/277 [02:48<00:04,  1.55it/s]
Probabilities for choices: {'0': 0.43017578125, '1': 0.56982421875}
Predicted Label: 1, True Label: 1
Evaluating:  98%|                                          | 271/277 [02:48<00:03,  1.64it/s]
Probabilities for choices: {'0': 0.38671875, '1': 0.61328125}
Predicted Label: 1, True Label: 1
Evaluating:  98%|                                          | 272/277 [02:49<00:02,  1.79it/s]
Probabilities for choices: {'0': 0.4111328125, '1': 0.5888671875}
Predicted Label: 1, True Label: 0
Evaluating:  99%|                                          | 273/277 [02:49<00:02,  1.81it/s]
Probabilities for choices: {'0': 0.486328125, '1': 0.513671875}
Predicted Label: 1, True Label: 0
Evaluating:  99%|                                          | 274/277 [02:50<00:01,  1.73it/s]
Probabilities for choices: {'0': 0.416748046875, '1': 0.5830078125}
Predicted Label: 1, True Label: 0
Evaluating:  99%|                                          | 275/277 [02:50<00:01,  1.68it/s]
Probabilities for choices: {'0': 0.490234375, '1': 0.509765625}
Predicted Label: 1, True Label: 0
Evaluating: 100%|                                          | 276/277 [02:51<00:00,  1.73it/s]
Probabilities for choices: {'0': 0.31396484375, '1': 0.68603515625}
Predicted Label: 1, True Label: 1
Evaluating: 100%|                                          | 277/277 [02:52<00:00,  1.61it/s]
Evaluation completed. Accuracy: 0.4621 (128/277)
Logits saved!
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/RTE#
```

13b pretrain acc: **0.5740**


```
Probabilities for choices: {'0': 0.52734375, '1': 0.47265625}
Predicted Label: 0, True Label: 0
Evaluating:  97%|                                          | 269/277 [01:22<00:02,  2.80it/s]
Probabilities for choices: {'0': 0.5078125, '1': 0.4921875}
Predicted Label: 0, True Label: 1
Evaluating:  97%|                                          | 270/277 [01:23<00:02,  3.03it/s]
Probabilities for choices: {'0': 0.6533203125, '1': 0.346923828125}
Predicted Label: 0, True Label: 1
Evaluating:  98%|                                          | 271/277 [01:23<00:01,  3.23it/s]
Probabilities for choices: {'0': 0.59619140625, '1': 0.403564453125}
Predicted Label: 0, True Label: 1
Evaluating:  98%|                                          | 272/277 [01:23<00:01,  3.59it/s]
Probabilities for choices: {'0': 0.53515625, '1': 0.46484375}
Predicted Label: 0, True Label: 0
Evaluating:  99%|                                          | 273/277 [01:23<00:01,  3.66it/s]
Probabilities for choices: {'0': 0.5, '1': 0.5}
Predicted Label: 0, True Label: 0
Evaluating:  99%|                                          | 274/277 [01:24<00:00,  3.51it/s]
Probabilities for choices: {'0': 0.640625, '1': 0.359375}
Predicted Label: 0, True Label: 0
Evaluating:  99%|                                          | 275/277 [01:24<00:00,  3.37it/s]
Probabilities for choices: {'0': 0.529296875, '1': 0.470703125}
Predicted Label: 0, True Label: 0
Evaluating: 100%|                                          | 276/277 [01:24<00:00,  3.44it/s]
Probabilities for choices: {'0': 0.394287109375, '1': 0.60595703125}
Predicted Label: 1, True Label: 1
Evaluating: 100%|                                          | 277/277 [01:25<00:00,  3.25it/s]
Evaluation completed. Accuracy: 0.5740 (159/277)
Logits saved!
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/RTE#
```

7b loratune acc: **0.8628**


```
Probabilities for choices: {'0': 0.85986328125, '1': 0.140380859375}
Predicted Label: 0, True Label: 0
Evaluating:  97%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 269/277 [01:24<00:02,  2.91it/s]
Probabilities for choices: {'0': 0.1800537109375, '1': 0.81982421875}
Predicted Label: 1, True Label: 1
Evaluating:  97%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 270/277 [01:24<00:02,  3.14it/s]
Probabilities for choices: {'0': 0.5849609375, '1': 0.414794921875}
Predicted Label: 0, True Label: 1
Evaluating:  98%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 271/277 [01:24<00:01,  3.33it/s]
Probabilities for choices: {'0': 0.219970703125, '1': 0.77978515625}
Predicted Label: 1, True Label: 1
Evaluating:  98%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 272/277 [01:25<00:01,  3.68it/s]
Probabilities for choices: {'0': 0.91943359375, '1': 0.08038330078125}
Predicted Label: 0, True Label: 0
Evaluating:  99%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 273/277 [01:25<00:01,  3.73it/s]
Probabilities for choices: {'0': 0.89306640625, '1': 0.106689453125}
Predicted Label: 0, True Label: 0
Evaluating:  99%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 274/277 [01:25<00:00,  3.56it/s]
Probabilities for choices: {'0': 0.943359375, '1': 0.056640625}
Predicted Label: 0, True Label: 0
Evaluating:  99%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 275/277 [01:25<00:00,  3.44it/s]
Probabilities for choices: {'0': 0.76611328125, '1': 0.2337646484375}
Predicted Label: 0, True Label: 0
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 276/277 [01:26<00:00,  3.54it/s]
Probabilities for choices: {'0': 0.0009698867797851562, '1': 0.9990234375}
Predicted Label: 1, True Label: 1
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 277/277 [01:26<00:00,  3.20it/s]
Evaluation completed. Accuracy: 0.8628 (239/277)
Logits saved!
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/RTE#
```

Proxy tuning acc: **0.8448**

Consistent proxy tuning acc: **0.8484**

13b loratune acc: **0.8664**

Gaussian Process tuning acc: **0.8628**

230 / 2490 (9.24%)

Gaussian Process with filter tuning acc: **0.8448**

Config: input_threshold=0.149, output_threshold=0.4

46 / 2490 (1.85%)