

7b pretrain acc: **0.4621**

```
root@e8e99ab71e21: ~/Work:
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**

```
root@e8e99ab71e21: ~/Work:
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**

```
root@e8e99ab71e21: ~
Probabilities for choices: {'0': 0.85986328125, '1': 0.14013671875}
Predicted Label: 0, True Label: 0
Evaluating: 97% | 269/277 [01:24:00:02, 2.91it/s]
Probabilities for choices: {'0': 0.1800537109375, '1': 0.81984621875}
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': 0.08033330078125}
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': 0.23376640625}
Predicted Label: 0, True Label: 0
Evaluating: 100% | 276/277 [01:26:00:00, 3.54it/s]
Probabilities for choices: {'0': 0.009698867777851562, '1': 0.9903011328125}
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**

```
root@e8e99ab71e21: ~  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.020332353189587593, '1': 0.9796676635742188}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.13660839200019836, '1': 0.863391637802124}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.9891815781593323, '1': 0.010818448849022388}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.00985985528677702, '1': 0.9901401400566101}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.8289388418197632, '1': 0.1710612028837204}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.13660839200019836, '1': 0.863391637802124}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.4073334336280823, '1': 0.5926666259765625}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.12421300262212753, '1': 0.8757869601249695}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.8840392827987671, '1': 0.11596071728123291}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.8887588381767273, '1': 0.11124120652675629}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.8887588381767273, '1': 0.11124120652675629}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.7431679368019104, '1': 0.25683197379112244}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.0007321813027374446, '1': 0.9992678761482239}  
Predicted Answer: 1, True Answer: 1  
Evaluating: 100%|#####| 277/277 [00:00<00:00, 1792.94it/s]  
Evaluation completed. Accuracy: 0.8448 (234/277)  
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/RTE#
```

Consistent proxy tuning acc: **0.8484**

```
root@e8e99ab71e21: ~  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.0007672925130464137, '1': 0.9992327690124512}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.0002611903182696551, '1': 0.9997387528419495}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.9999538660049438, '1': 4.611274562194012e-05}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.00018814207578543574, '1': 0.9998118281364441}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.9999023675918579, '1': 9.7615658887662e-05}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.00018522574100643396, '1': 0.9998148083686829}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.002800926798954606, '1': 0.9971990585327148}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.0011879054363816977, '1': 0.9988120794296265}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.9999125003814697, '1': 8.750300185056403e-05}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.9999617338180542, '1': 3.822911094175652e-05}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.9998726844787598, '1': 0.00012731103925034404}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.9964063763618469, '1': 0.0035936026833951473}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 8.88808426680043e-05, '1': 0.9999110698699951}  
Predicted Answer: 1, True Answer: 1  
Evaluating: 100%|#####| 277/277 [00:00<00:00, 2094.72it/s]  
Evaluation completed. Accuracy: 0.8484 (235/277)  
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/RTE#
```

13b loratune acc: **0.8664**

```
Predicted Label: 1, True Label: 1  
Evaluating: 97%|#####| 269/277 [03:20<00:06, 1.24it/s]  
Probabilities for choices: {'0': 0.89501953125, '1': 0.105224609375}  
Predicted Label: 0, True Label: 0  
Evaluating: 97%|#####| 270/277 [03:21<00:05, 1.34it/s]  
Probabilities for choices: {'0': 0.00926971435546875, '1': 0.99072265625}  
Predicted Label: 1, True Label: 1  
Evaluating: 98%|#####| 271/277 [03:22<00:04, 1.41it/s]  
Probabilities for choices: {'0': 0.05145263671875, '1': 0.94873046875}  
Predicted Label: 1, True Label: 1  
Evaluating: 98%|#####| 272/277 [03:22<00:03, 1.56it/s]  
Probabilities for choices: {'0': 0.140380859375, '1': 0.85986328125}  
Predicted Label: 1, True Label: 1  
Evaluating: 99%|#####| 273/277 [03:23<00:02, 1.58it/s]  
Probabilities for choices: {'0': 0.6513671875, '1': 0.3486328125}  
Predicted Label: 0, True Label: 0  
Evaluating: 99%|#####| 274/277 [03:24<00:01, 1.50it/s]  
Probabilities for choices: {'0': 0.892578125, '1': 0.107421875}  
Predicted Label: 0, True Label: 0  
Evaluating: 99%|#####| 275/277 [03:24<00:01, 1.45it/s]  
Probabilities for choices: {'0': 0.81298828125, '1': 0.1871337890625}  
Predicted Label: 0, True Label: 0  
Evaluating: 100%|#####| 276/277 [03:25<00:00, 1.50it/s]  
Probabilities for choices: {'0': 0.65478515625, '1': 0.34521484375}  
Predicted Label: 0, True Label: 0  
Evaluating: 100%|#####| 277/277 [03:26<00:00, 1.34it/s]  
Probabilities for choices: {'0': 0.00971221923828125, '1': 0.990234375}  
Predicted Label: 1, True Label: 1  
Evaluation completed. Accuracy: 0.8664 (240/277)  
Logits saved!
```

Gaussian Process tuning acc: **0.8628**

230 / 2490 (9.24%)

```
root@e8e99ab71e21: ~  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.0004878571489825845, '1': 0.9995121955871582}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.00012931562378071249, '1': 0.9998706579208374}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.9998488426208496, '1': 0.0001511819864390418}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 8.34962775115855e-05, '1': 0.9999165534973145}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.9998260140419006, '1': 0.00017400544311385602}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 3.219233258278109e-05, '1': 0.9999678134918213}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.0007672925130464137, '1': 0.9992327690124512}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.0006771942134946585, '1': 0.999322772026062}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.9997997879981995, '1': 0.00020027377468068153}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.9998557567596436, '1': 0.00014425987319555134}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.9998557567596436, '1': 0.00014425987319555134}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.09670578688383102, '1': 0.9032942056655884}  
Predicted Answer: 1, True Answer: 0  
Probabilities for choices: {'0': 3.944258423871361e-05, '1': 0.9999605417251587}  
Predicted Answer: 1, True Answer: 1  
Evaluating: 100%|#####| 277/277 [00:00<00:00, 2246.36it/s]  
Evaluation completed. Accuracy: 0.8628 (239/277)  
(base) root@e8e99ab71e21: ~/Workspace/SF_CPT/RTE#
```

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

Config: input\_threshold=0.149, output\_threshold=0.4

46 / 2490 (1.85%)

```
root@e8e99ab71e21: ~/Works  
root@e8e99ab71e21: ~  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.0007672925130464137, '1': 0.9992327690124512}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.00023782206699252129, '1': 0.999762237071991}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.9999251365661621, '1': 7.484622619813308e-05}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.0001823546044761315, '1': 0.9998176693916321}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.99991774559021, '1': 8.220189192797989e-05}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.00016346470511052758, '1': 0.9998365640640259}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.0014103584690019488, '1': 0.9985895752906799}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.000656372809316963, '1': 0.9993435740470886}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.9999165534973145, '1': 8.34962775115855e-05}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.999937891960144, '1': 6.205048703122884e-05}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.9998623132705688, '1': 0.00013765462790615857}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.9984512329101562, '1': 0.0015487611526623368}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.0010889690020121634, '1': 0.998910427093506}  
Predicted Answer: 1, True Answer: 1  
Evaluating: 100%|#####| 277/277 [00:00<00:00, 2249.99it/s]  
Evaluation completed. Accuracy: 0.8448 (234/277)  
(base) root@e8e99ab71e21: ~/Workspace/SF_CPT/RTE#
```