Benchmark:

Model & Method	# Trainable Parameters	MNLI	SST-2	MRPC	CoLA	QNLI	QQP	RTE	STS-B	Avg.
RoB$_{base}$ (FT)*	125.0M	**87.6**	94.8	90.2	**63.6**	92.8	**91.9**	78.7	91.2	86.4
RoB$_{base}$ (BitFit)*	0.1M	84.7	93.7	**92.7**	62.0	91.8	84.0	81.5	90.8	85.2
RoB$_{base}$ (Adpt^D)*	0.3M	$87.1_{\pm.0}$	$94.2_{\pm.1}$	$88.5_{\pm1.1}$	$60.8_{\pm.4}$	$93.1_{\pm.1}$	$90.2_{\pm.0}$	$71.5_{\pm2.7}$	$89.7_{\pm.3}$	84.4
RoB$_{base}$ (Adpt^D)*	0.9M	$87.3_{\pm.1}$	$94.7_{\pm.3}$	$88.4_{\pm.1}$	$62.6_{\pm.9}$	$93.0_{\pm.2}$	$90.6_{\pm.0}$	$75.9_{\pm2.2}$	$90.3_{\pm.1}$	85.4
RoB$_{base}$ (LoRA)	0.3M	$87.5_{\pm.3}$	$\mathbf{95.1}_{\pm.2}$	$89.7_{\pm.7}$	$63.4_{\pm1.2}$	$\mathbf{93.3}_{\pm.3}$	$90.8_{\pm.1}$	$\mathbf{86.6}_{\pm.7}$	$\mathbf{91.5}_{\pm.2}$	**87.2**
RoB$_{large}$ (FT)*	355.0M	90.2	**96.4**	90.9	68.0	94.7	**92.2**	86.6	92.4	88.9
RoB$_{large}$ (LoRA)	0.8M	$\mathbf{90.6}_{\pm.2}$	$96.2_{\pm.5}$	$\mathbf{90.9}_{\pm1.2}$	$\mathbf{68.2}_{\pm1.9}$	$\mathbf{94.9}_{\pm.3}$	$91.6_{\pm.1}$	$\mathbf{87.4}_{\pm2.5}$	$\mathbf{92.6}_{\pm.2}$	**89.0**
RoB$_{large}$ (Adpt^P)†	3.0M	$90.2_{\pm.3}$	$96.1_{\pm.3}$	$90.2_{\pm.7}$	$\mathbf{68.3}_{\pm1.0}$	$94.8_{\pm.2}$	$\mathbf{91.9}_{\pm.1}$	$83.8_{\pm2.9}$	$92.1_{\pm.7}$	88.4
RoB$_{large}$ (Adpt^P)†	0.8M	$\mathbf{90.5}_{\pm.3}$	$\mathbf{96.6}_{\pm.2}$	$89.7_{\pm1.2}$	$67.8_{\pm2.5}$	$94.8_{\pm.3}$	$91.7_{\pm.2}$	$80.1_{\pm2.9}$	$91.9_{\pm.4}$	87.9
RoB$_{large}$ (Adpt^H)†	6.0M	$89.9_{\pm.5}$	$96.2_{\pm.3}$	$88.7_{\pm2.9}$	$66.5_{\pm4.4}$	$94.7_{\pm.2}$	$92.1_{\pm.1}$	$83.4_{\pm1.1}$	$91.0_{\pm1.7}$	87.8
RoB$_{large}$ (Adpt^H)†	0.8M	$90.3_{\pm.3}$	$96.3_{\pm.5}$	$87.7_{\pm1.7}$	$66.3_{\pm2.0}$	$94.7_{\pm.2}$	$91.5_{\pm.1}$	$72.9_{\pm2.9}$	$91.5_{\pm.5}$	86.4
RoB$_{large}$ (LoRA)†	0.8M	$\mathbf{90.6}_{\pm.2}$	$96.2_{\pm.5}$	$\mathbf{90.2}_{\pm1.0}$	$68.2_{\pm1.9}$	$94.8_{\pm.3}$	$91.6_{\pm.2}$	$85.2_{\pm1.1}$	$\mathbf{92.3}_{\pm.5}$	88.6
DeB$_{XXL}$ (FT)*	1500.0M	91.8	**97.2**	92.0	72.0	**96.0**	92.7	93.9	92.9	91.1
DeB$_{XXL}$ (LoRA)	4.7M	$\mathbf{91.9}_{\pm.2}$	$96.9_{\pm.2}$	$\mathbf{92.6}_{\pm.6}$	$\mathbf{72.4}_{\pm1.1}$	$\mathbf{96.0}_{\pm.1}$	$\mathbf{92.9}_{\pm.1}$	$\mathbf{94.9}_{\pm.4}$	$\mathbf{93.0}_{\pm.2}$	**91.3**

Table 2: RoBERTa$_{base}$, RoBERTa$_{large}$, and DeBERTa$_{XXL}$ with different adaptation methods on the GLUE benchmark. We report the overall (matched and mismatched) accuracy for MNLI, Matthew's correlation for CoLA, Pearson correlation for STS-B, and accuracy for other tasks. Higher is better for all metrics. * indicates numbers published in prior works. † indicates runs configured in a setup similar to Houlsby et al. (2019) for a fair comparison.

7b pretrain acc: **0.4107**


```
Evaluating: 100%|          | 2994/3000 [14:49<00:01,  3.54it/s]
Probabilities for choices: {'1': 0.521484375, '0': 0.478515625}
Predicted Label: 1, True Label: 1
Evaluating: 100%|          | 2995/3000 [14:49<00:01,  3.61it/s]
Probabilities for choices: {'1': 0.437744140625, '0': 0.56201171875}
Predicted Label: 0, True Label: 1
Evaluating: 100%|          | 2996/3000 [14:49<00:01,  3.66it/s]
Probabilities for choices: {'1': 0.494140625, '0': 0.505859375}
Predicted Label: 0, True Label: 0
Evaluating: 100%|          | 2997/3000 [14:49<00:00,  3.64it/s]
Probabilities for choices: {'1': 0.541015625, '0': 0.458984375}
Predicted Label: 1, True Label: 0
Evaluating: 100%|          | 2998/3000 [14:50<00:00,  3.29it/s]
Probabilities for choices: {'1': 0.57177734375, '0': 0.42822265625}
Predicted Label: 1, True Label: 0
Evaluating: 100%|          | 2999/3000 [14:50<00:00,  3.41it/s]
Probabilities for choices: {'1': 0.46875, '0': 0.53125}
Predicted Label: 0, True Label: 1
Evaluating: 100%|          | 3000/3000 [14:50<00:00,  3.37it/s]
Probabilities for choices: {'1': 0.478515625, '0': 0.521484375}
Predicted Label: 0, True Label: 0
Evaluation completed. Accuracy: 0.4107 (1232/3000)

Logits saved!
```

13b pretrain acc: **0.3930**


```
Probabilities for choices: {'1': 0.5546875, '0': 0.445556640625}
Predicted Label: 1, True Label: 1
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2992/3000 [22:45<00:03,  2.06it/s]
Probabilities for choices: {'1': 0.529296875, '0': 0.470703125}
Predicted Label: 1, True Label: 1
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2993/3000 [22:46<00:03,  2.12it/s]
Probabilities for choices: {'1': 0.5888671875, '0': 0.4111328125}
Predicted Label: 1, True Label: 1
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2994/3000 [22:46<00:02,  2.18it/s]
Probabilities for choices: {'1': 0.63525390625, '0': 0.36474609375}
Predicted Label: 1, True Label: 1
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2995/3000 [22:46<00:02,  2.22it/s]
Probabilities for choices: {'1': 0.55810546875, '0': 0.441650390625}
Predicted Label: 1, True Label: 0
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2996/3000 [22:47<00:01,  2.24it/s]
Probabilities for choices: {'1': 0.56005859375, '0': 0.439697265625}
Predicted Label: 1, True Label: 0
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2997/3000 [22:47<00:01,  2.27it/s]
Probabilities for choices: {'1': 0.64404296875, '0': 0.355712890625}
Predicted Label: 1, True Label: 0
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2998/3000 [22:48<00:00,  2.15it/s]
Probabilities for choices: {'1': 0.57763671875, '0': 0.422607421875}
Predicted Label: 1, True Label: 1
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2999/3000 [22:48<00:00,  2.19it/s]
Probabilities for choices: {'1': 0.56005859375, '0': 0.439697265625}
Predicted Label: 1, True Label: 0
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 3000/3000 [22:49<00:00,  2.19it/s]
Evaluation completed. Accuracy: 0.3930 (1179/3000)
Logits saved!
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/QQP#
```

7b lora-tune acc: **0.6337**


```
Probabilities for choices: {'1': 0.82666015625, '0': 0.17333984375}
Predicted Label: 1, True Label: 1
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2992/3000 [11:41<00:02,  3.98it/s]
Probabilities for choices: {'1': 0.865234375, '0': 0.134765625}
Predicted Label: 1, True Label: 1
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2993/3000 [11:41<00:01,  4.10it/s]
Probabilities for choices: {'1': 0.85009765625, '0': 0.1500244140625}
Predicted Label: 1, True Label: 1
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2994/3000 [11:41<00:01,  4.20it/s]
Probabilities for choices: {'1': 0.80322265625, '0': 0.19677734375}
Predicted Label: 1, True Label: 1
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2995/3000 [11:41<00:01,  4.27it/s]
Probabilities for choices: {'1': 0.352294921875, '0': 0.64794921875}
Predicted Label: 0, True Label: 0
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2996/3000 [11:41<00:00,  4.33it/s]
Probabilities for choices: {'1': 0.52734375, '0': 0.47265625}
Predicted Label: 1, True Label: 0
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2997/3000 [11:42<00:00,  4.38it/s]
Probabilities for choices: {'1': 0.4111328125, '0': 0.5888671875}
Predicted Label: 0, True Label: 0
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2998/3000 [11:42<00:00,  4.06it/s]
Probabilities for choices: {'1': 0.87744140625, '0': 0.12249755859375}
Predicted Label: 1, True Label: 1
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 2999/3000 [11:42<00:00,  4.15it/s]
Probabilities for choices: {'1': 0.56982421875, '0': 0.43017578125}
Predicted Label: 1, True Label: 0
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 3000/3000 [11:42<00:00,  4.27it/s]
Evaluation completed. Accuracy: 0.6337 (1901/3000)
Logits saved!
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/QQP#
```

13b lora-tune acc: **0.8653**

```
Predicted Label: 1, True Label: 1
Evaluating: 100%|          | 2992/3000 [26:16<00:04,  1.82it/s]
Probabilities for choices: {'1': 0.8681640625, '0': 0.132080078125}
Predicted Label: 1, True Label: 1
Evaluating: 100%|          | 2993/3000 [26:16<00:03,  1.88it/s]
Probabilities for choices: {'1': 0.8623046875, '0': 0.1375732421875}
Predicted Label: 1, True Label: 1
Evaluating: 100%|          | 2994/3000 [26:17<00:03,  1.93it/s]
Probabilities for choices: {'1': 0.8232421875, '0': 0.1766357421875}
Predicted Label: 1, True Label: 1
Evaluating: 100%|          | 2995/3000 [26:17<00:02,  1.96it/s]
Probabilities for choices: {'1': 0.8310546875, '0': 0.1688232421875}
Predicted Label: 1, True Label: 1
Evaluating: 100%|          | 2996/3000 [26:18<00:02,  1.98it/s]
Probabilities for choices: {'1': 0.070068359375, '0': 0.93017578125}
Predicted Label: 0, True Label: 0
Evaluating: 100%|          | 2997/3000 [26:18<00:01,  2.00it/s]
Probabilities for choices: {'1': 0.03790283203125, '0': 0.9619140625}
Predicted Label: 0, True Label: 0
Evaluating: 100%|          | 2998/3000 [26:19<00:01,  1.87it/s]
Probabilities for choices: {'1': 0.1044921875, '0': 0.8955078125}
Predicted Label: 0, True Label: 0
Evaluating: 100%|          | 2999/3000 [26:19<00:00,  1.92it/s]
Probabilities for choices: {'1': 0.763671875, '0': 0.236572265625}
Predicted Label: 1, True Label: 1
Evaluating: 100%|          | 3000/3000 [26:20<00:00,  1.90it/s]
Probabilities for choices: {'1': 0.1365966796875, '0': 0.86328125}
Predicted Label: 0, True Label: 0
Evaluation completed. Accuracy: 0.8653 (2596/3000)

Logits saved!
```

Proxy tuning acc: **0.6100**

```
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.8688268065452576, '0': 0.13117322325706482}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.048857778310775576, '0': 0.9511421918869019}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.6261242032051086, '0': 0.37387582659721375}
Predicted Answer: 1, True Answer: 0
Probabilities for choices: {'1': 0.84594243764877322, '0': 0.1540575623512268}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.8558511734008789, '0': 0.14414885640144348}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.879974365234375, '0': 0.120025634765625}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.8661325573921204, '0': 0.13386750221252441}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.859663724899292, '0': 0.14033624529838562}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.3812195658683777, '0': 0.6187804341316223}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.5389832258224487, '0': 0.46101677417755127}
Predicted Answer: 1, True Answer: 0
Probabilities for choices: {'1': 0.4493926465511322, '0': 0.5506073832511902}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.8991213440895081, '0': 0.10087861865758896}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.6113819479942322, '0': 0.38861799240112305}
Predicted Answer: 1, True Answer: 0
Evaluating: 100%|♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦♦| 3000/3000 [00:01<00:00, 2159.80it/s]
Evaluation completed. Accuracy: 0.6100 (1830/3000)
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/QQP#
```

Consistent proxy tuning acc: **0.8603**

GP tuning acc: **0.8643**

900 / 10000 (9%)

GP with filter tuning acc: **0.8623**

Config: input_threshold=0.16, output_threshold=0.45

241 / 10000 (2.41%)

```
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.7057850360870361, '0': 0.29421496391296387}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.005220125894993544, '0': 0.9947799444198608}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.6406359076499939, '0': 0.3593641519546509}
Predicted Answer: 1, True Answer: 0
Probabilities for choices: {'1': 0.9149009585380554, '0': 0.08509904891252518}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.9683812260627747, '0': 0.031618766486644745}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.9489172697067261, '0': 0.05108273774385452}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.917302668094635, '0': 0.08269734680652618}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.894789457321167, '0': 0.10521053522825241}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.01590639166533947, '0': 0.9840936064720154}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.005554925184696913, '0': 0.9944451451301575}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.02262919954955578, '0': 0.9773707389831543}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'1': 0.9161096215248108, '0': 0.0838903859257698}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'1': 0.033589500933885574, '0': 0.9664104580879211}
Predicted Answer: 0, True Answer: 0
Evaluating: 100%|◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆◆| 3000/3000 [00:01<00:00, 2274.83it/s]
Evaluation completed. Accuracy: 0.8623 (2587/3000)
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/QQP#
```