

Benchmark from LoRA

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 ^P)*	0.3M	87.1 \pm 0	94.2 \pm 1	88.5 \pm 1.1	60.8 \pm 4	93.1 \pm 1	90.2 \pm 0	71.5 \pm 2.7	89.7 \pm 3	84.4
RoB _{base} (Adpt ^P)*	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 \pm 2.2	90.3 \pm 1	85.4
RoB _{base} (LoRA)	0.3M	87.5 \pm 3	95.1\pm2	89.7 \pm 7	63.4 \pm 1.2	93.3\pm3	90.8 \pm 1	86.6\pm7	91.5\pm2	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	90.6\pm2	96.2 \pm 5	90.9\pm1.2	68.2\pm1.9	94.9\pm3	91.6 \pm 1	87.4\pm2.5	92.6\pm2	89.0
RoB _{large} (Adpt ^P)†	3.0M	90.2 \pm 3	96.1 \pm 3	90.2 \pm 7	68.3\pm1.0	94.8\pm2	91.9\pm1	83.8 \pm 2.9	92.1 \pm 7	88.4
RoB _{large} (Adpt ^P)†	0.8M	90.5\pm3	96.6\pm2	89.7 \pm 1.2	67.8 \pm 2.5	94.8\pm3	91.7 \pm 2	80.1 \pm 2.9	91.9 \pm 4	87.9
RoB _{large} (Adpt ^H)†	6.0M	89.9 \pm 5	96.2 \pm 3	88.7 \pm 2.9	66.5 \pm 4.4	94.7 \pm 2	92.1 \pm 1	83.4 \pm 1.1	91.0 \pm 1.7	87.8
RoB _{large} (Adpt ^H)†	0.8M	90.3 \pm 3	96.3 \pm 5	87.7 \pm 1.7	66.3 \pm 2.0	94.7 \pm 2	91.5 \pm 1	72.9 \pm 2.9	91.5 \pm 5	86.4
RoB _{large} (LoRA)†	0.8M	90.6\pm2	96.2 \pm 5	90.2\pm1.0	68.2 \pm 1.9	94.8\pm3	91.6 \pm 2	85.2\pm1.1	92.3\pm5	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	91.9\pm2	96.9 \pm 2	92.6\pm6	72.4\pm1.1	96.0\pm1	92.9\pm1	94.9\pm4	93.0\pm2	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.

text1 string · lengths	text2 string · lengths	label int64	idx int64	label_text string · classes
 1 270	 3 2.46k	 0 1	 0 105k	 2 values
When did the third Digimon series begin?	Unlike the two seasons before it an...	1	0	not entailment
Which missile batteries often have individual...	When MANPADS is operated by...	1	1	not entailment
What two things does Popper argue Tarski's...	He bases this interpretation on th...	0	2	entailment
What is the name of the village 9 miles north of...	On 31 December 1853, the Ottoman forces a...	0	3	entailment
What famous palace is located in London?	London contains four World Heritage Sites...	1	4	not entailment
When is the term 'German dialects' used in regard...	When talking about the German language,...	0	5	entailment

7b pretrain accuracy: **0.4997**

```
root@e8e99ab71e21: ~  
Probabilities for choices: {'0': 0.466796875, '1': 0.533203125}  
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 2992/3000 [29:03<00:04, 1.89it/s]  
Probabilities for choices: {'0': 0.43408203125, '1': 0.56591796875}  
Predicted Label: 1, True Label: 0  
Evaluating: 100%|#####| 2993/3000 [29:03<00:03, 1.75it/s]  
Probabilities for choices: {'0': 0.69921875, '1': 0.30078125}  
Predicted Label: 0, True Label: 0  
Evaluating: 100%|#####| 2994/3000 [29:04<00:03, 1.77it/s]  
Probabilities for choices: {'0': 0.38671875, '1': 0.61328125}  
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 2995/3000 [29:04<00:02, 1.78it/s]  
Probabilities for choices: {'0': 0.375732421875, '1': 0.62451171875}  
Predicted Label: 1, True Label: 0  
Evaluating: 100%|#####| 2996/3000 [29:05<00:02, 1.79it/s]  
Probabilities for choices: {'0': 0.42431640625, '1': 0.57568359375}  
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 2997/3000 [29:05<00:01, 1.79it/s]  
Probabilities for choices: {'0': 0.533203125, '1': 0.466796875}  
Predicted Label: 0, True Label: 1  
Evaluating: 100%|#####| 2998/3000 [29:06<00:01, 1.80it/s]  
Probabilities for choices: {'0': 0.6669921875, '1': 0.332763671875}  
Predicted Label: 0, True Label: 1  
Evaluating: 100%|#####| 2999/3000 [29:07<00:00, 1.70it/s]  
Probabilities for choices: {'0': 0.5947265625, '1': 0.405517578125}  
Predicted Label: 0, True Label: 1  
Evaluating: 100%|#####| 3000/3000 [29:07<00:00, 1.72it/s]  
Evaluation completed. Accuracy: 0.4997 (1499/3000)  
Logits saved!  
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/QNLI#
```

13b pretrain accuracy: **0.5210**

```
root@e8e99ab71e21: ~/Works  
Probabilities for choices: {'0': 0.716796875, '1': 0.282958984375}  
Predicted Label: 0, True Label: 1  
Evaluating: 100%|#####| 2992/3000 [14:38<00:02, 3.61it/s]  
Probabilities for choices: {'0': 0.71875, '1': 0.281494140625}  
Predicted Label: 0, True Label: 0  
Evaluating: 100%|#####| 2993/3000 [14:38<00:02, 3.39it/s]  
Probabilities for choices: {'0': 0.74169921875, '1': 0.25830078125}  
Predicted Label: 0, True Label: 0  
Evaluating: 100%|#####| 2994/3000 [14:39<00:01, 3.43it/s]  
Probabilities for choices: {'0': 0.68603515625, '1': 0.31396484375}  
Predicted Label: 0, True Label: 1  
Evaluating: 100%|#####| 2995/3000 [14:39<00:01, 3.49it/s]  
Probabilities for choices: {'0': 0.64404296875, '1': 0.355712890625}  
Predicted Label: 0, True Label: 0  
Evaluating: 100%|#####| 2996/3000 [14:39<00:01, 3.52it/s]  
Probabilities for choices: {'0': 0.6689453125, '1': 0.3310546875}  
Predicted Label: 0, True Label: 1  
Evaluating: 100%|#####| 2997/3000 [14:39<00:00, 3.54it/s]  
Probabilities for choices: {'0': 0.66552734375, '1': 0.33447265625}  
Predicted Label: 0, True Label: 1  
Evaluating: 100%|#####| 2998/3000 [14:40<00:00, 3.54it/s]  
Probabilities for choices: {'0': 0.71044921875, '1': 0.289306640625}  
Predicted Label: 0, True Label: 1  
Evaluating: 100%|#####| 2999/3000 [14:40<00:00, 3.35it/s]  
Probabilities for choices: {'0': 0.546875, '1': 0.453369140625}  
Predicted Label: 0, True Label: 1  
Evaluating: 100%|#####| 3000/3000 [14:40<00:00, 3.41it/s]  
Evaluation completed. Accuracy: 0.5210 (1563/3000)  
Logits saved!  
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/QNLI# d
```

7b lora-tune accuracy: **0.8360**

```
root@e8e99ab71e21: ~  
Probabilities for choices: {'0': 0.1710205078125, '1': 0.8291015625}  
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 2992/3000 [15:25<00:02, 3.57it/s]  
Probabilities for choices: {'0': 0.20947265625, '1': 0.79052734375}  
Predicted Label: 1, True Label: 0  
Evaluating: 100%|#####| 2993/3000 [15:25<00:02, 3.35it/s]  
Probabilities for choices: {'0': 0.64404296875, '1': 0.355712890625}  
Predicted Label: 0, True Label: 0  
Evaluating: 100%|#####| 2994/3000 [15:25<00:01, 3.38it/s]  
Probabilities for choices: {'0': 0.127685546875, '1': 0.87255859375}  
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 2995/3000 [15:26<00:01, 3.43it/s]  
Probabilities for choices: {'0': 0.77197265625, '1': 0.2281494140625}  
Predicted Label: 0, True Label: 0  
Evaluating: 100%|#####| 2996/3000 [15:26<00:01, 3.43it/s]  
Probabilities for choices: {'0': 0.1461181640625, '1': 0.85400390625}  
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 2997/3000 [15:26<00:00, 3.41it/s]  
Probabilities for choices: {'0': 0.338134765625, '1': 0.662109375}  
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 2998/3000 [15:26<00:00, 3.39it/s]  
Probabilities for choices: {'0': 0.19189453125, '1': 0.80810546875}  
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 2999/3000 [15:27<00:00, 3.21it/s]  
Probabilities for choices: {'0': 0.25390625, '1': 0.74609375}  
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 3000/3000 [15:27<00:00, 3.23it/s]  
Evaluation completed. Accuracy: 0.8360 (2508/3000)  
Logits saved!  
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/QNLI#
```

13b lora-tune accuracy: **0.8780**

```
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 2992/3000 [32:45<00:04, 1.68it/s]  
Probabilities for choices: {'0': 0.1688232421875, '1': 0.8310546875}  
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 2993/3000 [32:46<00:04, 1.57it/s]  
Probabilities for choices: {'0': 0.11126708984375, '1': 0.888671875}  
Predicted Label: 1, True Label: 0  
Evaluating: 100%|#####| 2994/3000 [32:47<00:03, 1.58it/s]  
Probabilities for choices: {'0': 0.7705078125, '1': 0.2294921875}  
Predicted Label: 0, True Label: 0  
Evaluating: 100%|#####| 2995/3000 [32:47<00:03, 1.60it/s]  
Probabilities for choices: {'0': 0.06756591796875, '1': 0.9326171875}  
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 2996/3000 [32:48<00:02, 1.61it/s]  
Probabilities for choices: {'0': 0.83740234375, '1': 0.162353515625}  
Predicted Label: 0, True Label: 0  
Evaluating: 100%|#####| 2997/3000 [32:48<00:01, 1.61it/s]  
Probabilities for choices: {'0': 0.07696533203125, '1': 0.9228515625}  
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 2998/3000 [32:49<00:01, 1.62it/s]  
Probabilities for choices: {'0': 0.22265625, '1': 0.77734375}  
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 2999/3000 [32:50<00:00, 1.53it/s]  
Probabilities for choices: {'0': 0.466796875, '1': 0.533203125}  
Predicted Label: 1, True Label: 1  
Evaluating: 100%|#####| 3000/3000 [32:50<00:00, 1.52it/s]  
Probabilities for choices: {'0': 0.1812744140625, '1': 0.81884765625}  
Predicted Label: 1, True Label: 1  
Evaluation completed. Accuracy: 0.8780 (2634/3000)  
Logits saved!
```

Proxy tuning accuracy: **0.6937**

```
root@e8e99ab71e21: ~/Workspace
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.1233656257390976, '1': 0.8766343593597412}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.6757646203041077, '1': 0.3242354094982147}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'0': 0.3023911118507385, '1': 0.6976089477539062}
Predicted Answer: 1, True Answer: 0
Probabilities for choices: {'0': 0.14414885640144348, '1': 0.8558511734008789}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.08151975274085999, '1': 0.9184802770614624}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.09203287214040756, '1': 0.907967209815979}
Predicted Answer: 1, True Answer: 0
Probabilities for choices: {'0': 0.6076632142066956, '1': 0.3923368752002716}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'0': 0.05146276578307152, '1': 0.9485372304916382}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.5851011276245117, '1': 0.4148988425731659}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'0': 0.07055959850549698, '1': 0.929440438747406}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.24798743426799774, '1': 0.7520126104354858}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.167763352394104, '1': 0.8322367072105408}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.2845759987831116, '1': 0.7154240012168884}
Predicted Answer: 1, True Answer: 1
Evaluating: 100% | 3000/3000 [00:01<00:00, 2284.63it/s]
Evaluation completed. Accuracy: 0.6937 (2081/3000)
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/QNLI#
```

Consistent proxy tuning accuracy: **0.9140**

```
root@e8e99ab71e21: ~
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.002511463360860944, '1': 0.9974884986877441}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.9989995360374451, '1': 0.0010005044750869274}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'0': 0.9945307374000549, '1': 0.005469274241477251}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'0': 0.0034834241960197687, '1': 0.9965165853500366}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.0009850081987679005, '1': 0.9990149736404419}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.005554925184696913, '1': 0.9944451451301575}
Predicted Answer: 1, True Answer: 0
Probabilities for choices: {'0': 0.9994296431541443, '1': 0.0005703156348317862}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'0': 0.0007321813027374446, '1': 0.9992678761482239}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.9981324076652527, '1': 0.0018675660248845816}
Predicted Answer: 0, True Answer: 0
Probabilities for choices: {'0': 0.003429603762924671, '1': 0.996570348739624}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.002714985515922308, '1': 0.997285008430481}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.013636834919452667, '1': 0.9863631129264832}
Predicted Answer: 1, True Answer: 1
Probabilities for choices: {'0': 0.005911069456487894, '1': 0.9940890073776245}
Predicted Answer: 1, True Answer: 1
Evaluating: 100% | 3000/3000 [00:01<00:00, 1936.42it/s]
Evaluation completed. Accuracy: 0.9140 (2742/3000)
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/QNLI#
```

Gaussian Process tuning accuracy: **0.9193**

800 / 9000 (8.89%)

```
root@e8e99ab71e21: ~  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.0018102109897881746, '1': 0.9981897473335266}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.9990890026092529, '1': 0.0009110512328334153}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.9881309270858765, '1': 0.011869121342897415}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.0020190139766782522, '1': 0.9979810118675232}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.0009850081987679005, '1': 0.9990149736404419}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.0023231625091284513, '1': 0.9976769089698792}  
Predicted Answer: 1, True Answer: 0  
Probabilities for choices: {'0': 0.999220609664917, '1': 0.0007793660624884069}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.00036829791497439146, '1': 0.9996316432952881}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.9980121850967407, '1': 0.001987774157896638}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.0017007224960252643, '1': 0.9982993006706238}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.01590639166533947, '1': 0.9840936064720154}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.2628418505191803, '1': 0.7371581792831421}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.0061929901130497456, '1': 0.9938070178031921}  
Predicted Answer: 1, True Answer: 1  
Evaluating: 100%|#####| 3000/3000 [00:01<00:00, 1916.37it/s]  
Evaluation completed. Accuracy: 0.9193 (2758/3000)  
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/QNLI#
```

Gaussian Process with filter accuracy: **0.9180**

180 / 9000 (2%)

```
root@e8e99ab71e21: ~  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.0015487611526623368, '1': 0.9984512329101562}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.999045193195343, '1': 0.0009547316585667431}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.9891815781593323, '1': 0.010818448849022388}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.0019570167642086744, '1': 0.9980429410934448}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.0009253848693333566, '1': 0.9990745782852173}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.001782196224667132, '1': 0.998217761516571}  
Predicted Answer: 1, True Answer: 0  
Probabilities for choices: {'0': 0.9990302324295044, '1': 0.0009697519708424807}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.0003920417802873999, '1': 0.9996079802513123}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.9971103668212891, '1': 0.002889581024646759}  
Predicted Answer: 0, True Answer: 0  
Probabilities for choices: {'0': 0.0012065897462889552, '1': 0.998793363571167}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.007937993854284286, '1': 0.992061972618103}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.19682618975639343, '1': 0.8031737804412842}  
Predicted Answer: 1, True Answer: 1  
Probabilities for choices: {'0': 0.006388320587575436, '1': 0.9936116337776184}  
Predicted Answer: 1, True Answer: 1  
Evaluating: 100%|#####| 3000/3000 [00:01<00:00, 1931.02it/s]  
Evaluation completed. Accuracy: 0.9180 (2754/3000)  
(base) root@e8e99ab71e21:~/Workspace/SF_CPT/QNLI#
```