For finetuning, I used 40% of the data to train using AdamW optimizer.
I trained 4 models which had the hyperparameters:

1) 1e-5 Learning Rate, 16 Batch Size over 3 Epochs. (Base Case)
2) 1e-5 Learning Rate, 16 Batch Size over 1 Epoch. (Epoch Change)
3) 1e-5 Learning Rate, 32 Batch Size over 3 Epochs. (Batch Size Change)
4) 3e-5 Learning Rate, 16 Batch Size over 3 Epochs. (Learning Rate Change)

The rouge scores were:
Base Case:
{'rouge-1': {'f': 0.0841702417285644, 'p': 0.051084033100745045, 'r': 0.3079353399021223},
'rouge-2': {'f': 0.020264212268463918, 'p': 0.012273728389982757, 'r': 0.0889187865602414},
'rouge-l': {'f': 0.07986952382854193, 'p': 0.04838538159016817, 'r': 0.2949689895110542}}

Epoch Change Case:
{'rouge-1': {'f': 0.0841702417285644, 'p': 0.051084033100745045, 'r': 0.3079353399021223},
'rouge-2': {'f': 0.020264212268463918, 'p': 0.012273728389982757, 'r': 0.0889187865602414},
'rouge-l': {'f': 0.07986952382854193, 'p': 0.04838538159016817, 'r': 0.2949689895110542}}

Batch Size Change Case:
{'rouge-1': {'f': 0.08405307834185982, 'p': 0.051347044412385134, 'r': 0.3027681743296919},
'rouge-2': {'f': 0.02033616108539318, 'p': 0.012603284370296461, 'r': 0.08554887596751025},
'rouge-l': {'f': 0.07967384391155471, 'p': 0.04859775332820371, 'r': 0.2897710090147396}}

Learning Rate Change Case:
{'rouge-1': {'f': 0.0841702417285644, 'p': 0.051084033100745045, 'r': 0.3079353399021223},
'rouge-2': {'f': 0.020264212268463918, 'p': 0.012273728389982757, 'r': 0.0889187865602414},
'rouge-l': {'f': 0.07986952382854193, 'p': 0.04838538159016817, 'r': 0.2949689895110542}}