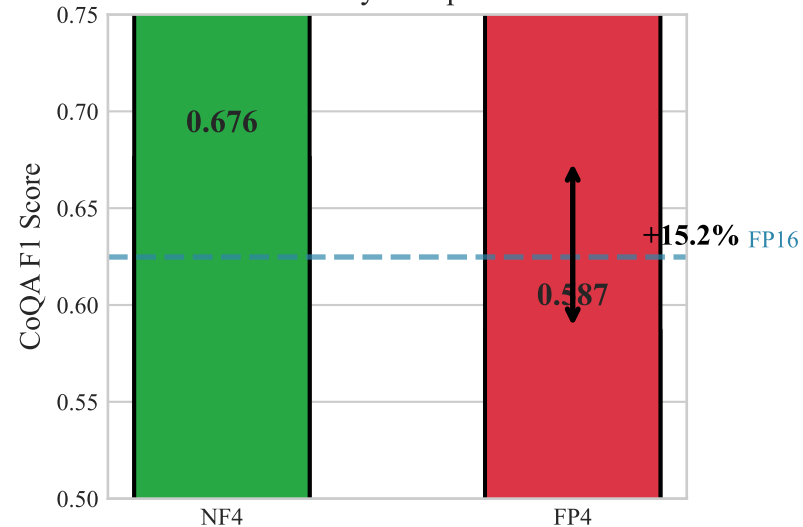


NF4 vs FP4: Better Accuracy at Same Memory Cost

Accuracy Comparison



Memory Usage

