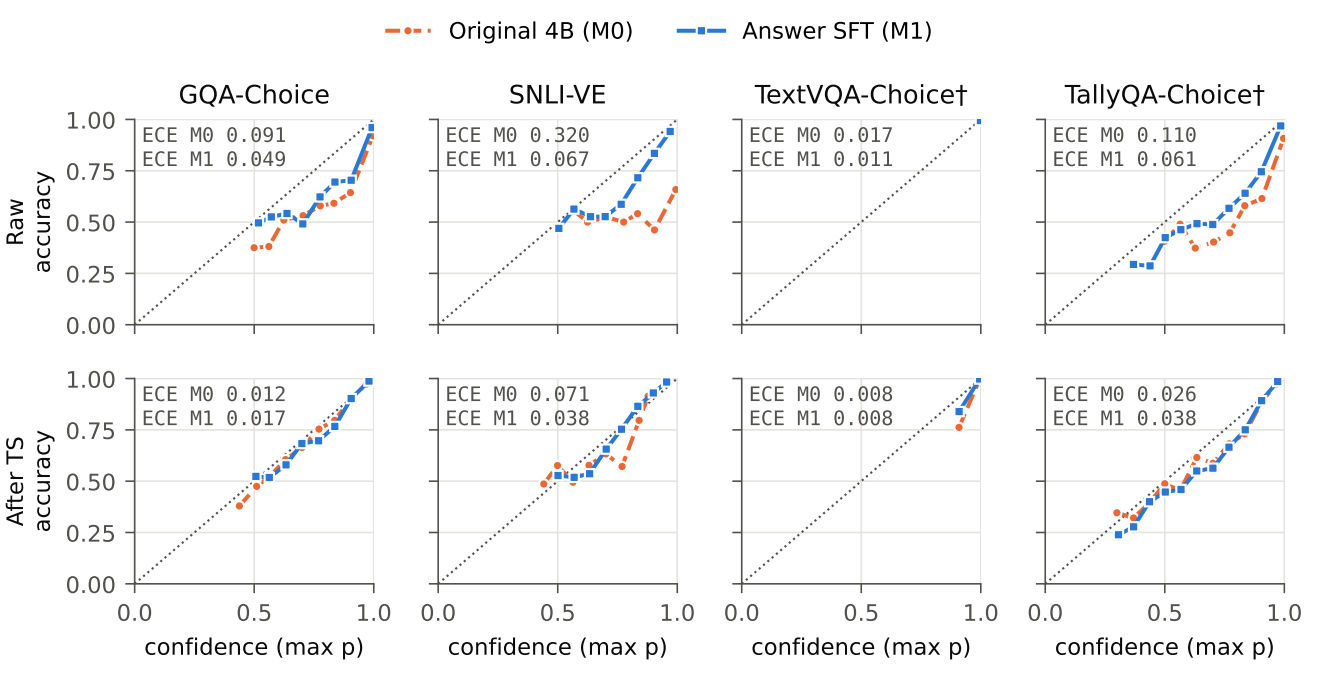

Original 4B (M0)
Answer SFT (M1)
GQA-Choice
SNLI-VE
TextVQA-Choice†
TallyQA-Choice†
Raw accuracy
After TS accuracy
ECE M0 0.091
ECE M1 0.049
ECE M0 0.320
ECE M1 0.067
ECE M0 0.017
ECE M1 0.011
ECE M0 0.110
ECE M1 0.061
ECE M0 0.012
ECE M1 0.017
ECE M0 0.071
ECE M1 0.038
ECE M0 0.008
ECE M1 0.008
ECE M0 0.026
ECE M1 0.038
1.00
0.75
0.50
0.25
0.00
0.0
0.5
1.0
confidence (max p)