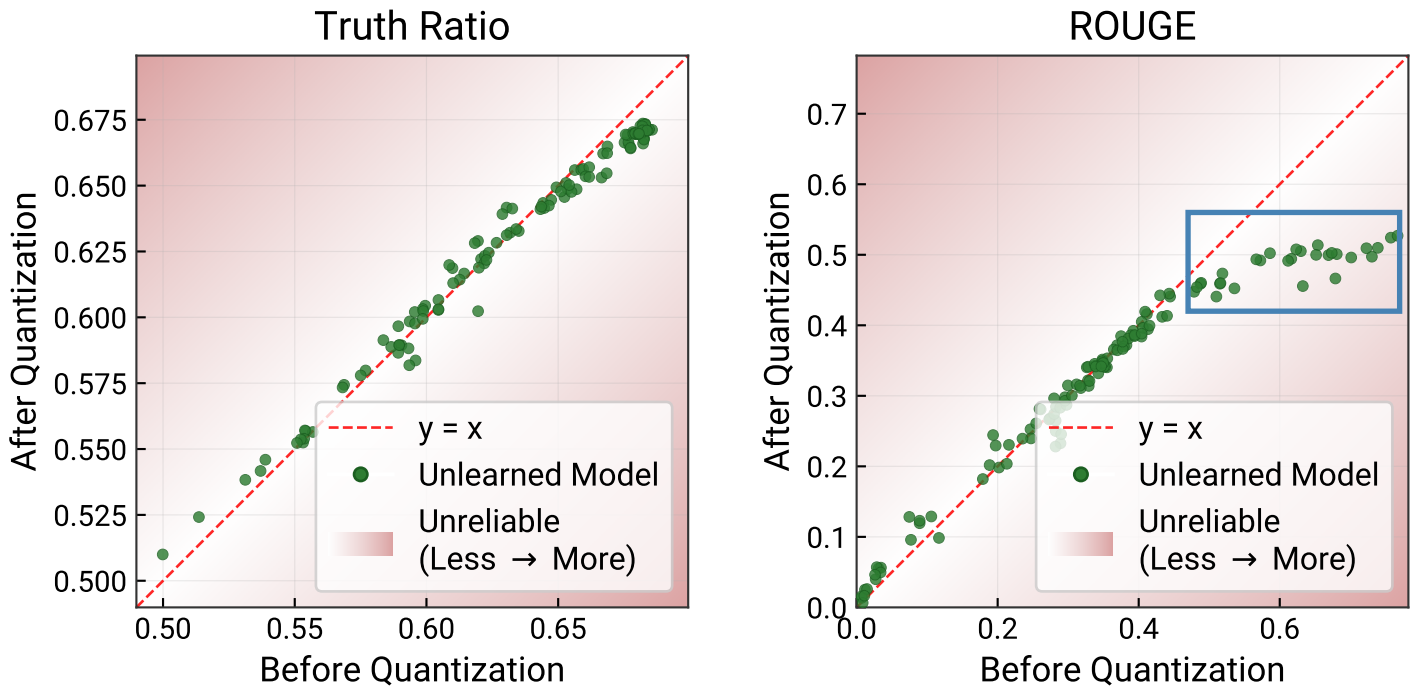
Truth Ratio
After Quantization
Before Quantization
0.675
0.650
0.625
0.600
0.575
0.550
0.525
0.500
0.50
0.55
0.60
0.65
y = x
Unlearned Model
Unreliable
(Less → More)
ROUGE
After Quantization
Before Quantization
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.0
0.2
0.4
0.6
y = x
Unlearned Model
Unreliable
(Less → More)