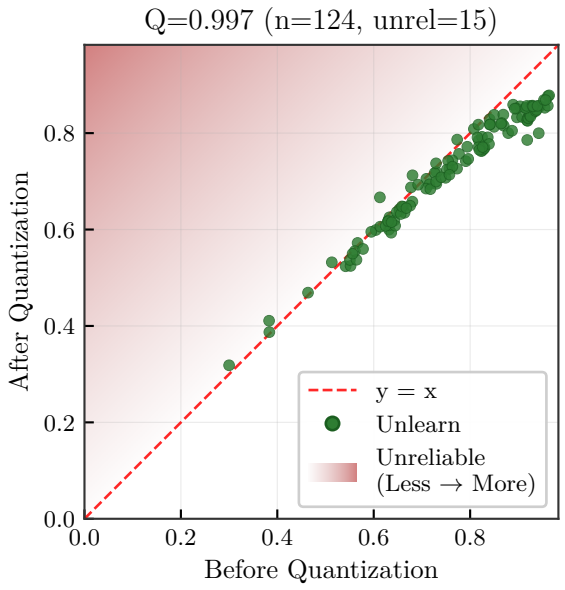
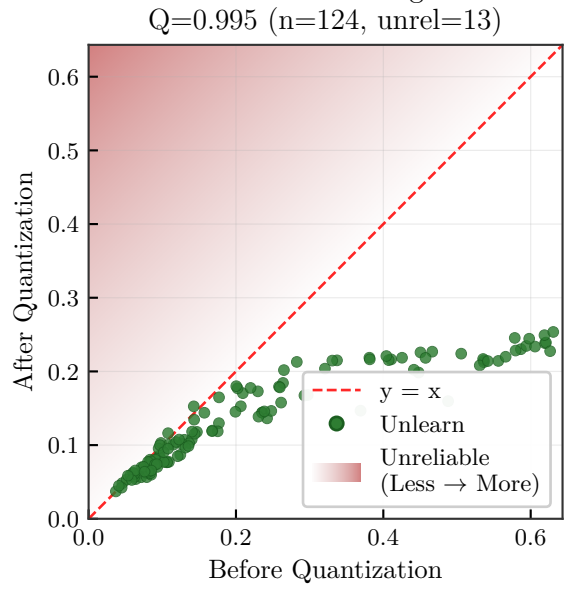
Quantization Robustness (13 Metrics + 4 Normalized MIA + Logit Lens)
(150 Unlearned Models; Utility Filtered)

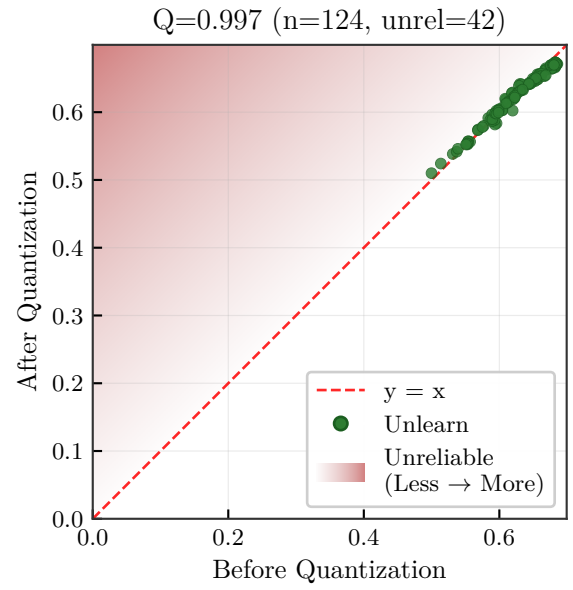
Exact Memorization
Q=0.997 (n=124, unrel=15)



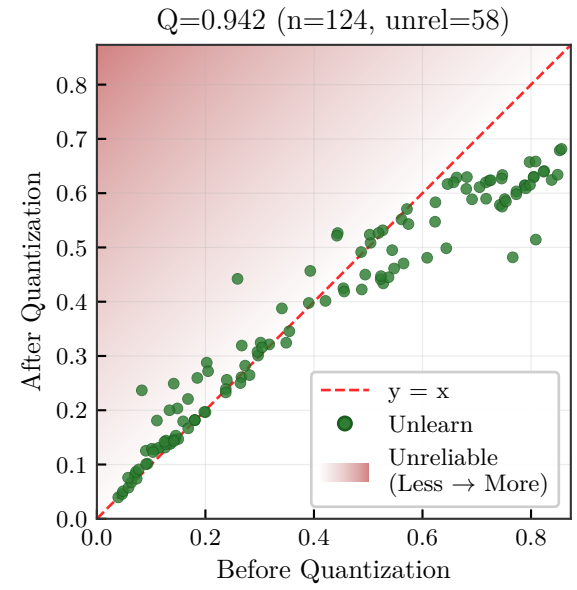

Extraction Strength
Q=0.995 (n=124, unrel=13)



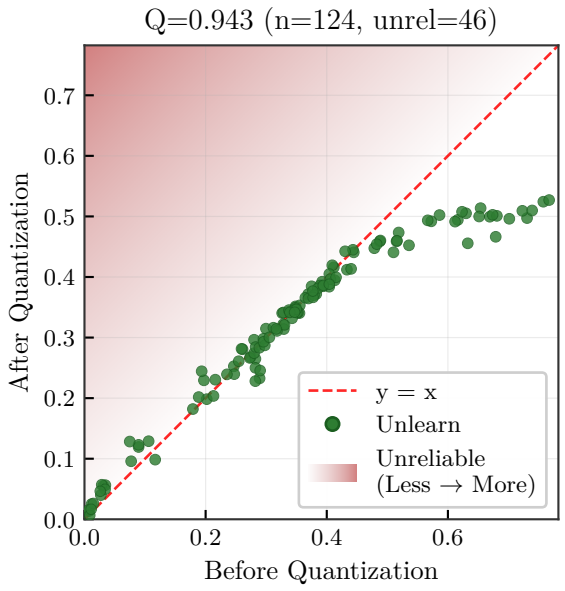

Truth Ratio
Q=0.997 (n=124, unrel=42)



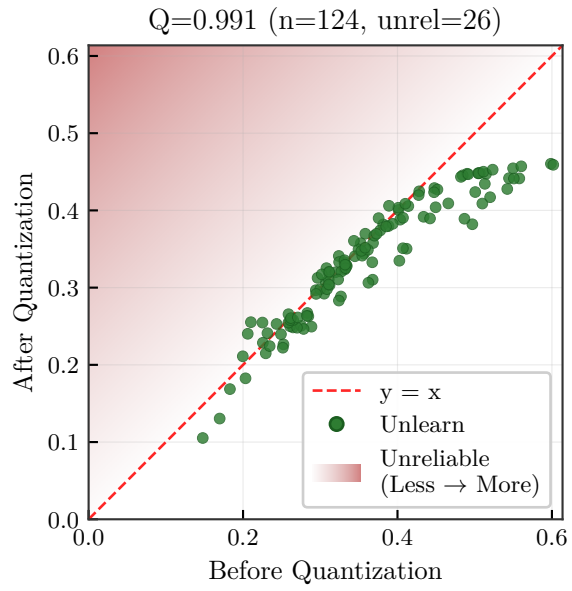

Prob.
Q=0.942 (n=124, unrel=58)



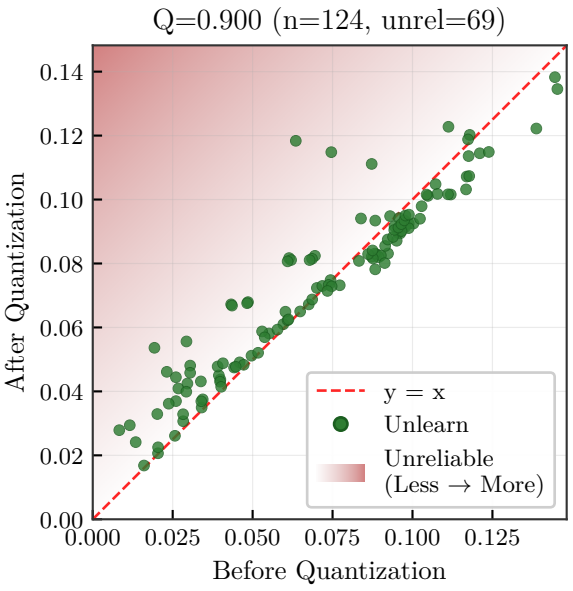

ROUGE
Q=0.943 (n=124, unrel=46)



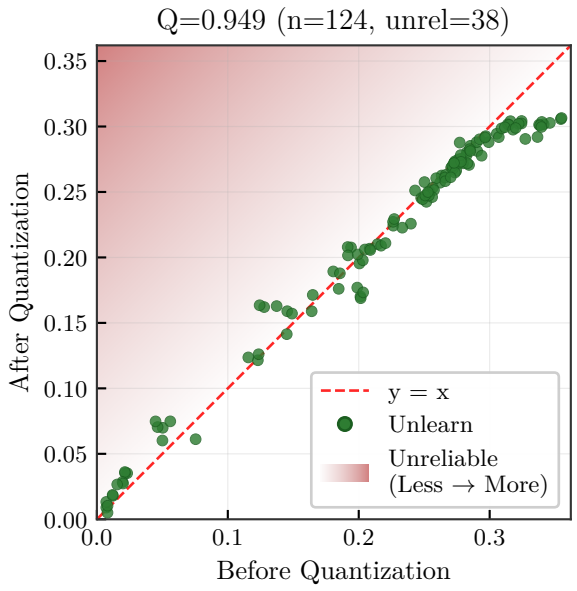

Jailbreak ROUGE
Q=0.991 (n=124, unrel=26)



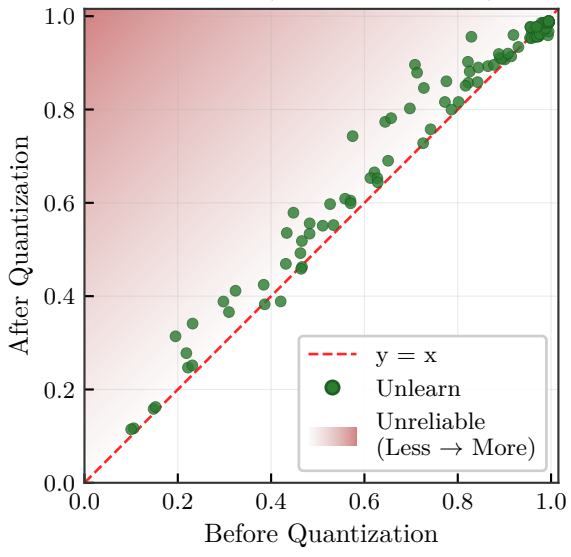

Para. Prob.
Q=0.900 (n=124, unrel=69)



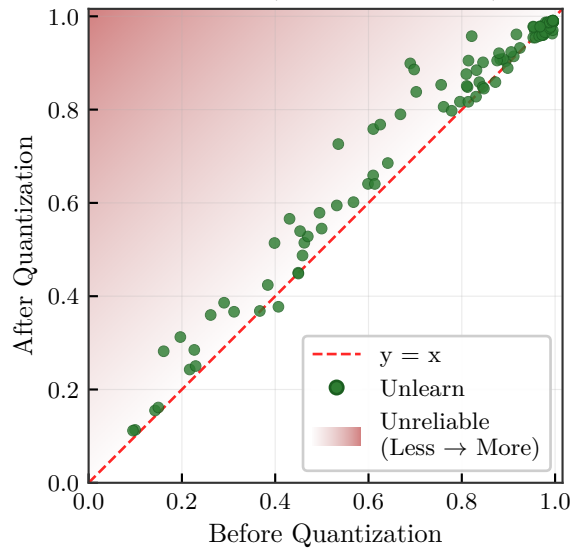

Para. ROUGE
Q=0.949 (n=124, unrel=38)



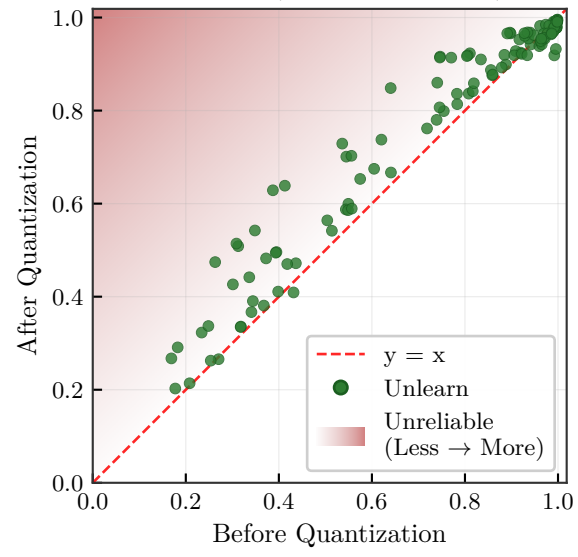

MIA-LOSS (raw AUC)
Q=0.953 (n=124, unrel=71)



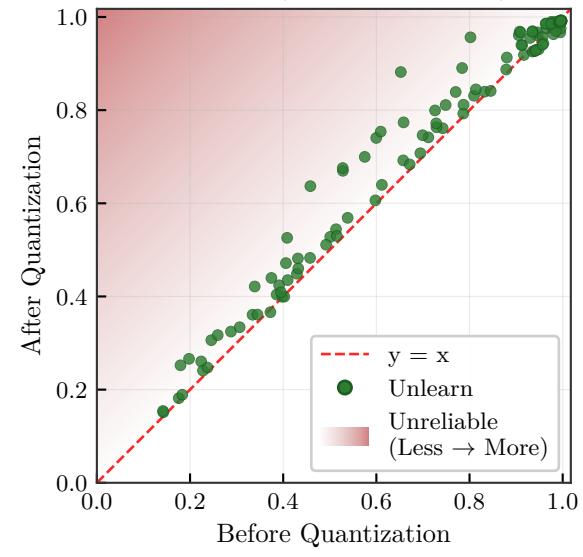

MIA-MinK (raw AUC)
Q=0.948 (n=124, unrel=72)



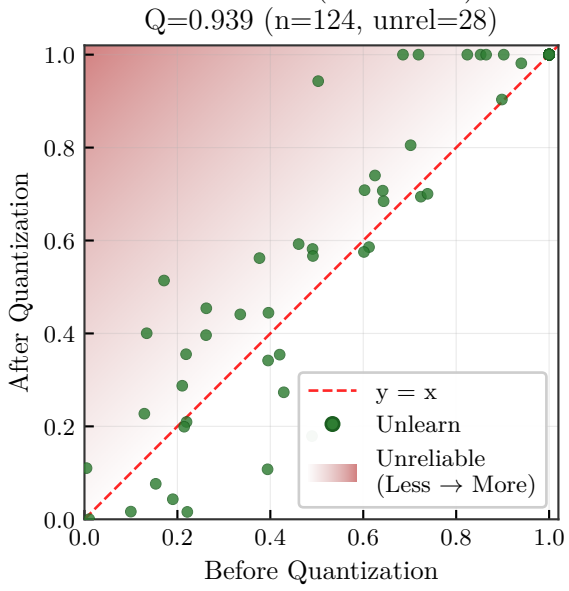

MIA-MinK++ (raw AUC)
Q=0.923 (n=124, unrel=81)



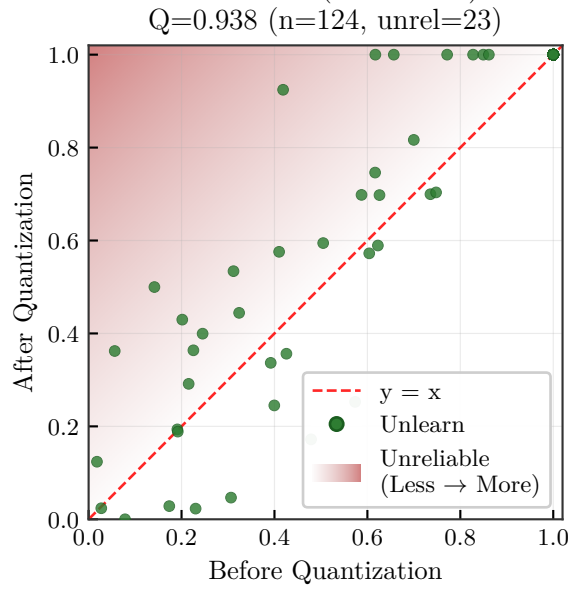

MIA-ZLib (raw AUC)
Q=0.949 (n=124, unrel=83)



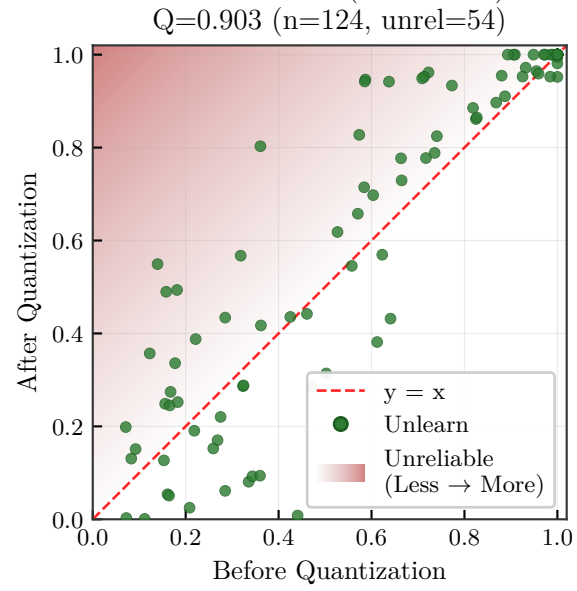

MIA-LOSS (normalized)
Q=0.939 (n=124, unrel=28)



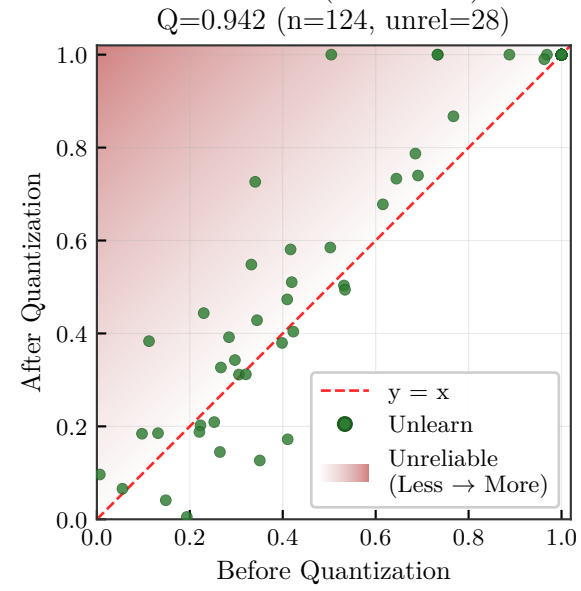

MIA-MinK (normalized)
Q=0.938 (n=124, unrel=23)



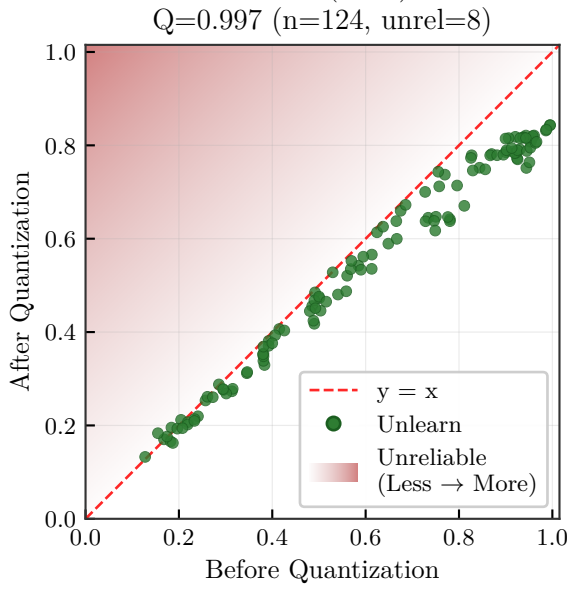

MIA-MinK++ (normalized)
Q=0.903 (n=124, unrel=54)




MIA-ZLib (normalized)
Q=0.942 (n=124, unrel=28)




1-UDS (Ours)
Q=0.997 (n=124, unrel=8)




Logit Lens
Q=1.000 (n=124, unrel=1)