Relearning Robustness (13 Metrics + 4 Normalized MIA + 3 Rep. Baselines)
(150 Unlearned Models; Utility + Before-Value Filtered)

Exact Memorization
R=0.707 (n=110, over=90, under=20)

Extraction Strength
R=0.384 (n=124, over=118, under=6)

Truth Ratio
R=0.389 (n=124, over=77, under=47)

Prob.
R=0.674 (n=124, over=95, under=29)

ROUGE
R=0.210 (n=124, over=124, under=0)

Jailbreak ROUGE
R=0.173 (n=124, over=120, under=4)

Para. Prob.
R=0.604 (n=124, over=30, under=94)

Para. ROUGE
R=0.077 (n=124, over=120, under=4)

MIA-LOSS (raw AUC)
R=0.650 (n=55, over=50, under=5)

MIA-MinK (raw AUC)
R=0.646 (n=51, over=45, under=6)

MIA-MinK++ (raw AUC)
R=0.583 (n=59, over=58, under=1)

MIA-ZLib (raw AUC)
R=0.605 (n=64, over=63, under=1)

MIA-LOSS (normalized)
R=0.719 (n=29, over=26, under=3)

MIA-MinK (normalized)
R=0.746 (n=23, over=21, under=2)

MIA-MinK++ (normalized)
R=0.634 (n=52, over=44, under=8)

MIA-ZLib (normalized)
R=0.783 (n=31, over=27, under=4)

1-CKA
R=0.000 (n=0, over=0, under=0)

1-Fisher (0.1%)
R=0.993 (n=24, over=19, under=5)

1-Logit Lens
R=0.863 (n=69, over=10, under=59)

1-UDS (Ours)
R=0.898 (n=67, over=28, under=39)

Legend (repeated in each panel):
- y = x + Δ_{ret}
- y = x
- Unlearn
- Retain
- Unreliable (Less → More)

Axes (each panel): x = Before Relearning, y = After Relearning