Lexical Ablation: Probe Robustness to Keyword Removal

Legend: Original Text, Keywords Removed

Metric	Original Text	Keywords Removed
AUROC	1.00	1.00
Accuracy	0.93	0.93
Separation		4.7

41 keywords removed (avg 13.5/pair)
AUROC unchanged: 1.0 → 1.0