

SAE Probes vs Best Baseline Method Across All Settings
Mean improvement: 0.0758 | 78.0% improved | n=2369 datasets

