Model-Agnostic Testing: Attribution Methods Across FR Architectures (n=500) 80 **Attribution Method** Grad-CAM 70 SHAP 60 (%) **Falsification Rate** 50 45.6 34.0 40 35.3 30 20 10 0 ArcFace CosFace SphereFace **Face Recognition Model**