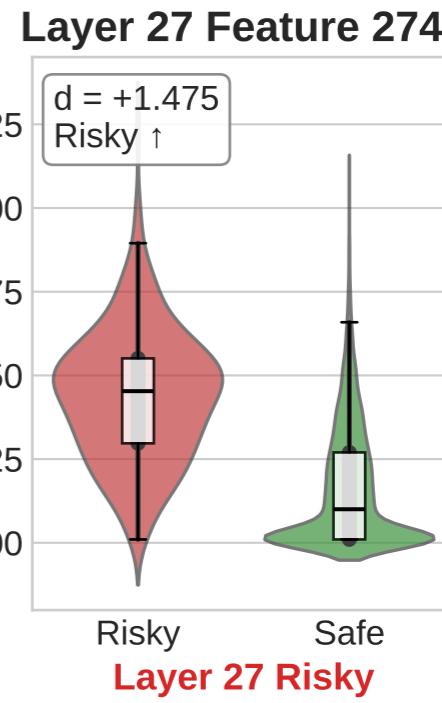
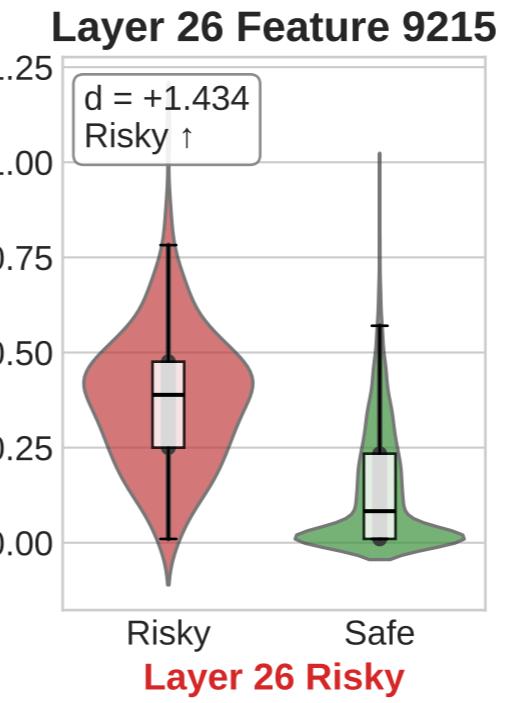
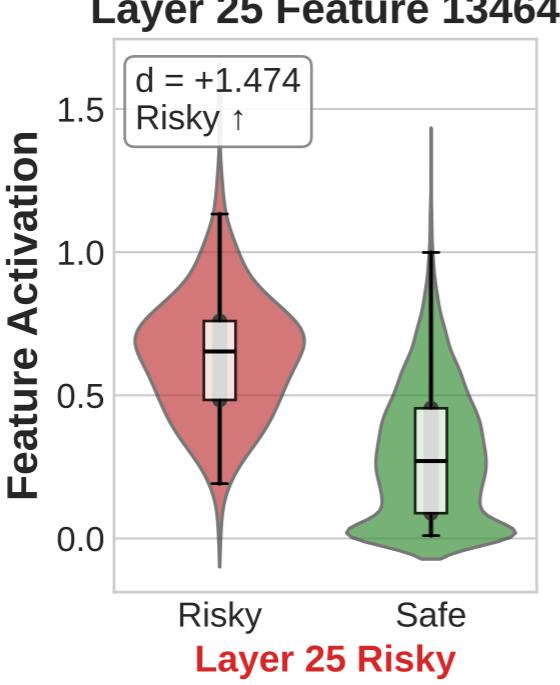
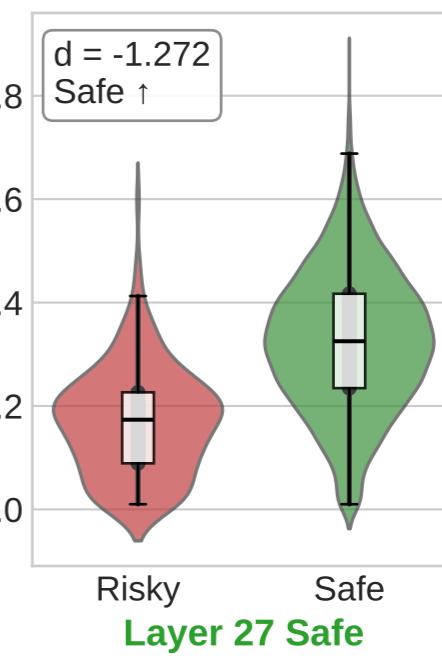
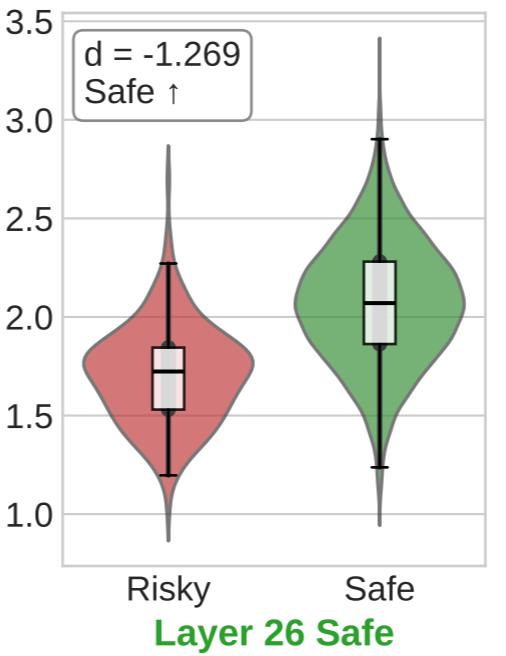
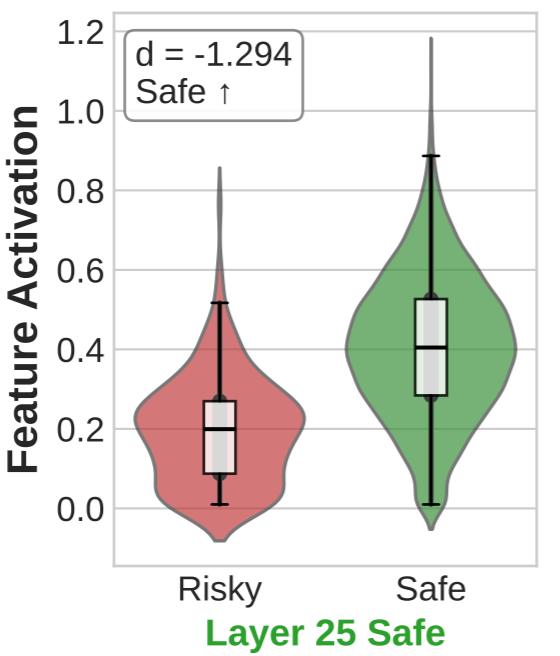


SAE Feature Separation Analysis - Complete Layer Analysis

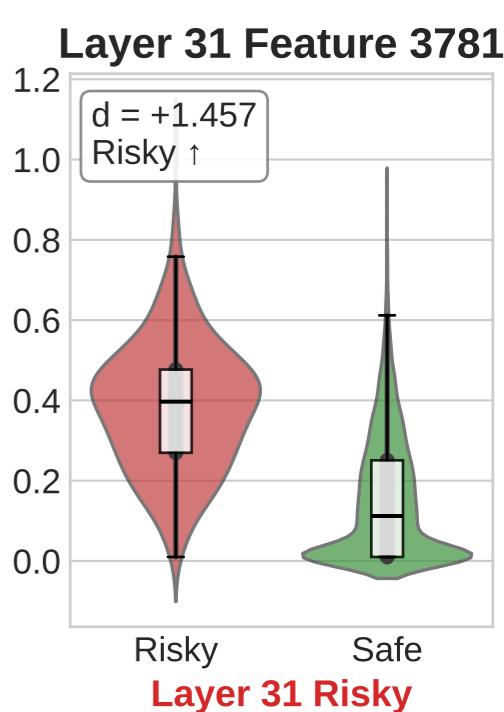
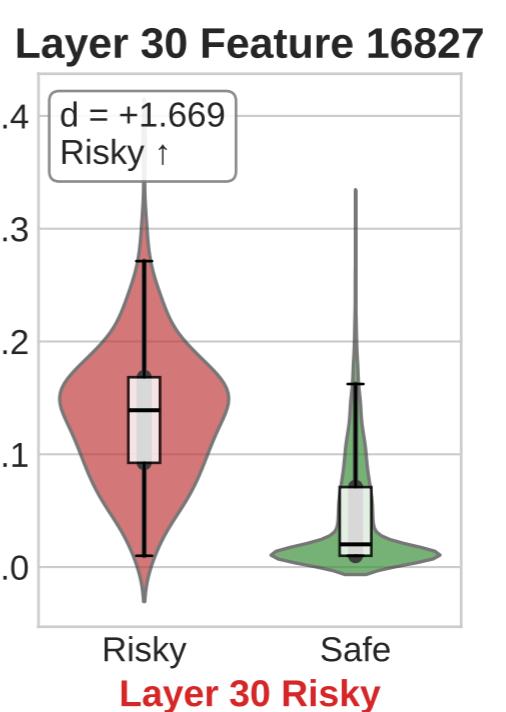
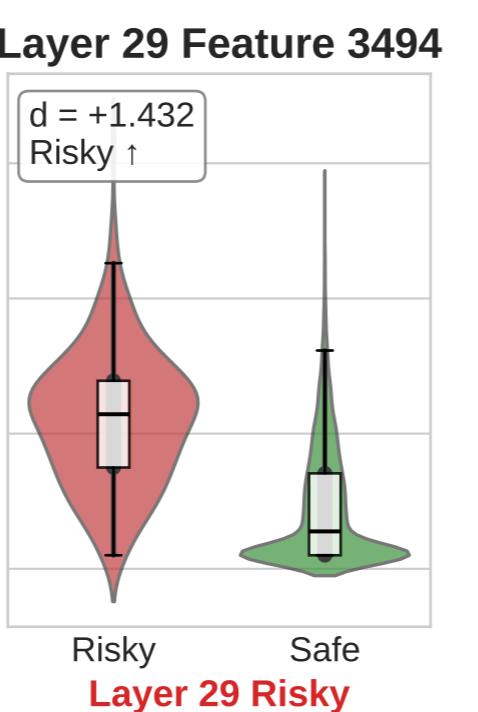
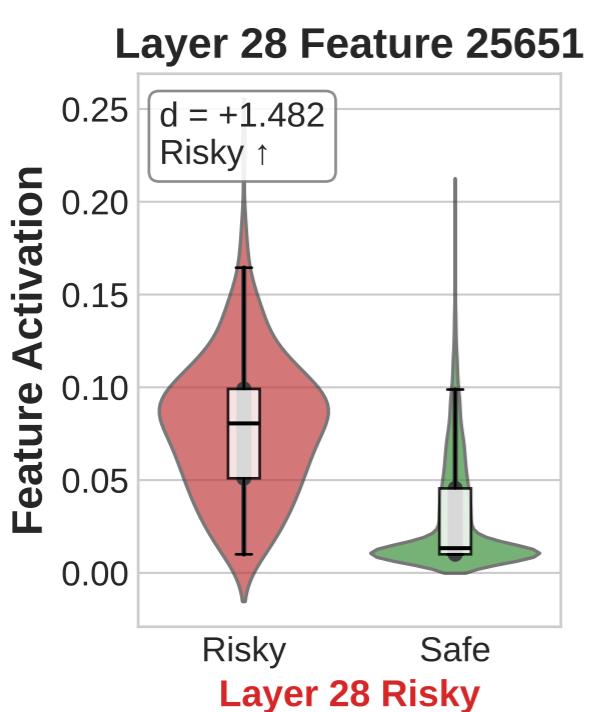
Layers 25-27
Risky Features



Layers 25-27
Safe Features



Layers 28-31
Risky Features



Layers 28-31
Safe Features

