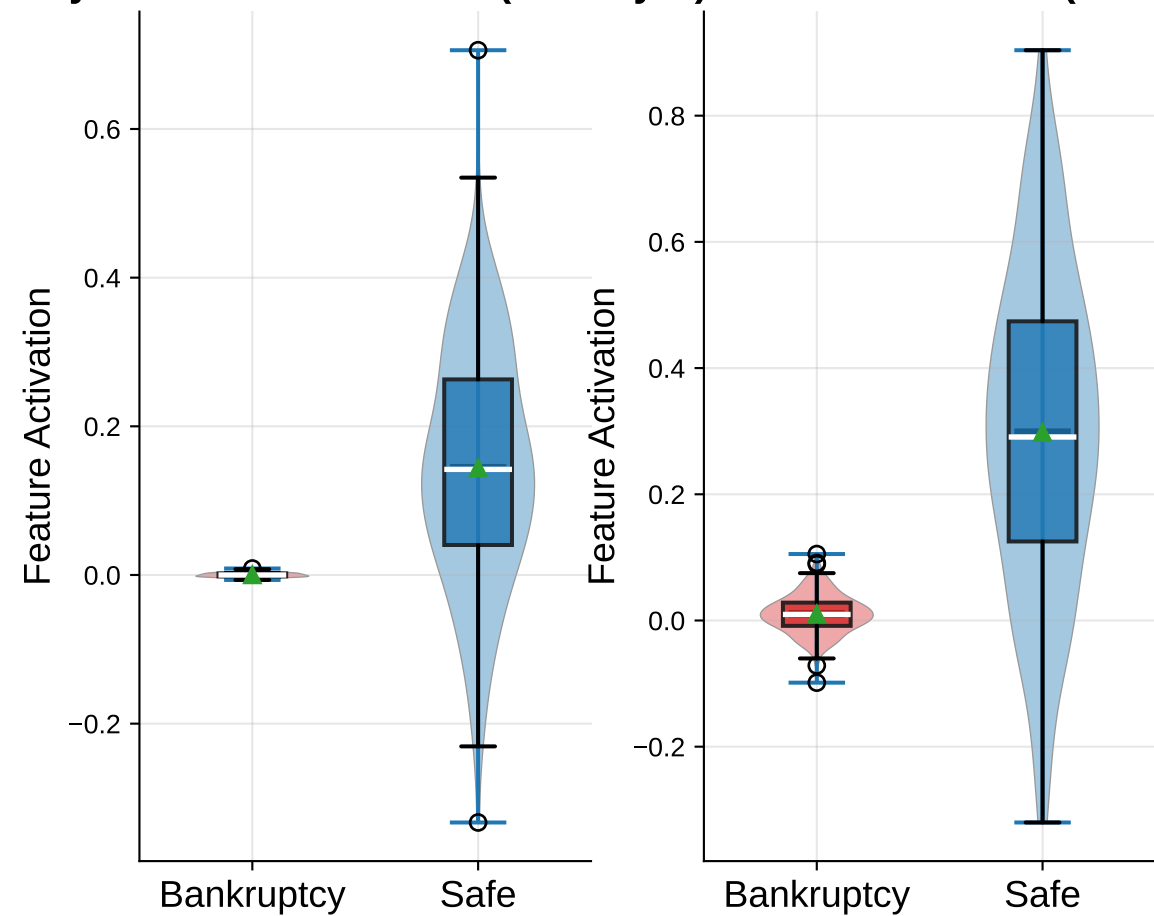
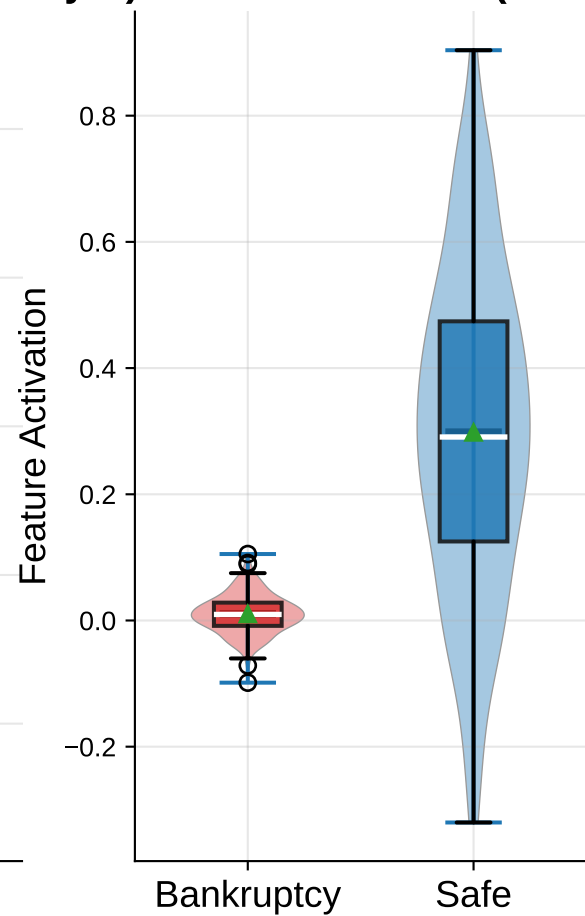


SAE Feature Activation Distributions: Bankruptcy vs Safe Groups

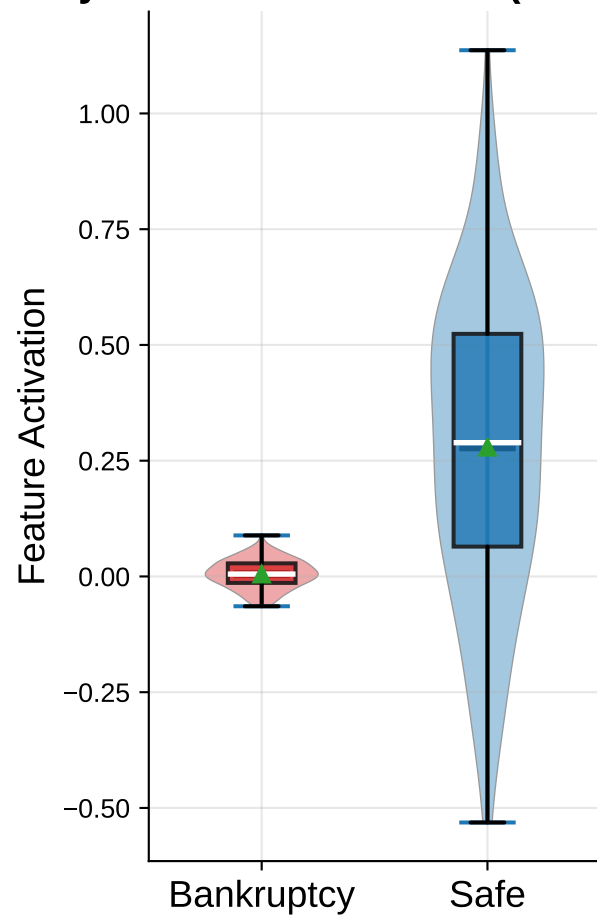
Layer 25 - Feature 17793 (d = -1.047)



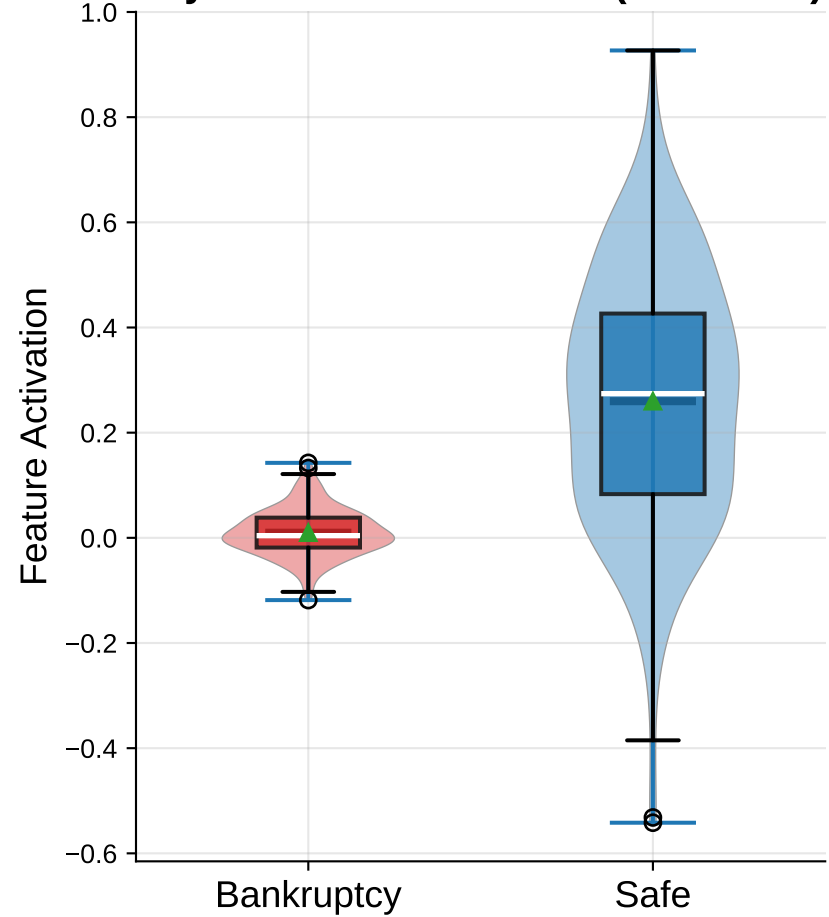
Layer 26 - Feature 31257 (d = -1.065)



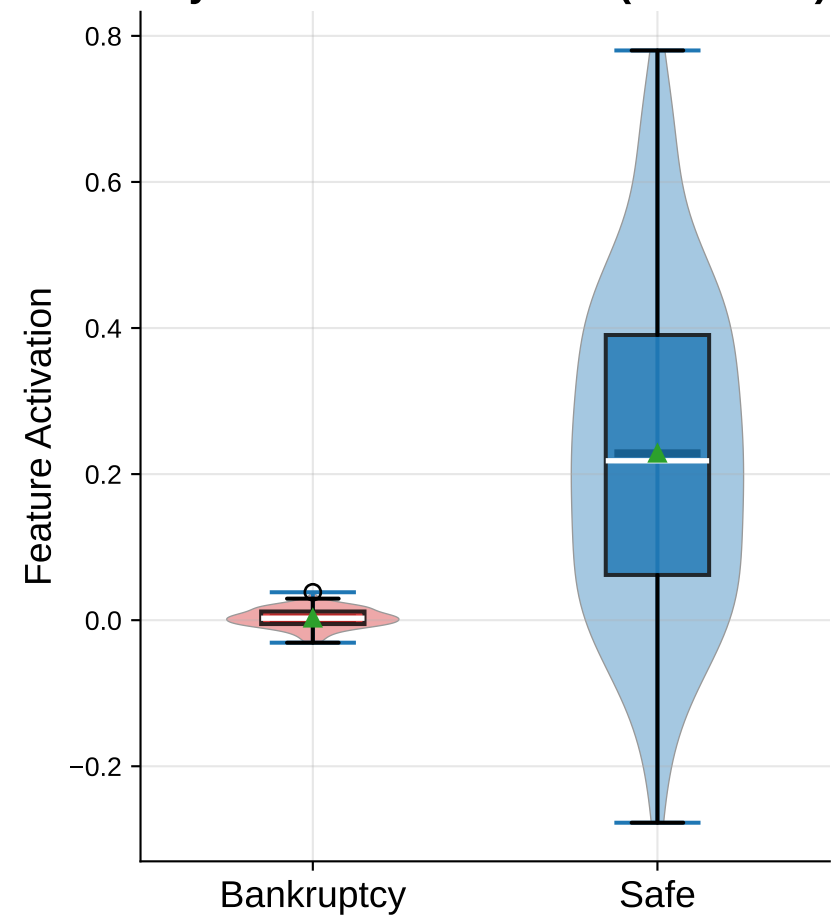
Layer 27 - Feature 32051 (d = -1.086)



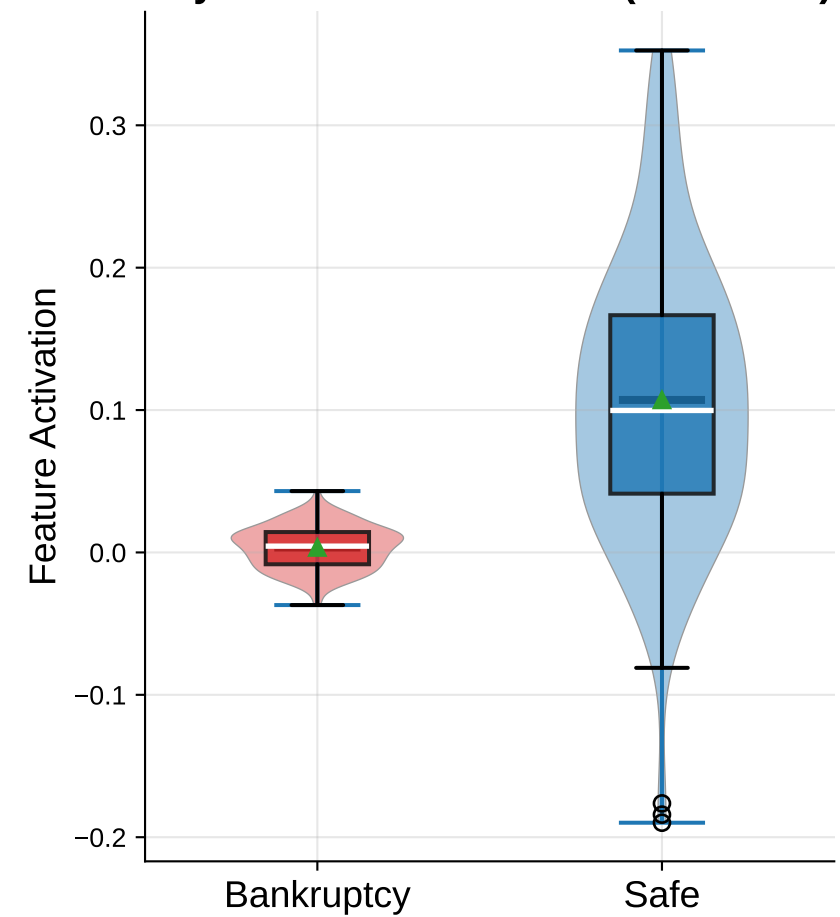
Layer 28 - Feature 21135 (d = -1.098)



Layer 29 - Feature 29399 (d = -0.998)



Layer 30 - Feature 17910 (d = -1.097)



Layer 31 - Feature 29127 (d = -1.116)

