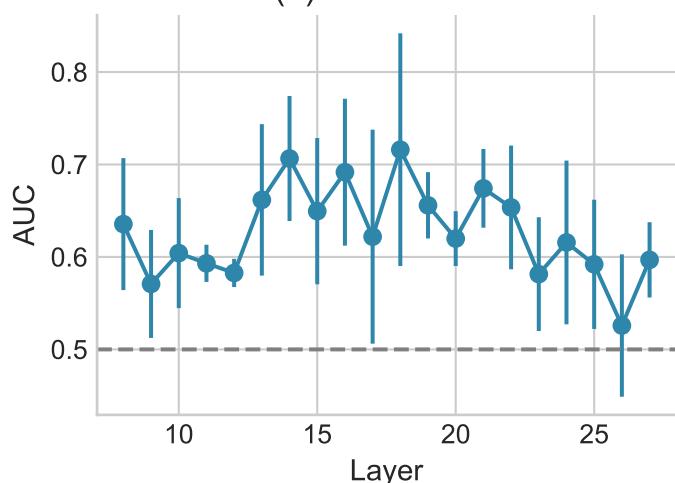
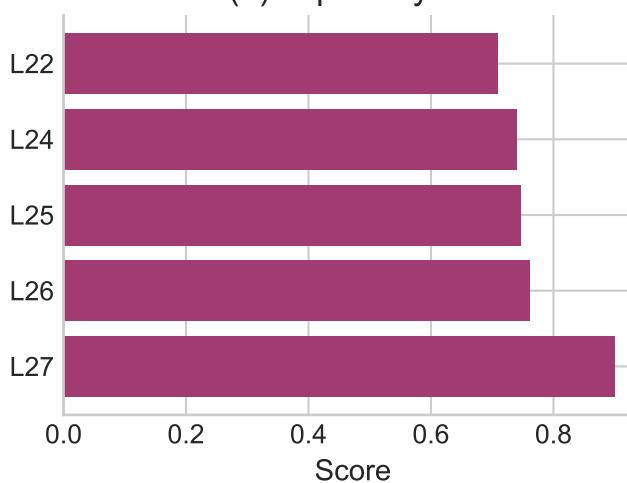


# Phase A: Reasoning Subspace Analysis Summary

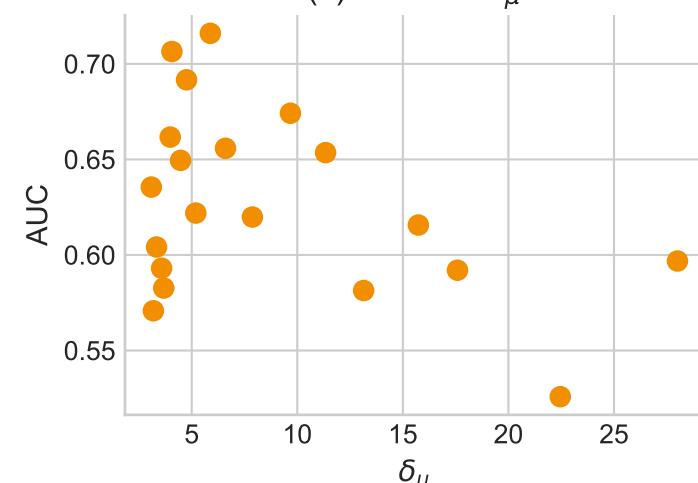
(a) Probe AUC



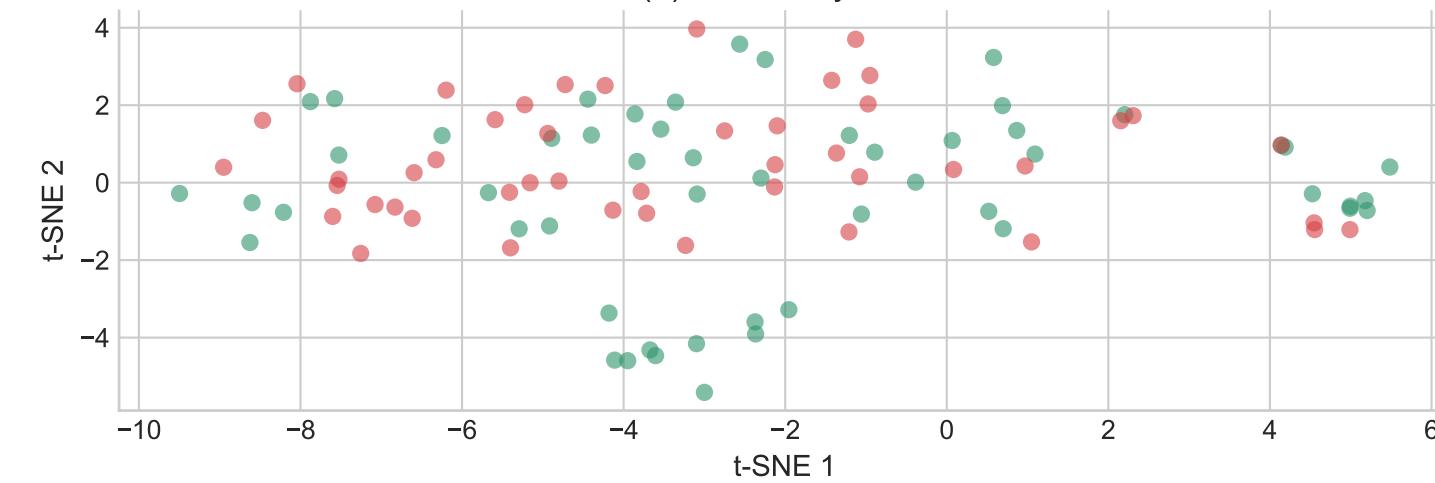
(b) Top 5 Layers



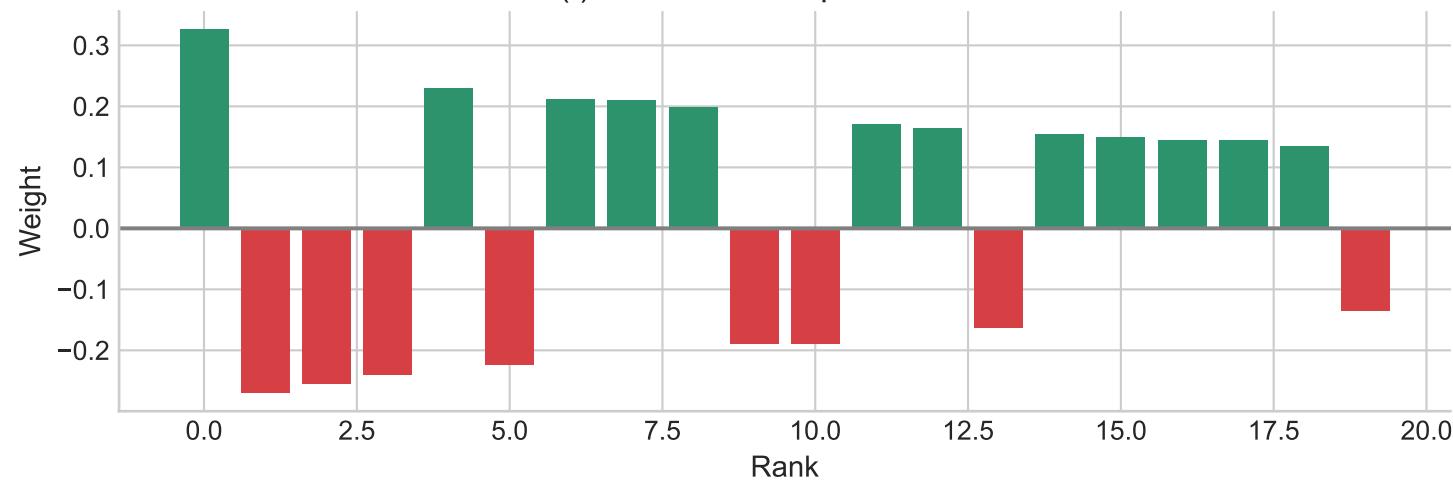
(c) AUC vs  $\delta_\mu$



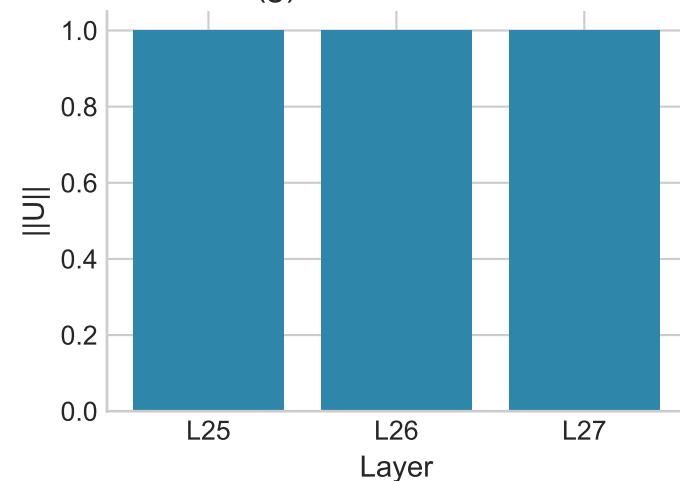
(d) t-SNE Layer 27



(f) Direction Components L27



(g) Direction Norm



(e) Labels

