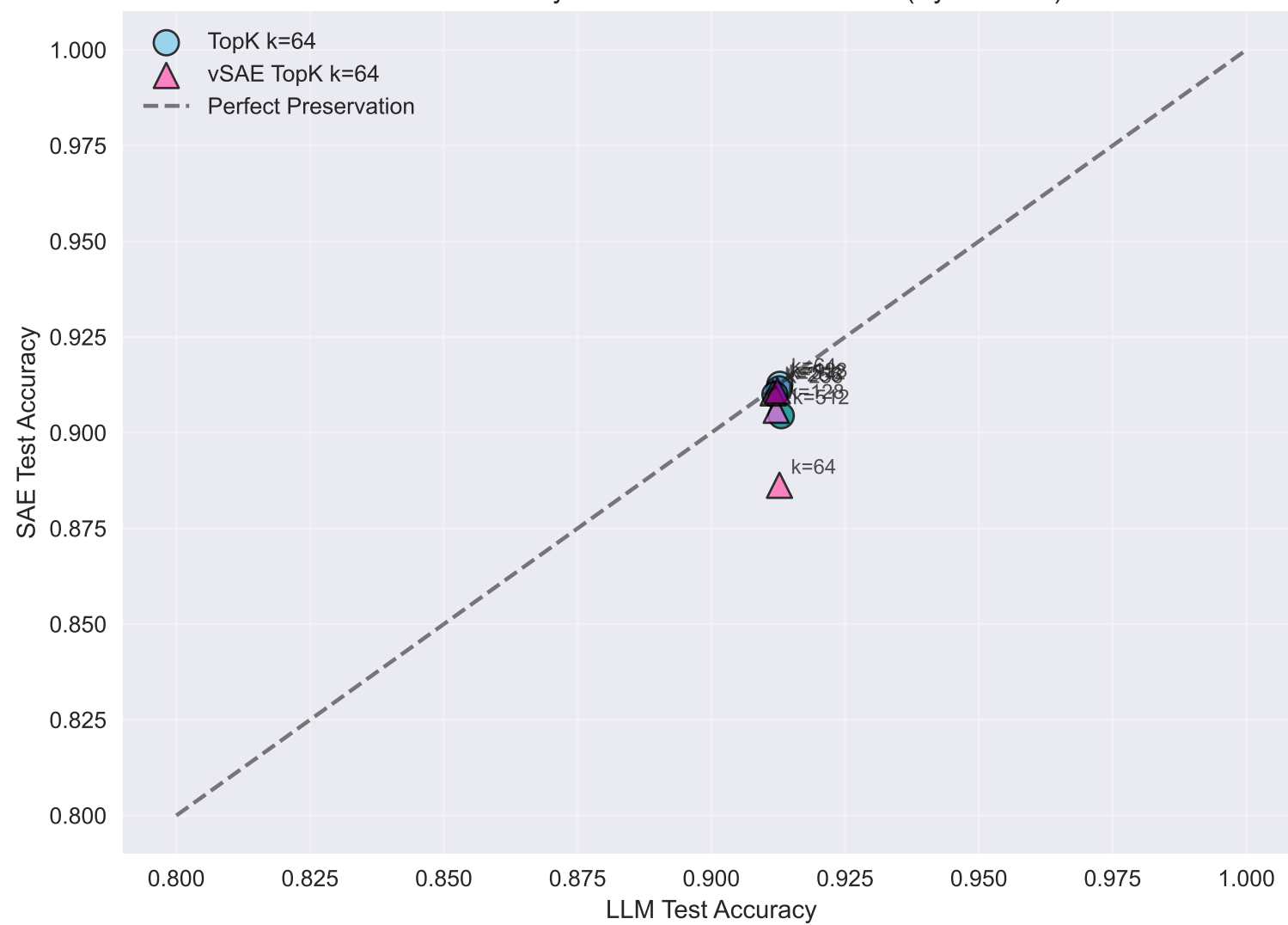
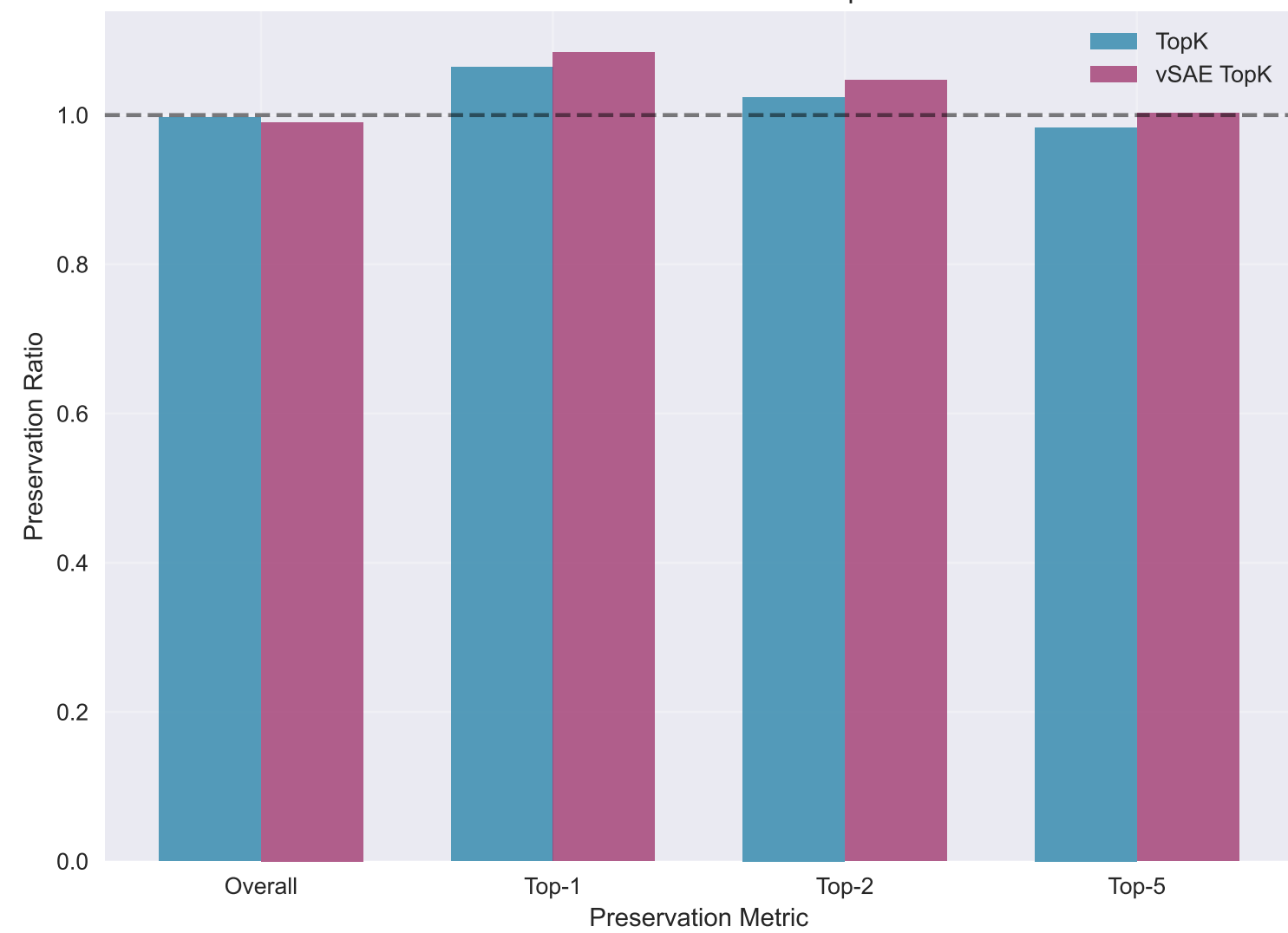


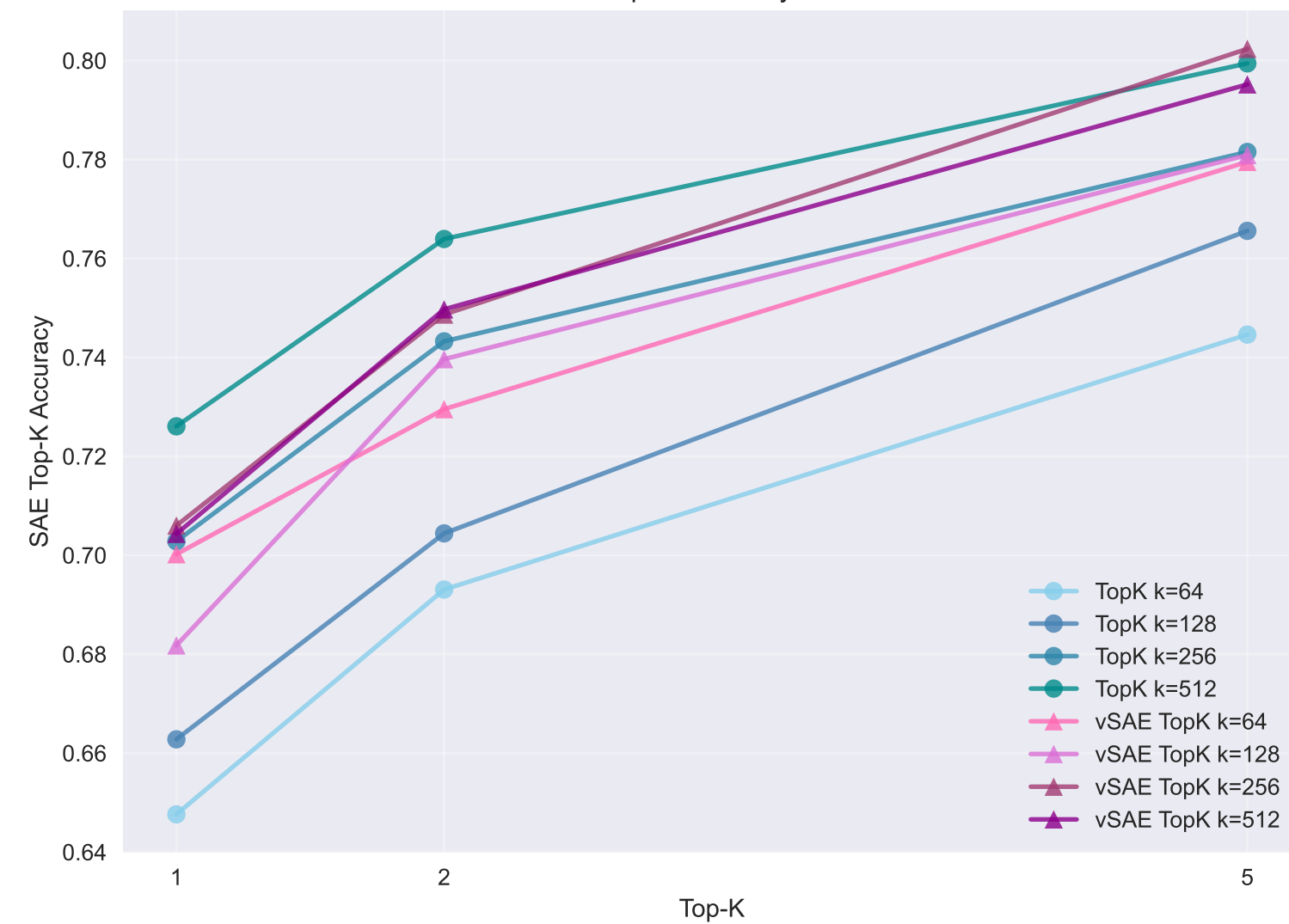
Overall Accuracy: SAE vs LLM Performance (Pythia-70M)



Performance Preservation Comparison



SAE Top-K Accuracy Trends



Performance Degradation by K Value

