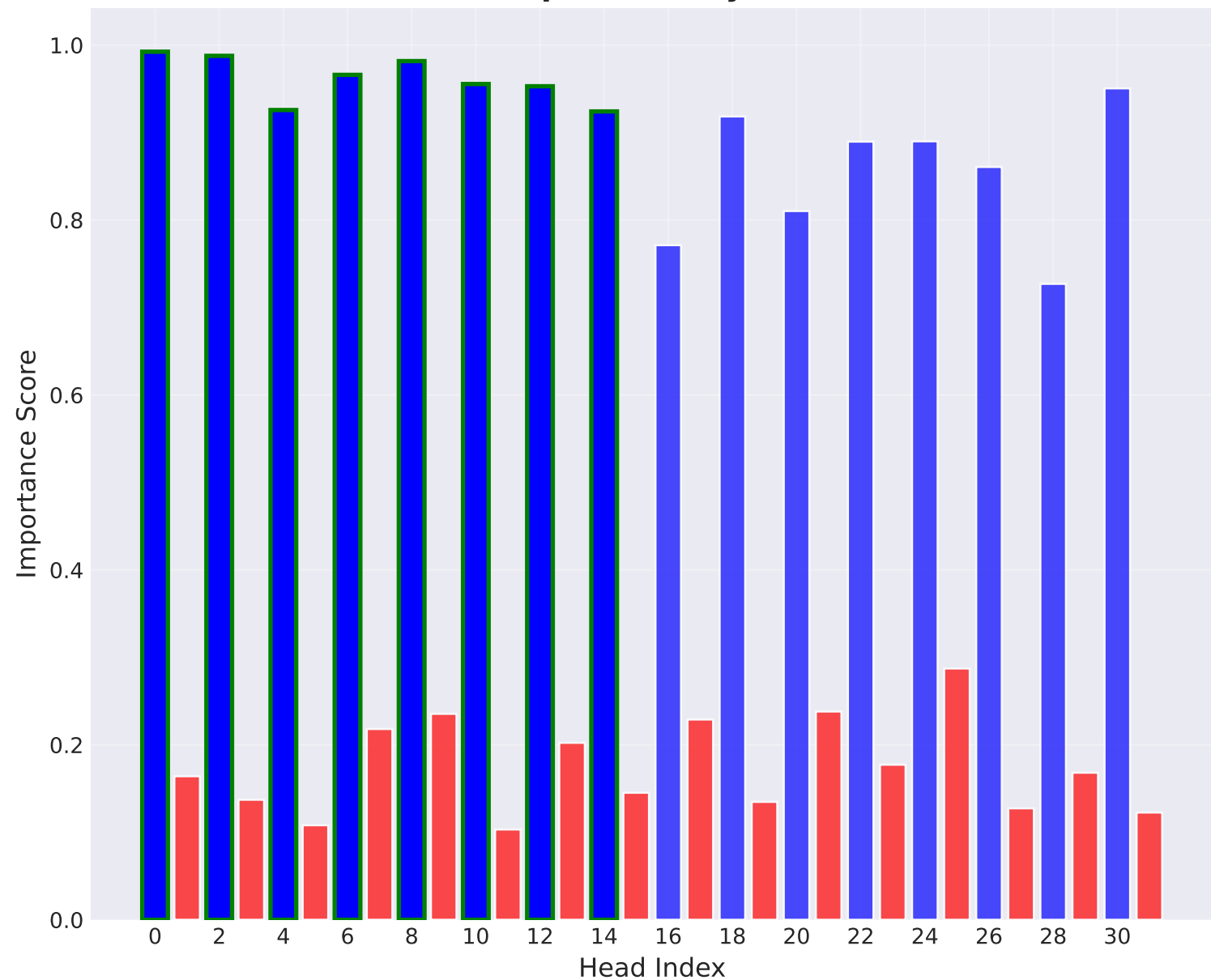
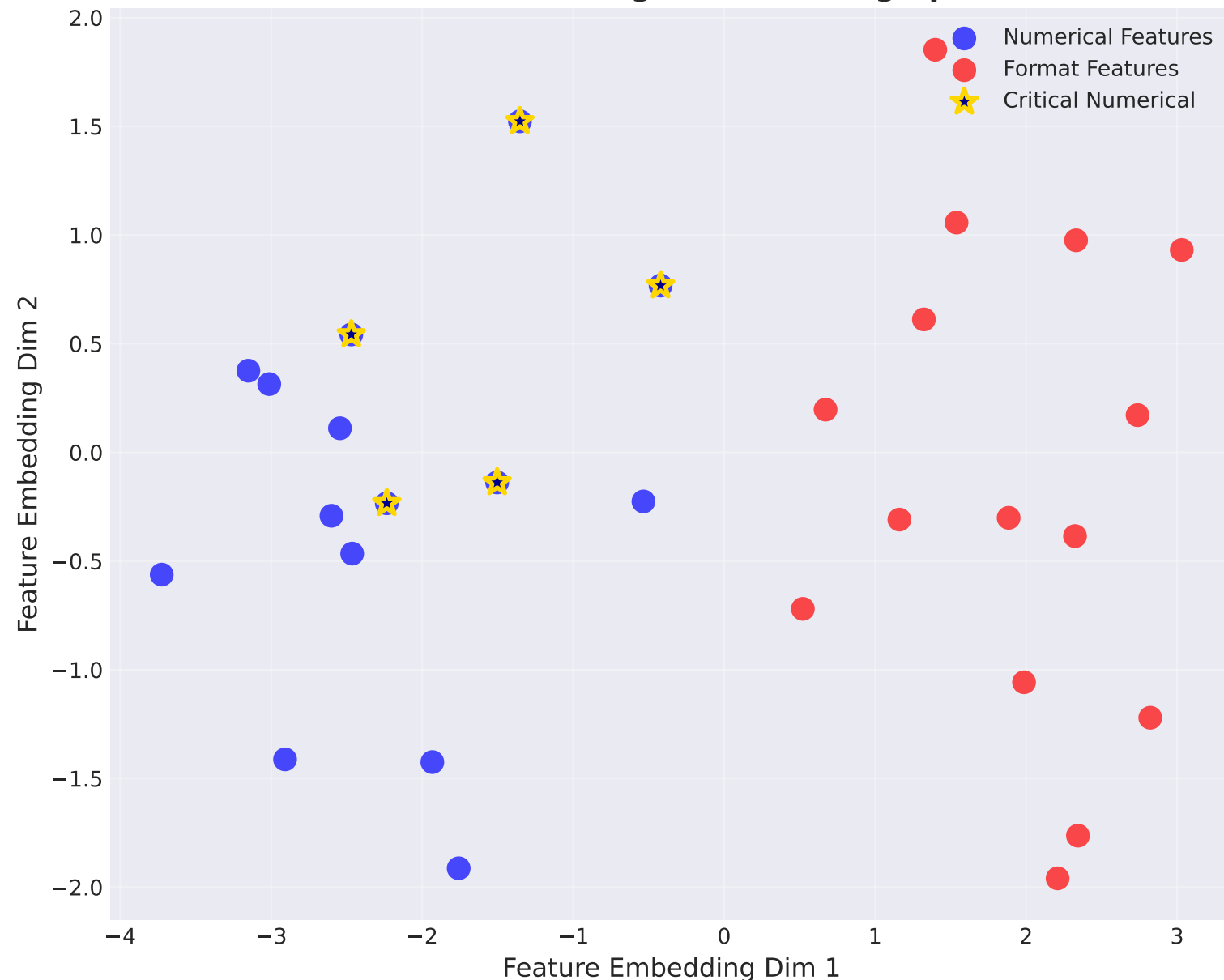


Detailed Feature-Head Correlation Analysis

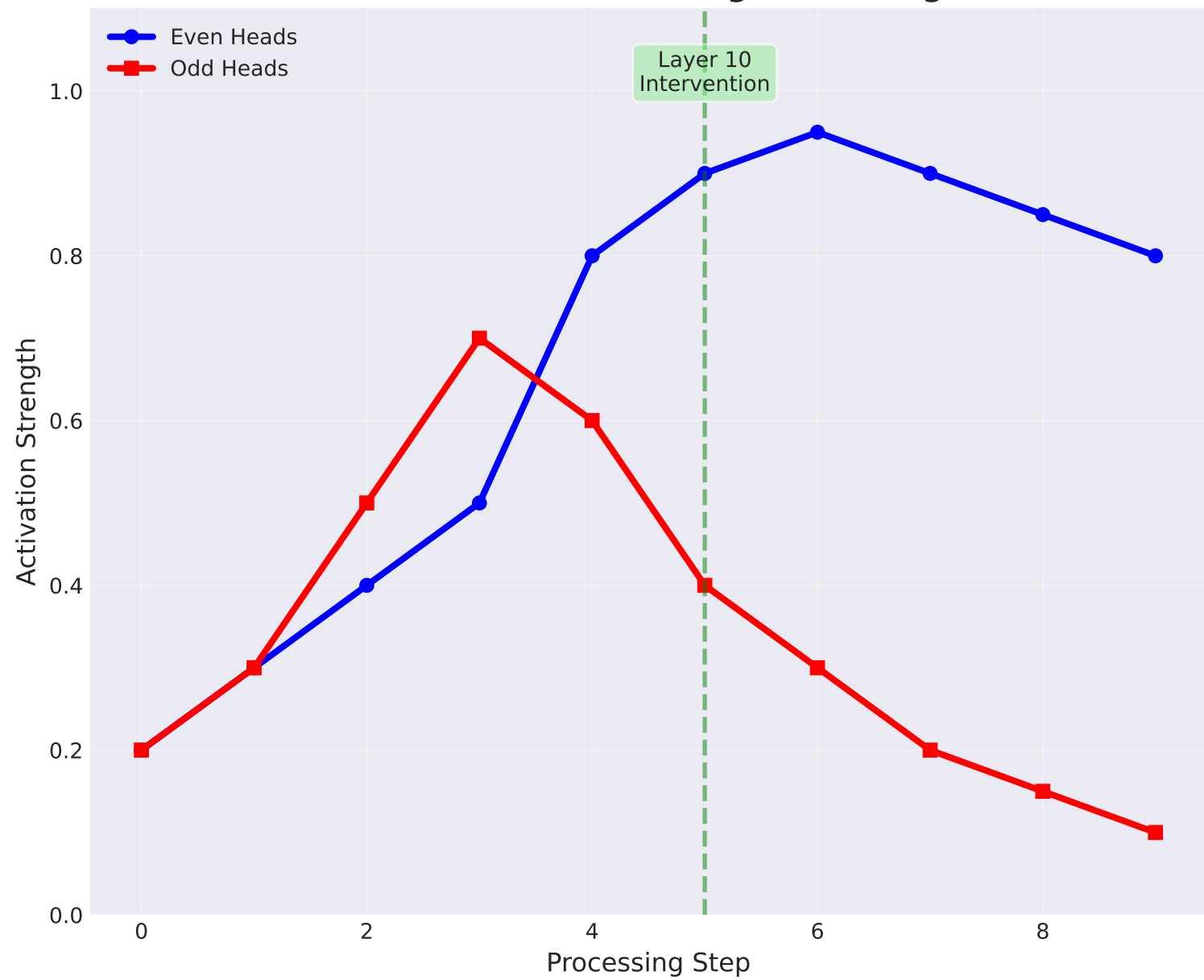
Feature Importance by Head Index



Feature Clustering in Embedding Space



Head Activation During Processing



Intervention Success by Head Configuration

