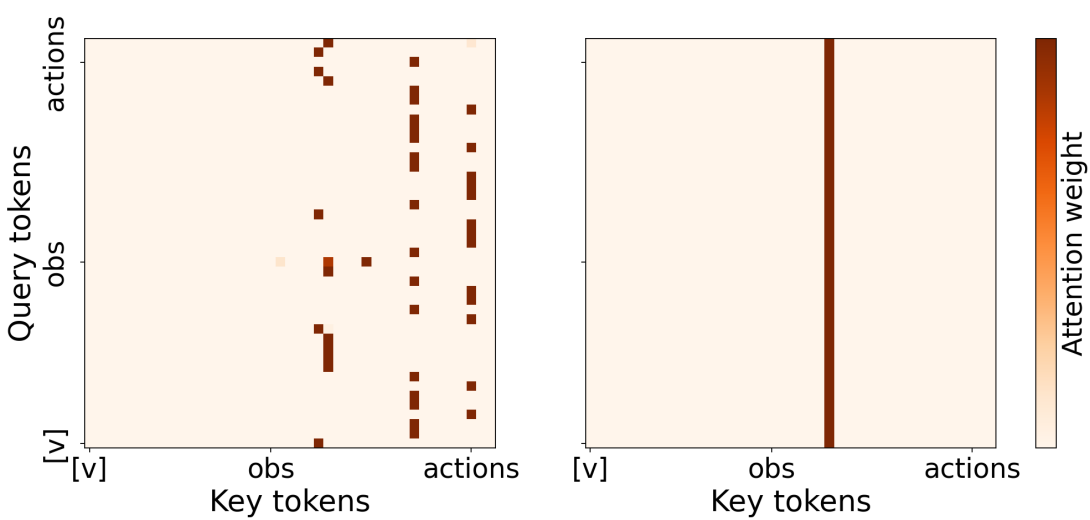


Layer 0 Step 0

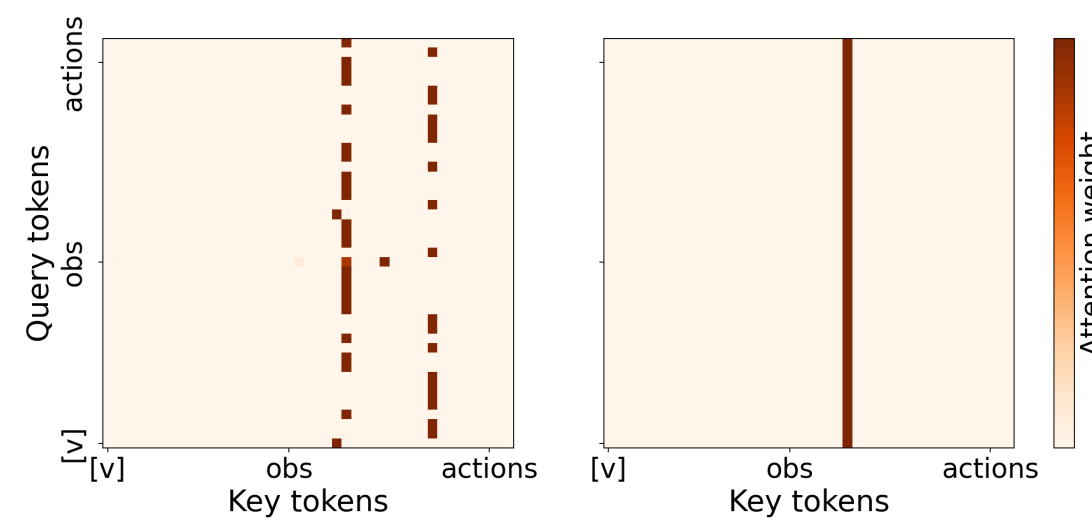
Layer 1



Step 100

Layer 0

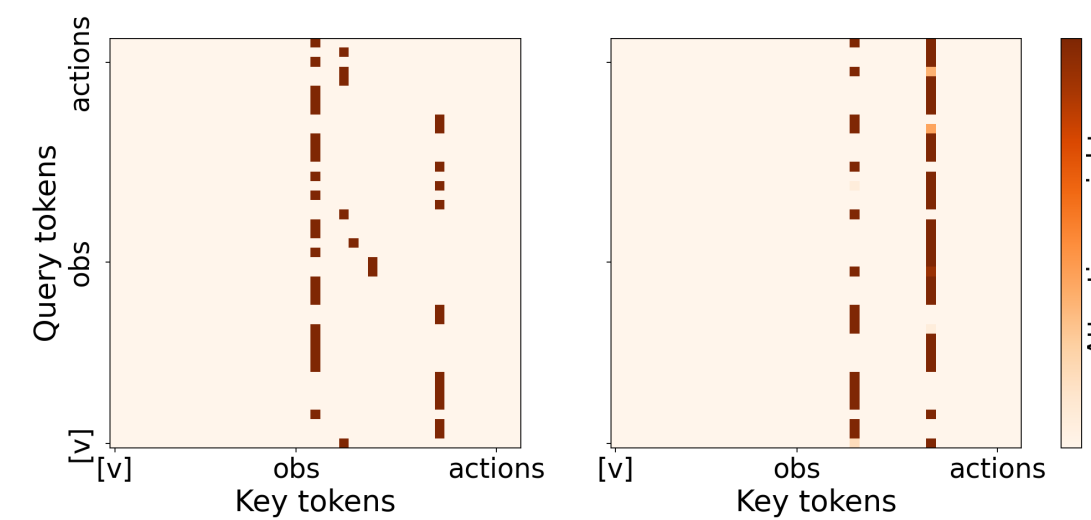
Layer 1



Step 200

Layer 0

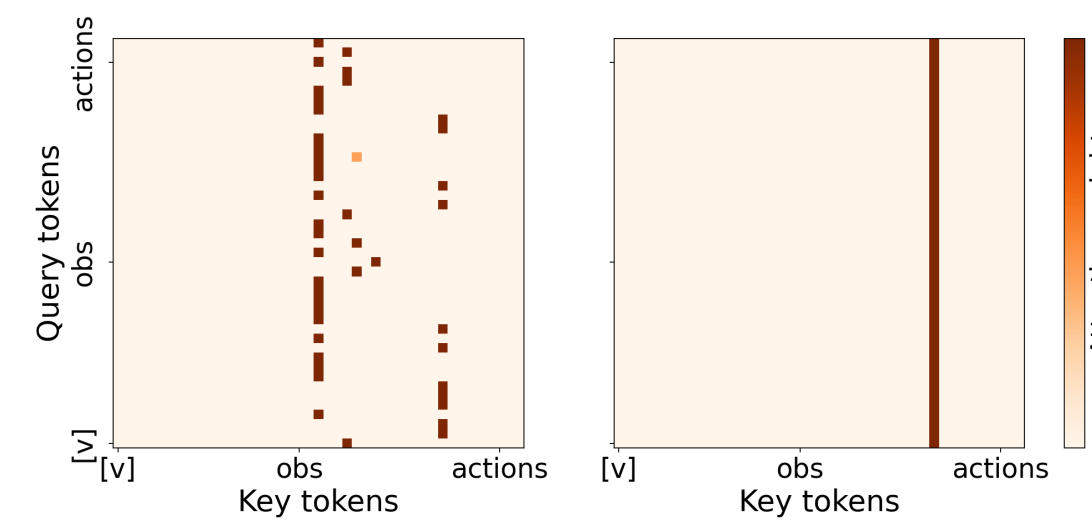
Layer 1



Step 300

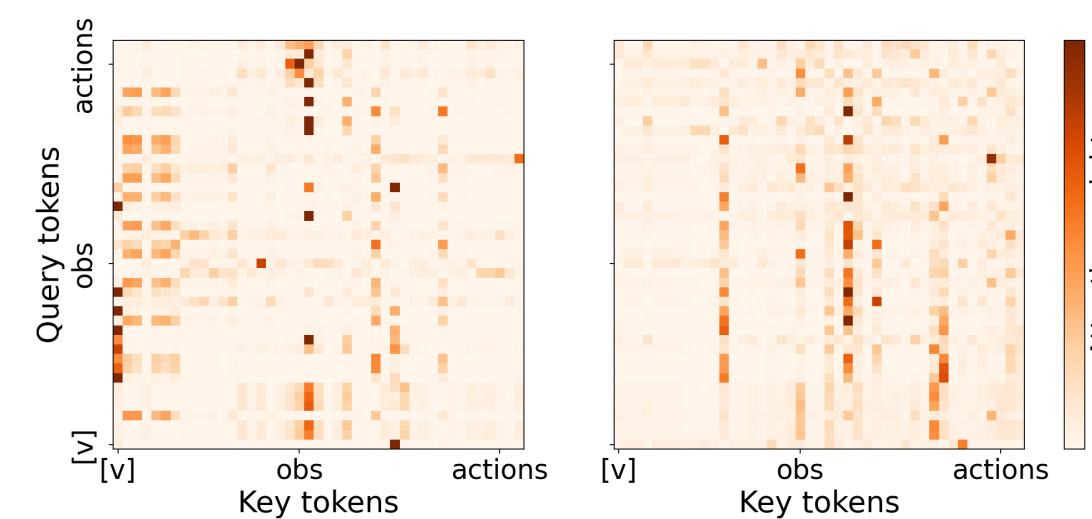
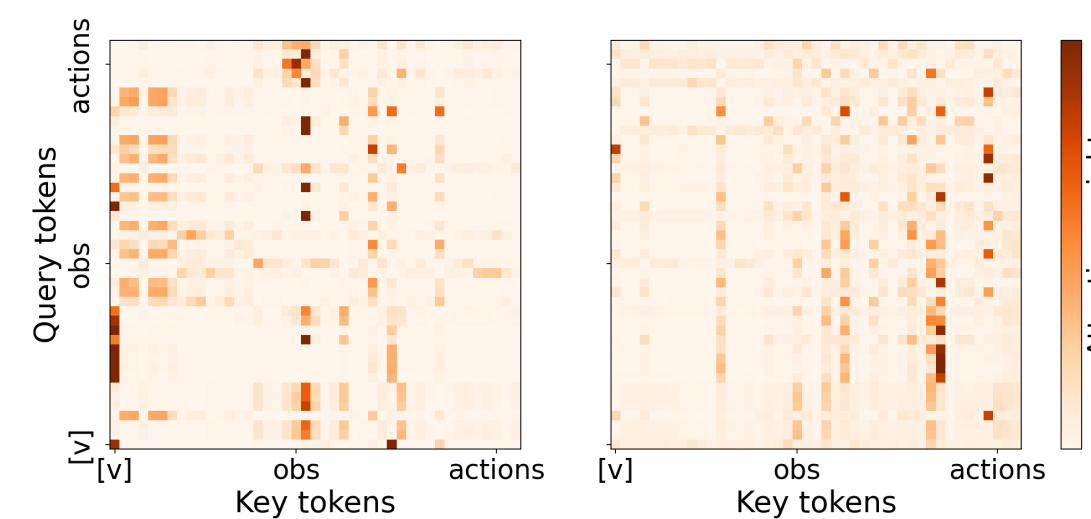
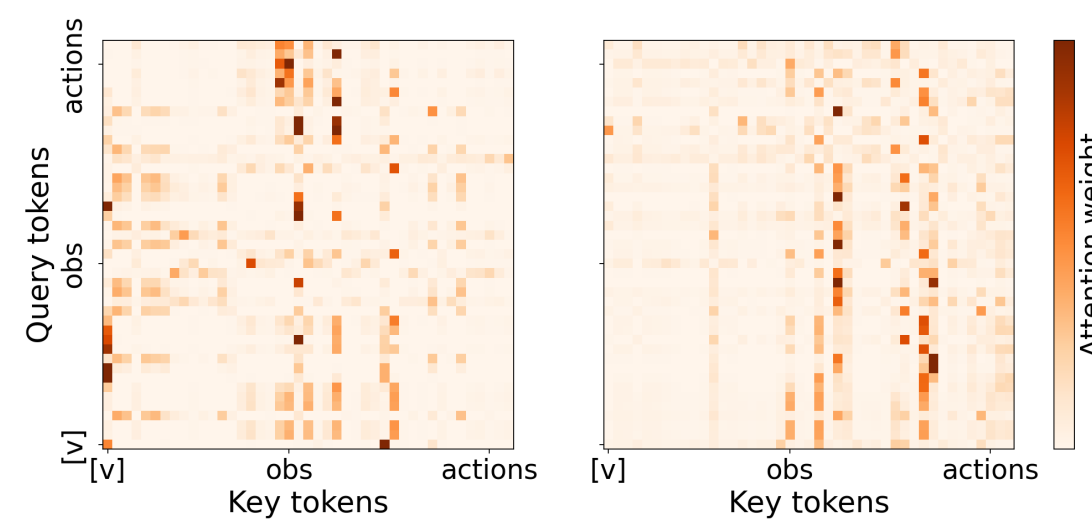
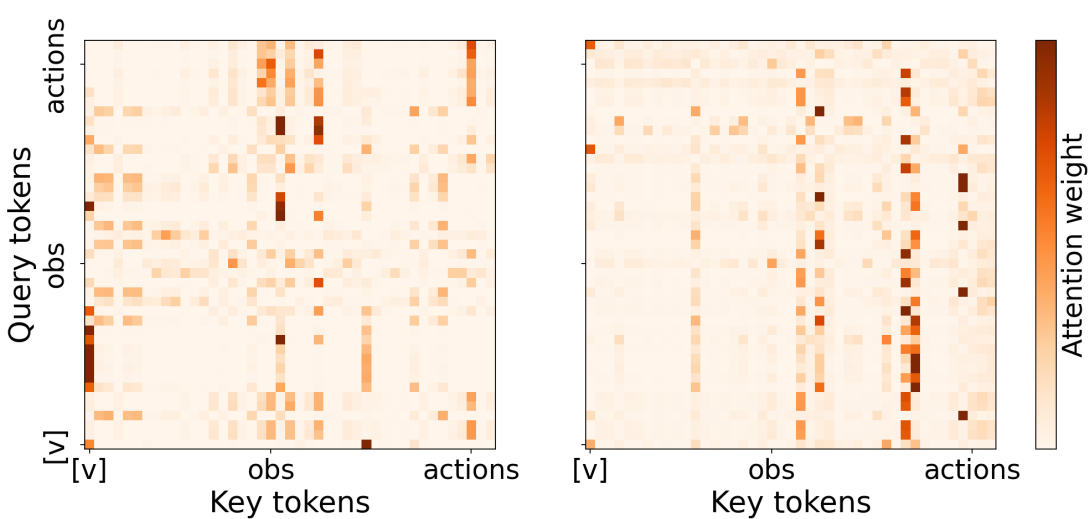
Layer 0

Layer 1



Transformer without TQL Attention Map

QK Norm Attention Map



$\sigma$ Reparam Attention Map

