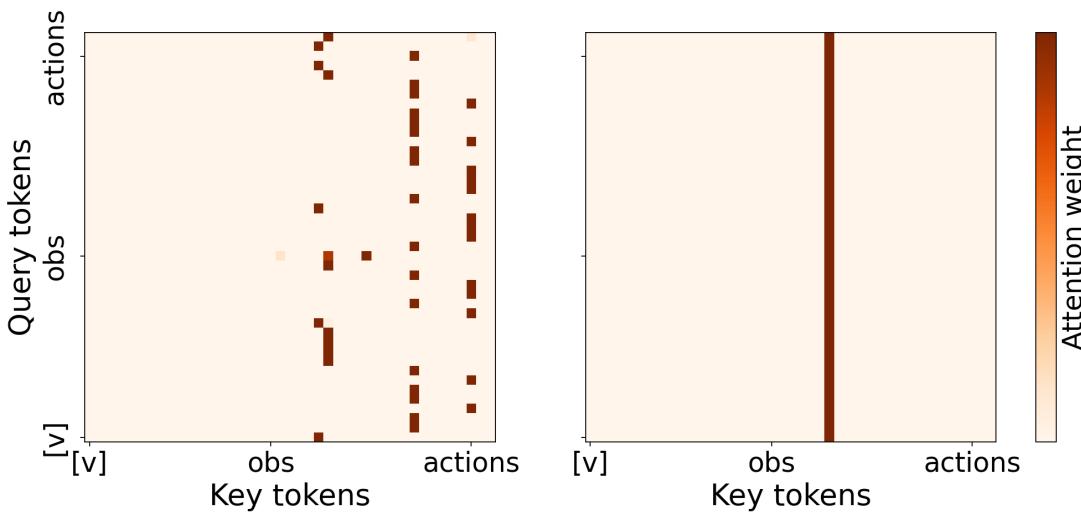


Layer 0 Step 0

Layer 1

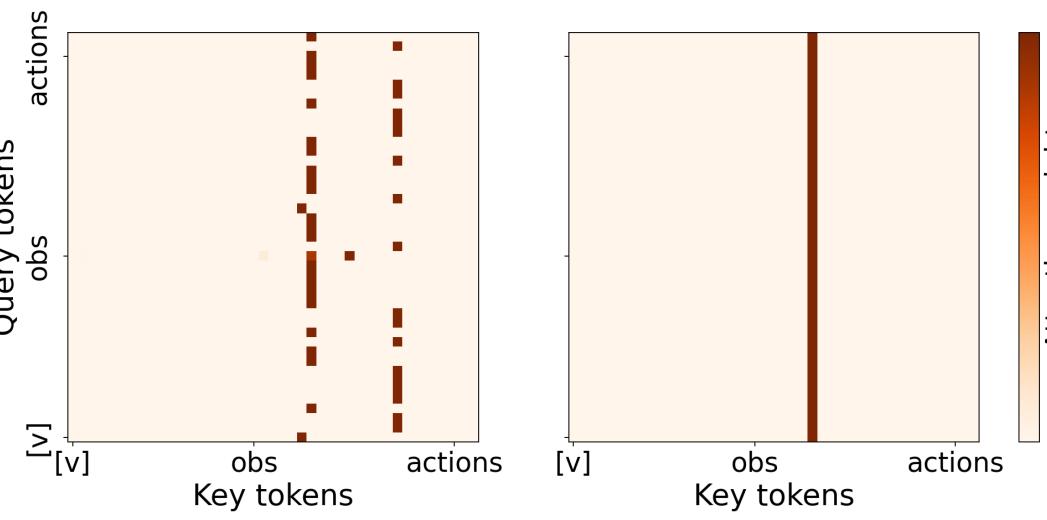


Step 100

Layer 0

Layer 1

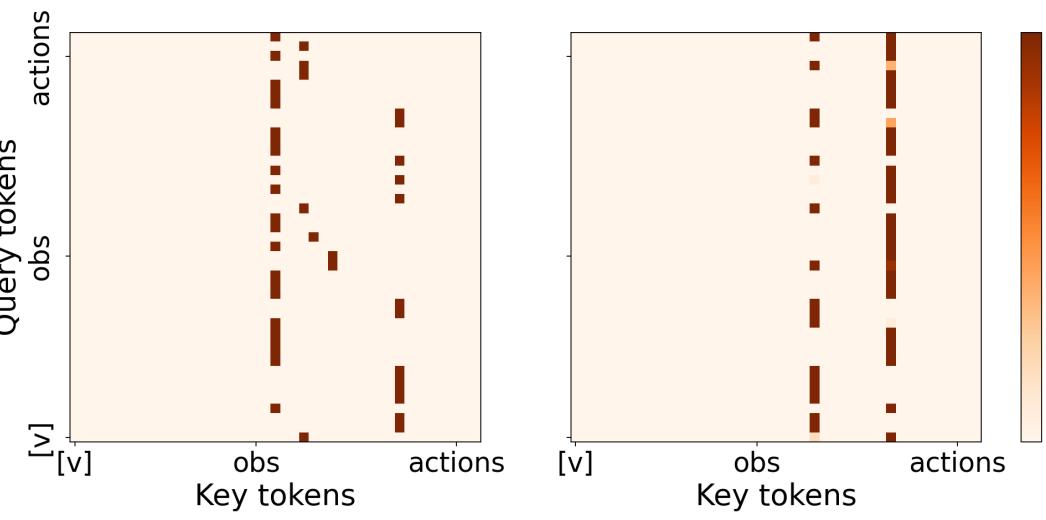
Transformer without TQL Attention Map



Step 200

Layer 0

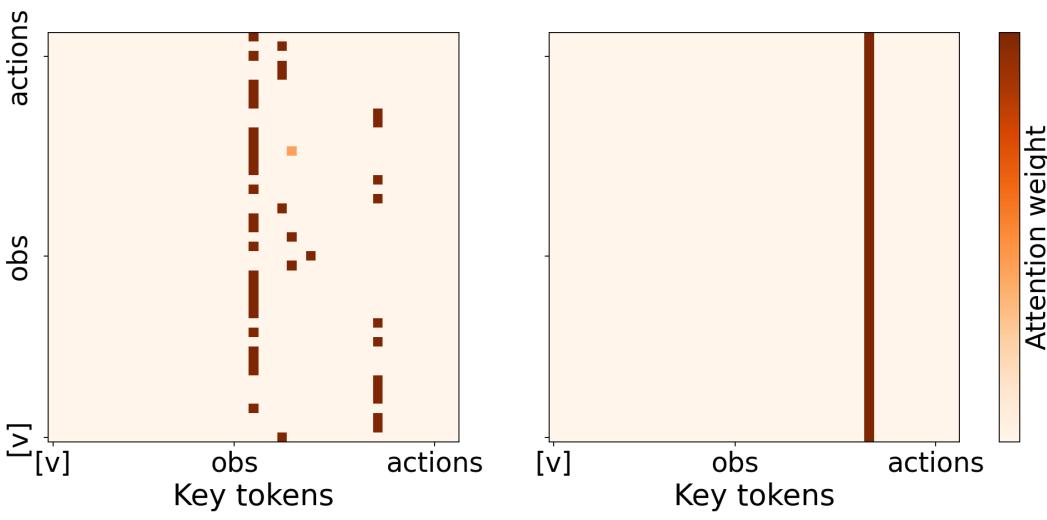
Layer 1



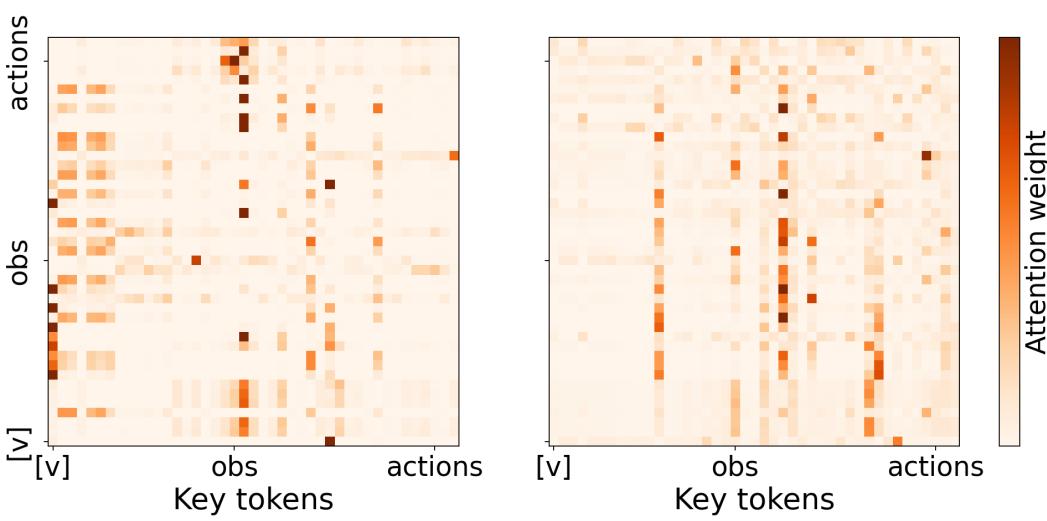
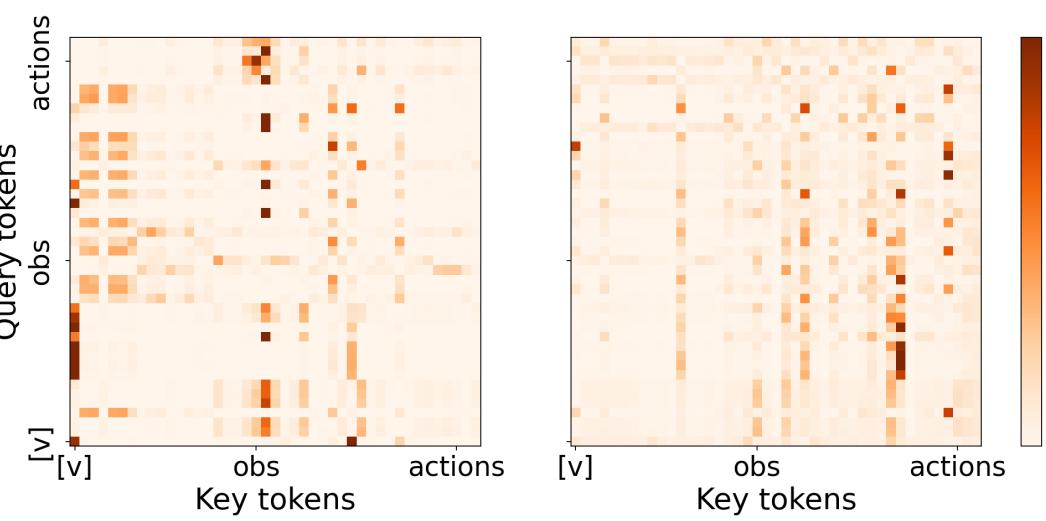
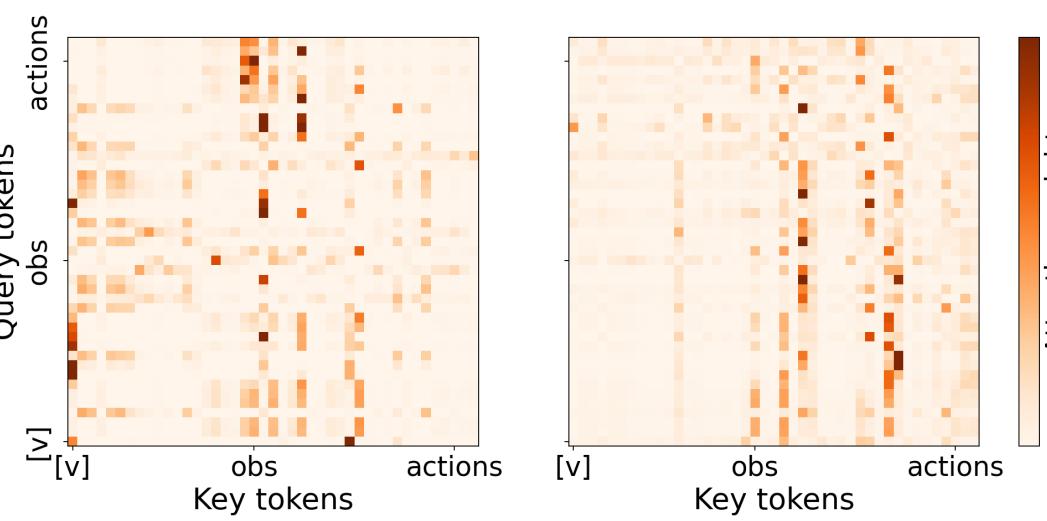
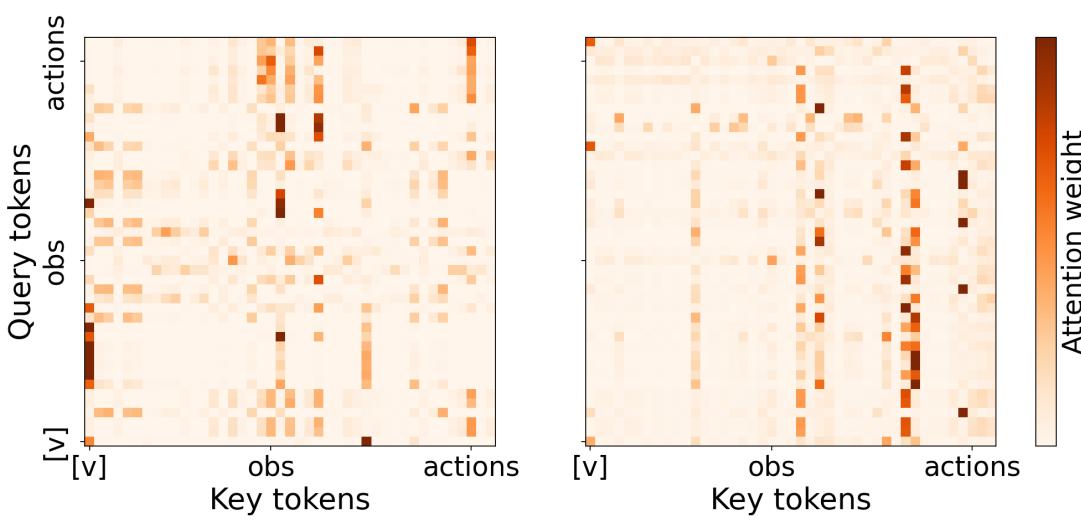
Step 300

Layer 0

Layer 1



QK Norm Attention Map



σ Reparam Attention Map

