

Step 0

Step 100

Step 200

Step 300

Layer 0

Layer 1

Layer 0

Layer 1

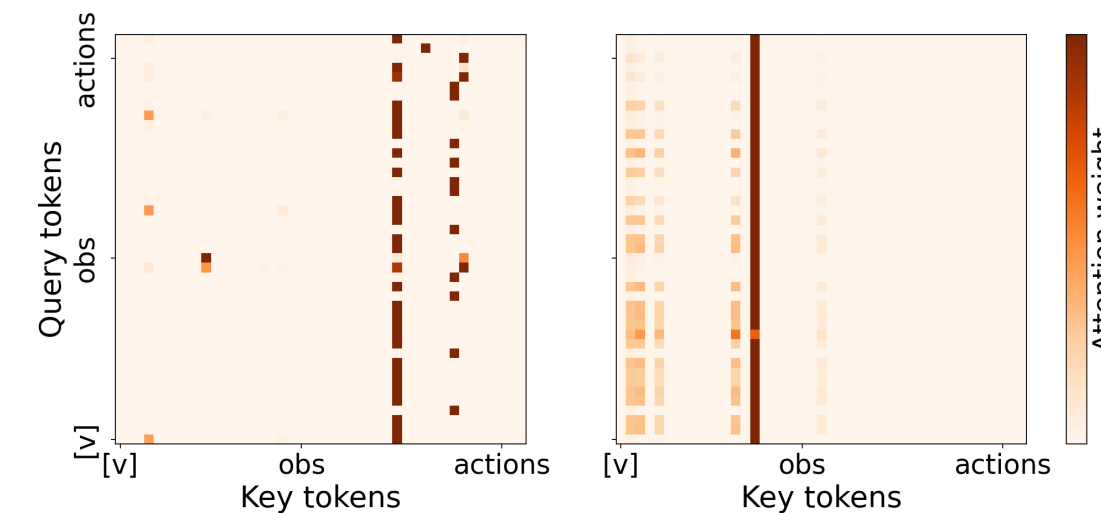
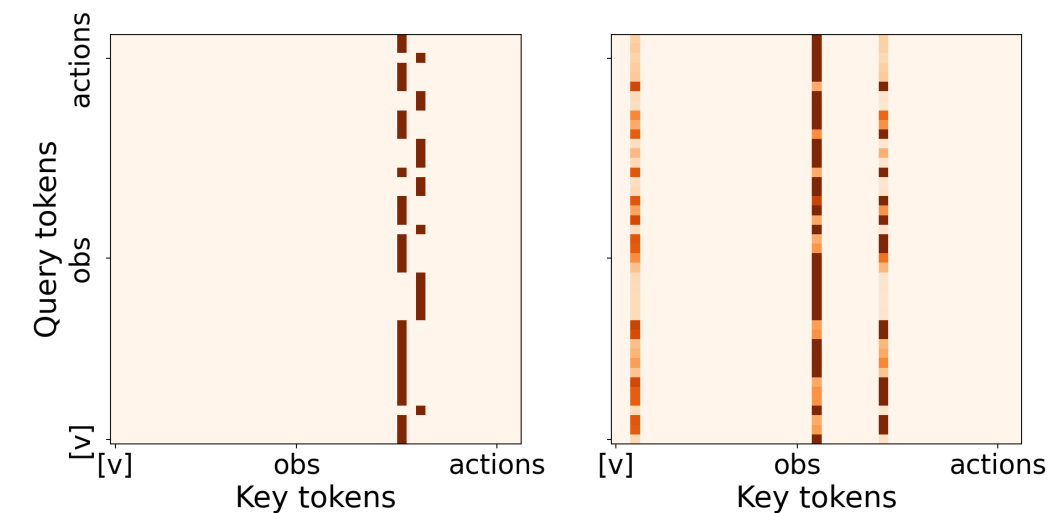
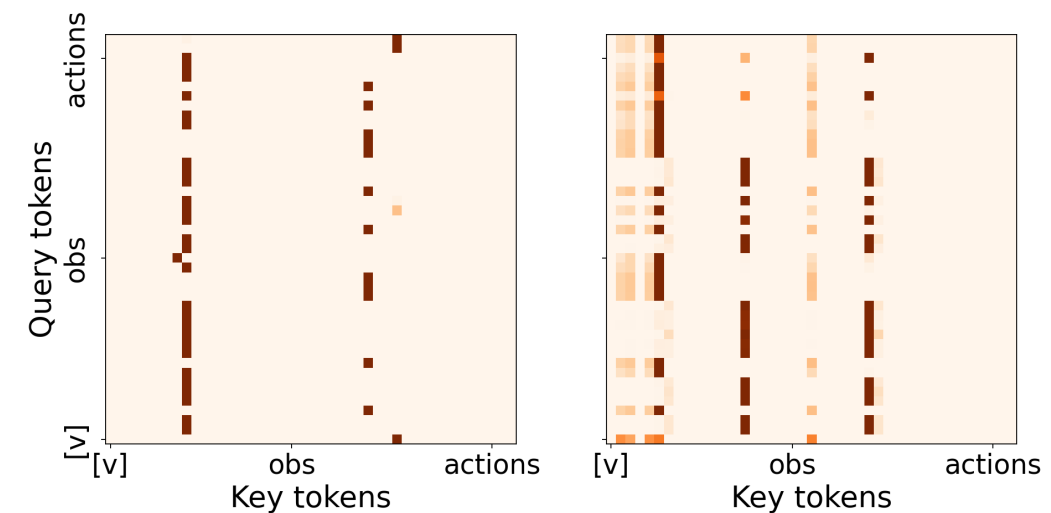
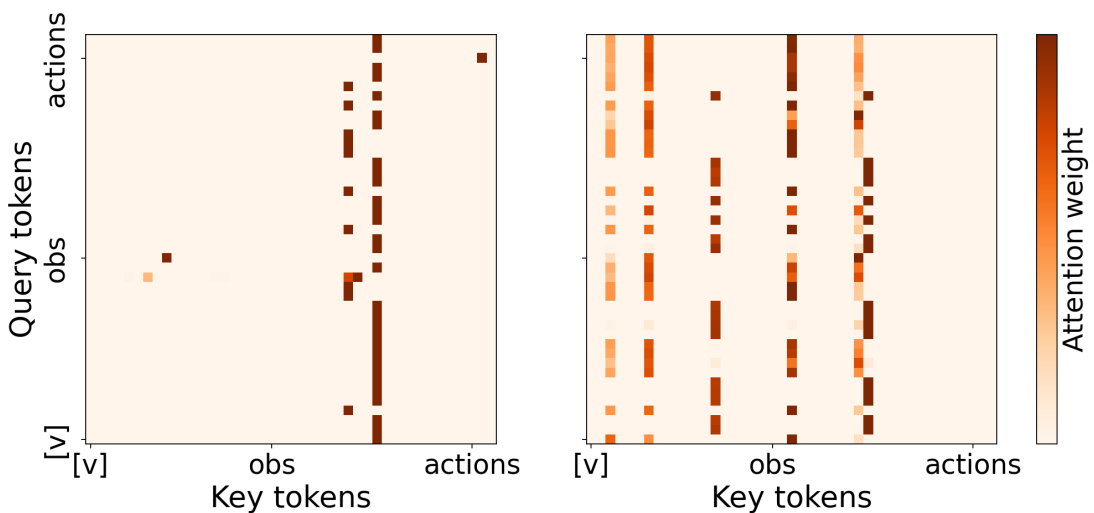
Layer 0

Layer 1

Layer 0

Layer 1

4-layer Transformer without TQL Attention Map



4-layer TQL Attention Map

