Attention Mechanism: Complete Walkthrough

Step 1: Q and K Matrices

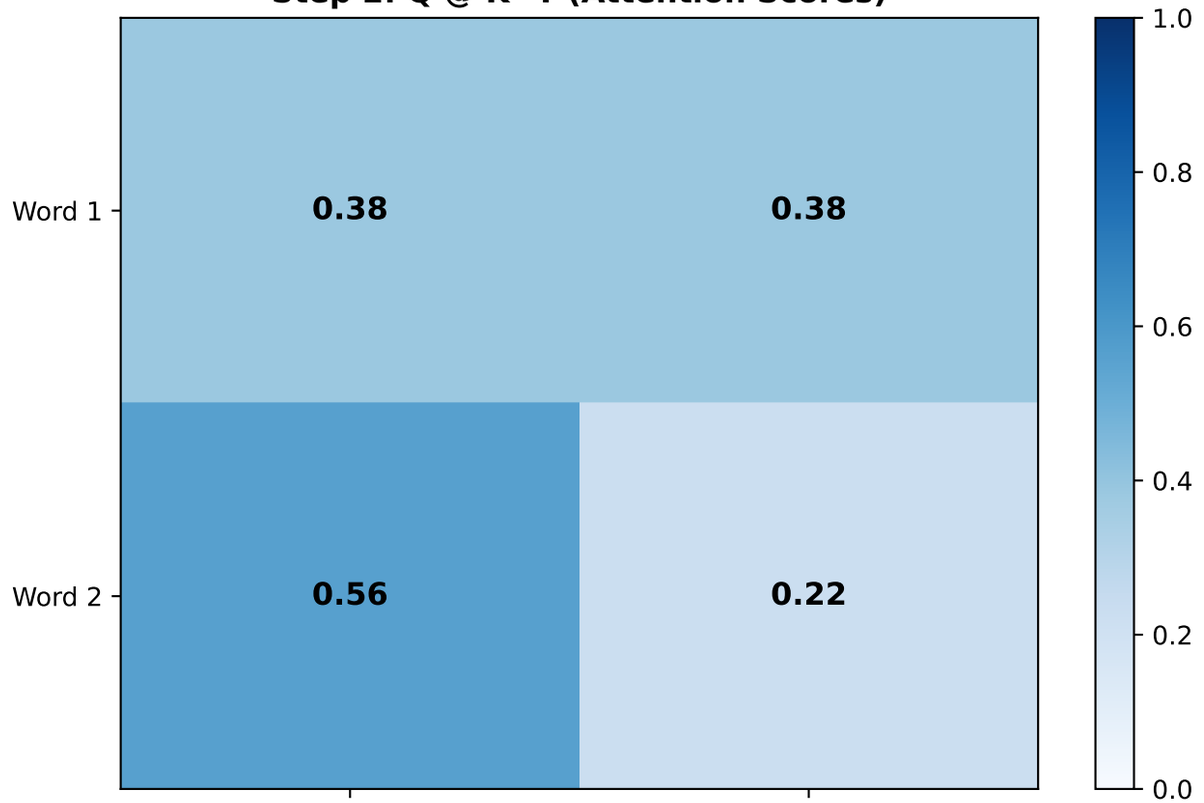
Query (Q)

```
Word 1: [0.5, 0.3]
Word 2: [0.2, 0.8]
```

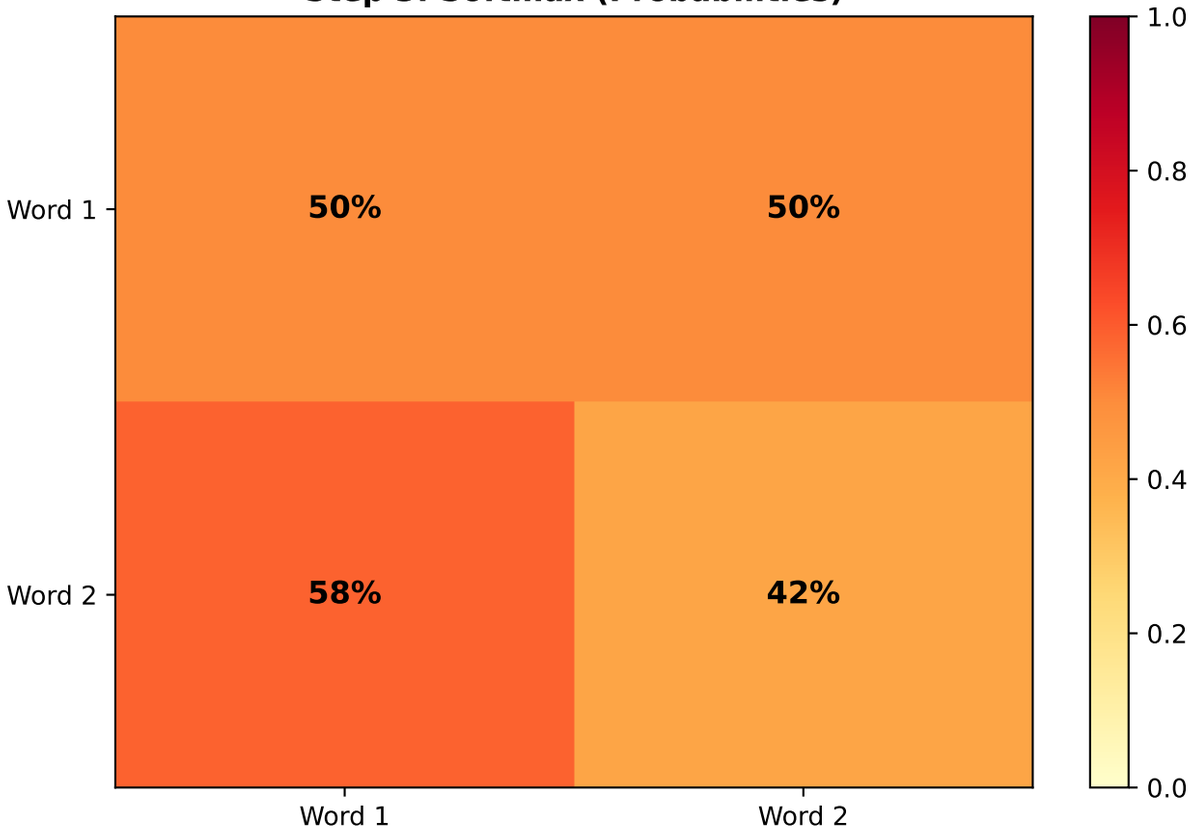
Key (K)

```
Word 1: [0.4, 0.6]
Word 2: [0.7, 0.1]
```

Step 2: Q @ K^T (Attention Scores)



Step 3: Softmax (Probabilities)



Step 4: Weighted Sum with V

Output = Attention @ V

```
Word 1 output:
[0.60, 0.45]

Word 2 output:
[0.65, 0.41]
```

This is the context-aware representation!