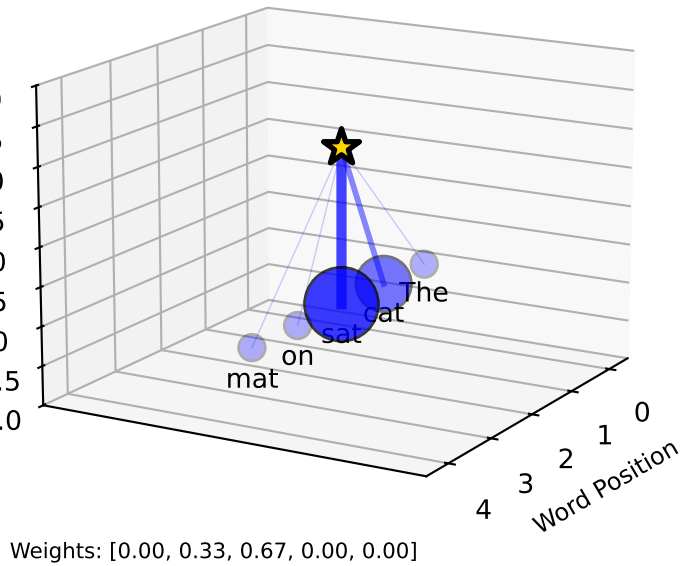
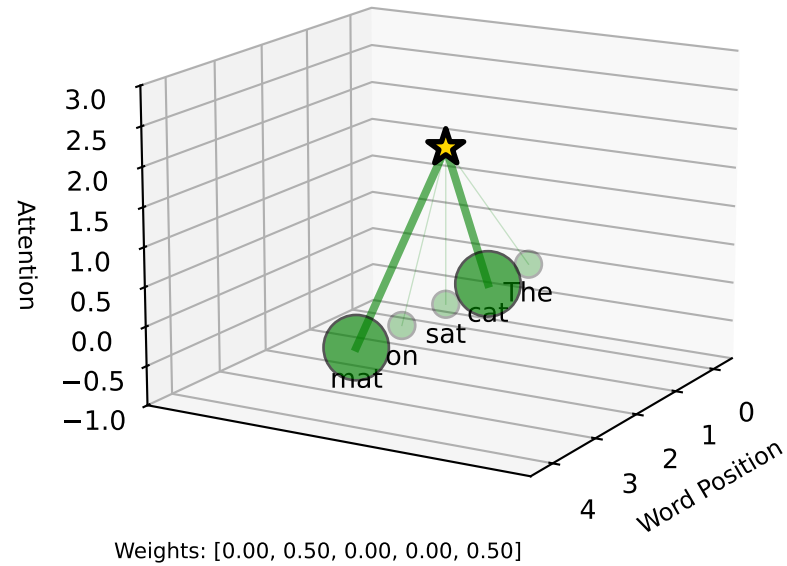


Multi-Head Attention: 4 Different Perspectives

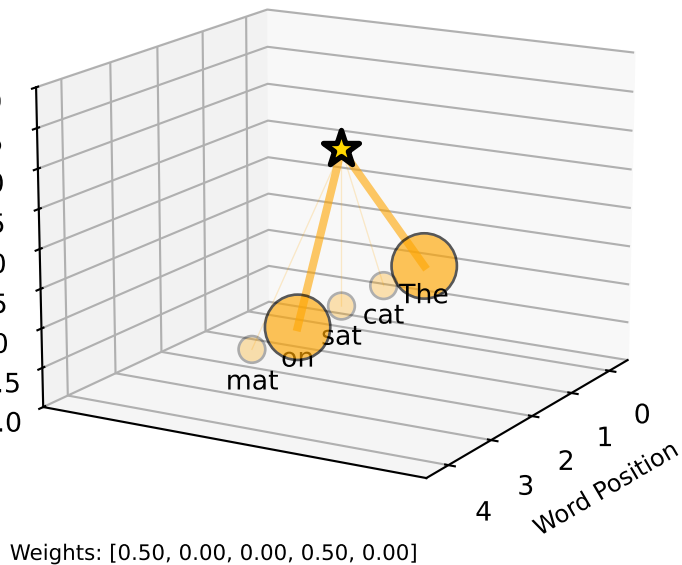
Head 1: Grammar Expert



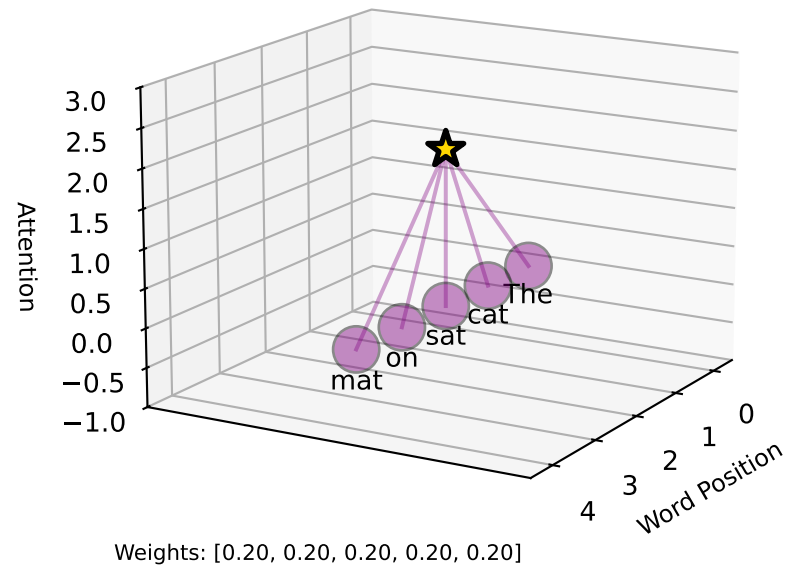
Head 2: Meaning Expert



Head 3: Position Expert



Head 4: Context Expert



Each head learns different attention patterns:

- Grammar: Subject-verb relationships
- Meaning: Semantic connections
- Position: Structural patterns
- Context: Global understanding