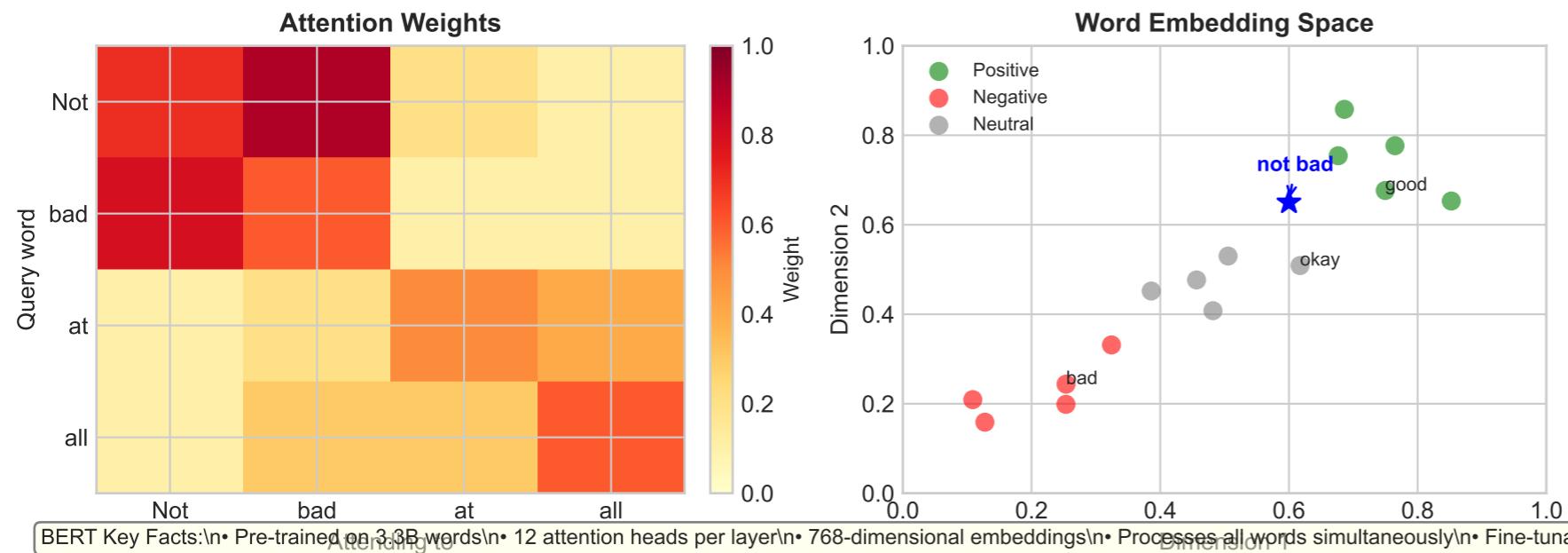
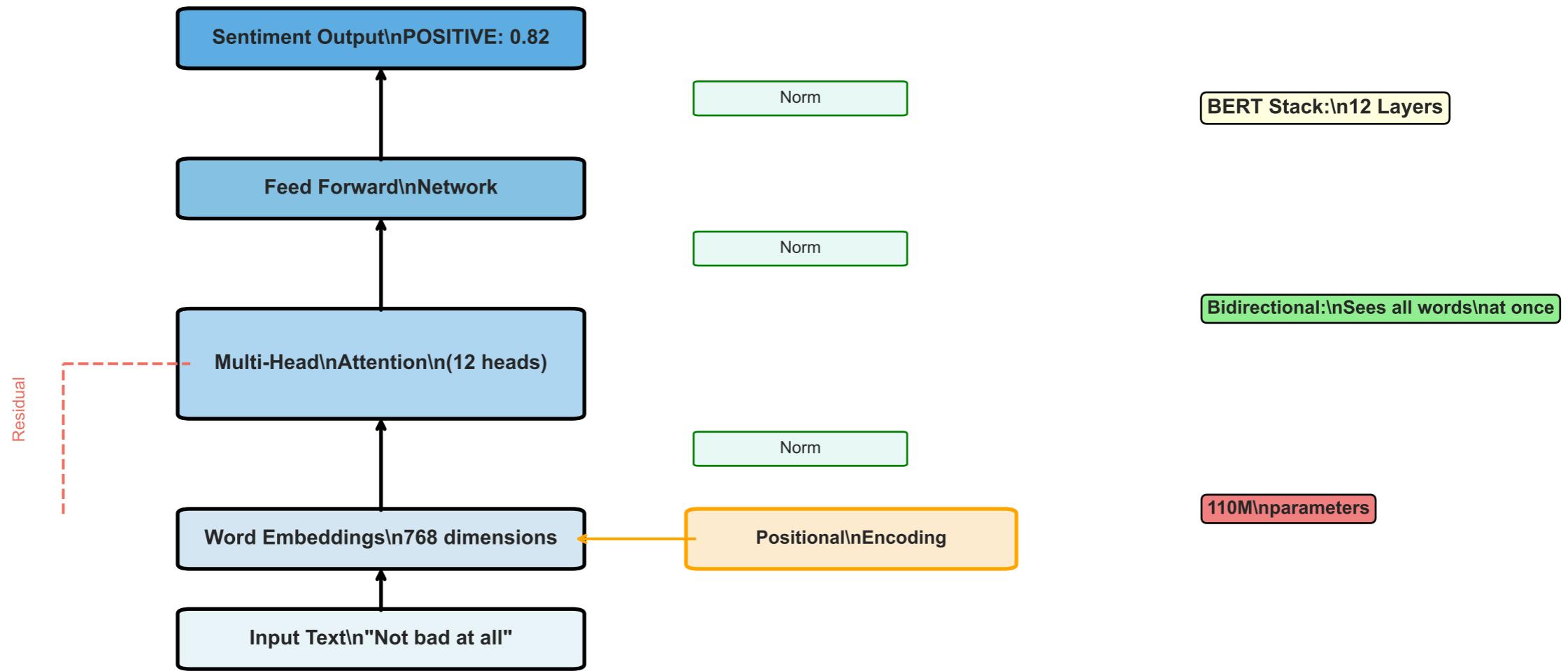


# How Transformers Process Text for Sentiment Analysis

## Transformer Architecture (Simplified)



**BERT Key Facts:**

- Pre-trained on 3.3B words
- 12 attention heads per layer
- 768-dimensional embeddings
- Processes all words simultaneously
- Fine-tunable for specific tasks