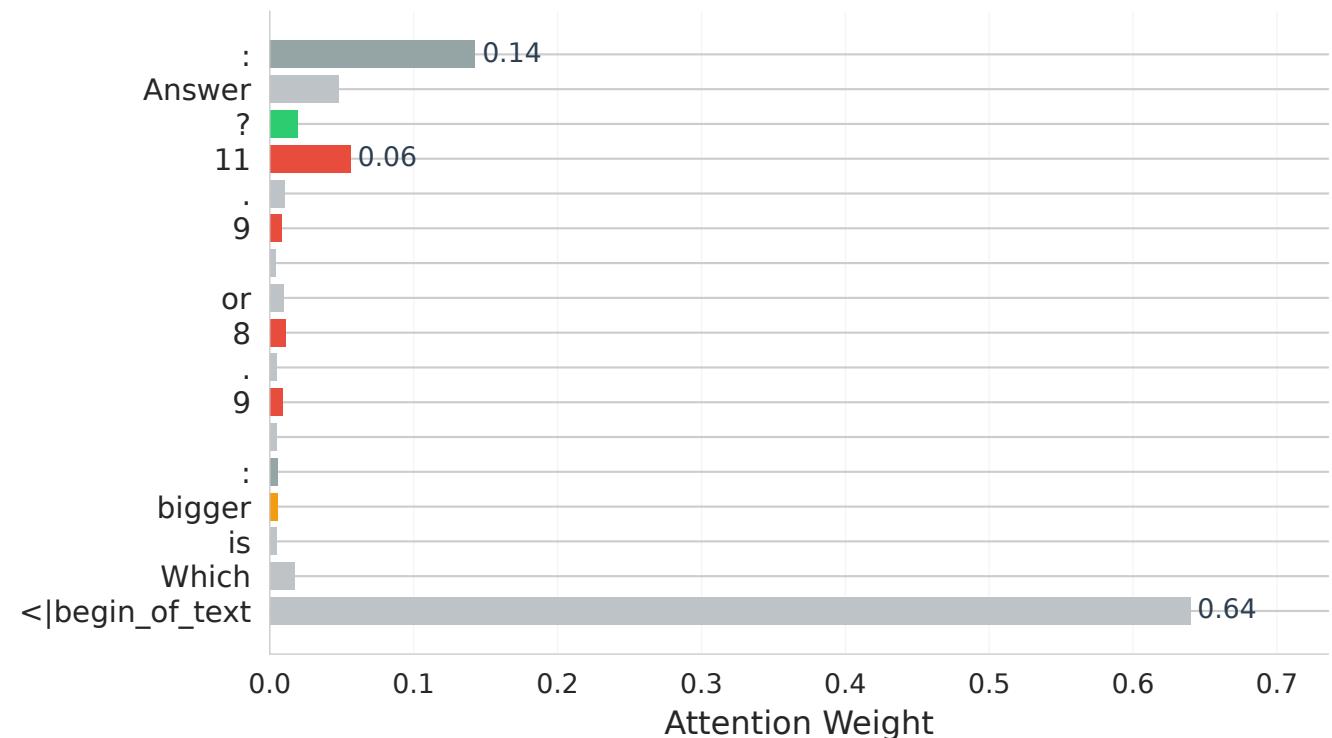
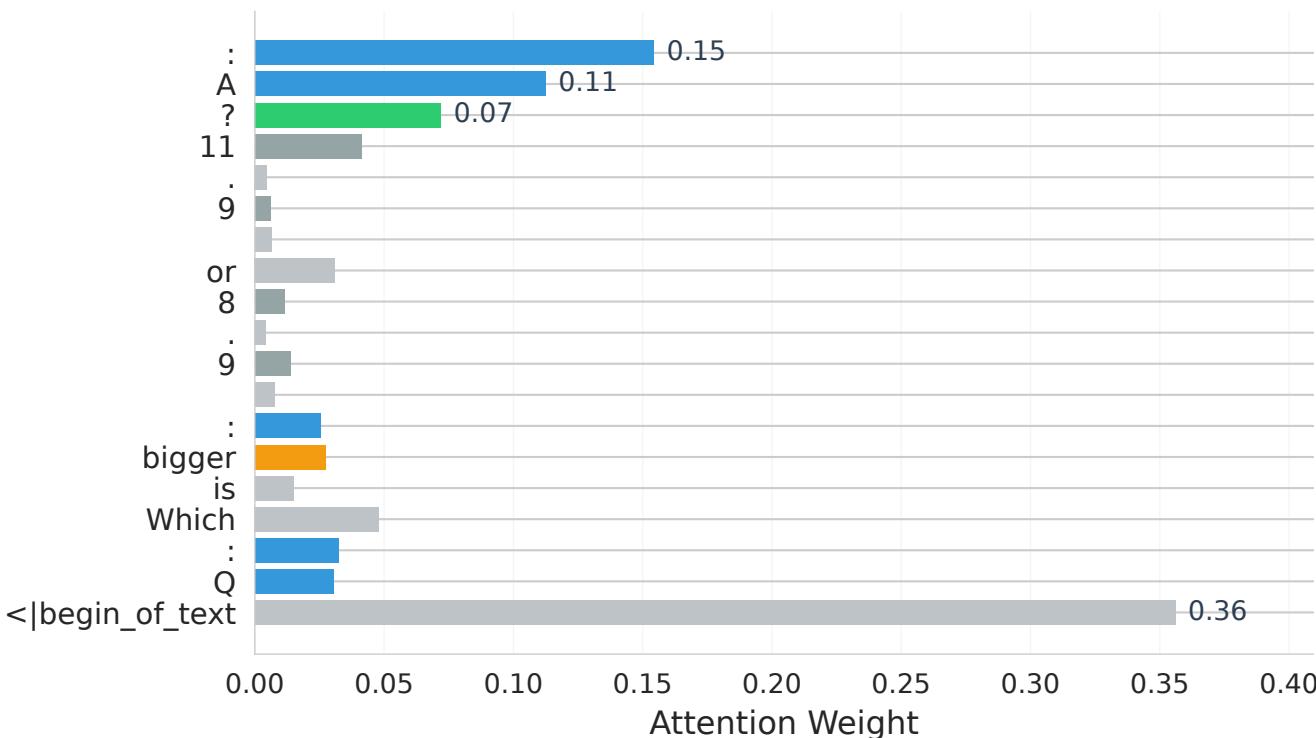


Attention Pattern Analysis: Layer 10, Head 17

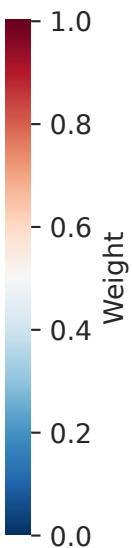
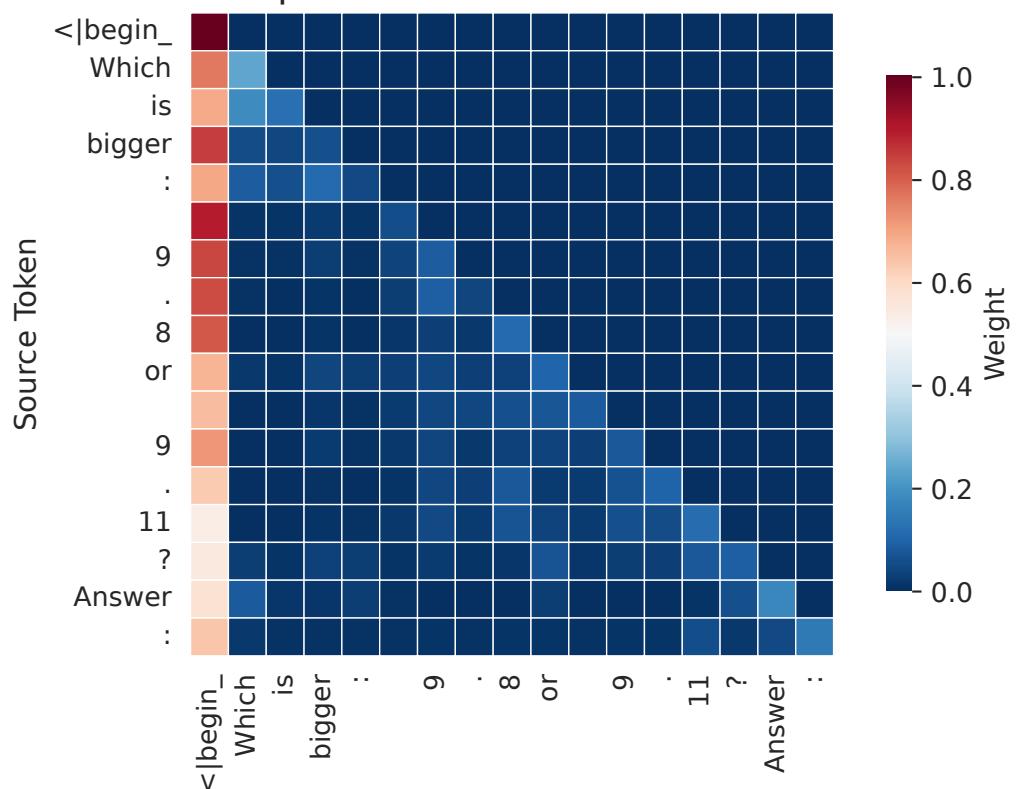
Simple Format (Correct Answer)



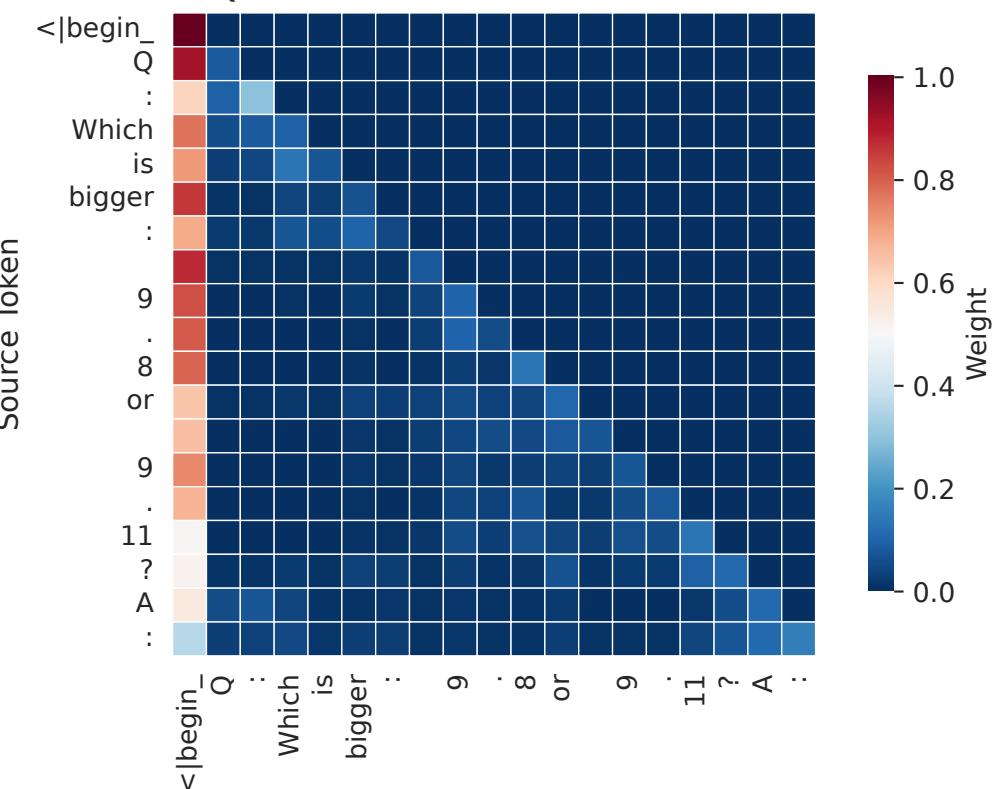
Q&A Format (Incorrect Answer)



Simple Format - Attention Matrix



Q&A Format - Attention Matrix



Last token attention distribution. Simple format: 8.4% on numbers. Q&A format: 35.5% on format tokens, 7.3% on numbers.

Number Tokens

Format Tokens

Query Token

Question Mark

Other Tokens