**“Attention is all you Need” - Response**

1. The authors are trying to suggest a new simple network architecture, the Transformer, that based on attention mechanisms only for a superior quality and time-saving performance with sequential input data such as natural language.
2. The authors wanted to write this paper because they wanted to explain their model, by breaking down the Transformer into particular elements and demonstrating their performance with translation tasks.
3. The authors are different from others as in they focus on attention mechanisms, or self-attention, instead of using recurrence or convolutions, which reduce the computations to a constant instead of proportional to the distances.