

Your 598 ID: 5107

Title of paper: Attention is All You Need by Polosukhin et al.

What is their primary result?

Why is this important?

What are their key ideas?

What are the limitations, either in performance or applicability?

What might be an interesting next step based on this work?

What's the architecture?

How did they train and evaluate it?

Did they implement something?

Grader's 598 ID: