

## Summary

The article “Attention is all you need” introduces the Transformer model, a novel architecture for sequence transduction tasks, which relies solely on attention mechanisms. It outperforms existing models in terms of translation quality, parallelizability, and training time efficiency. For instance, it achieves superior BLEU scores on English-to-German and English-to-French translation tasks, while requiring significantly less training time and resources. Additionally, it demonstrates strong generalization to other tasks like English constituency parsing.