Resources for understanding the paper “Attention is All You Need”

1. The paper itself

[[1706.03762] Attention Is All You Need](https://arxiv.org/abs/1706.03762)

1. A simplified explanation of the transformer architecture

[The Illustrated Transformer – Jay Alammar – Visualizing machine learning one concept at a time.](https://jalammar.github.io/illustrated-transformer/)

1. For the coding aspect

[The Annotated Transformer](https://nlp.seas.harvard.edu/2018/04/03/attention.html)