

Large Language Models

# Attention Mechanisms

Rodrygo Santos

Anisio Lacerda

# References

[Build a Large Language Model \(from Scratch\)](#), Ch. 3  
by Raschka (2024)

[Neural machine translation by jointly learning to align and translate](#)

Bahdanau et al., ICLR 2015

[Attention is all you need](#)

Vaswani et al., NIPS 2017