

Large Language Models

GPT Architecture

Rodrygo Santos

Anisio Lacerda

References

[Build a Large Language Model \(from Scratch\)](#), Ch. 4
by Raschka (2024)

[Let's build GPT: from scratch, in code, spelled out](#)
by Karpathy (2023)

References

[Improving language understanding by generative pre-training](#)

by Radford et al. (2018)

[Language models are unsupervised multitask learners](#)

by Radford et al. (2019)

[Language models are few-shot learners](#)

by Brown et al. (2020)

References

[Gaussian error linear units \(GELUs\)](#)

by Hendrycks and Gimpel (2016)

[Layer normalization](#)

by Ba et al. (2016)

[Identity mappings in deep residual networks](#)

by He et al. (2016)