

Large Language Models

Pretraining

Rodrygo Santos

Anisio Lacerda

References

[Build a Large Language Model \(from Scratch\)](#), Ch. 5

by Raschka (2024)

[Let's reproduce GPT-2 \(124M\)](#)

by Karpathy (2024)

[Decoupled weight decay regularization](#)

by Loshchilov and Frank Hutter (2017)