

# Treino T5 e Indexação

...

30 de março de 2023

# Conceitos Importantes

## Causal Language Modeling - ChatGPT

É uma técnica de processamento de linguagem natural em que um modelo de linguagem é treinado para prever a próxima palavra em uma sequência de palavras, dada a informação das palavras anteriores na sequência. Esse tipo de modelo é chamado de modelo de linguagem causal porque a ordem em que as palavras aparecem é levada em consideração na previsão da próxima palavra.

# Finetuning

- Modelo base = t5-base
- batch\_size = 32
- gradient\_accumulation\_steps = 8
- epochs = 15
- eval\_bleu = 20.3254

Step	Training Loss	Validation Loss	Bleu	Gen Len
50	5.115100	4.052029	1.245100	17.505000
100	3.027900	1.978187	8.370500	9.599000
150	2.007800	1.696618	14.699400	9.120000
200	1.811100	1.609012	16.123600	9.020000
250	1.685600	1.555242	16.980600	8.890000
300	1.594300	1.516577	19.023900	9.143000
350	1.496200	1.493860	19.064300	9.061000
400	1.413300	1.484577	18.957700	8.759000
450	1.328800	1.480291	19.946400	9.129000
500	1.254600	1.470809	20.325400	8.939000
550	1.179100	1.481003	20.307500	9.038000

# Problemas

- Tentativa de treinar em máquina com 8GB de RAM
  - Treino muito ruim com o resultados: bleu ~16 e NDCG ~ 0.60
- Muito demorado pra fazer expansão ~ 19h (batch 1)
  - Com batch maior demorava ainda mais

# Obrigado

Manoel Veríssimo

---