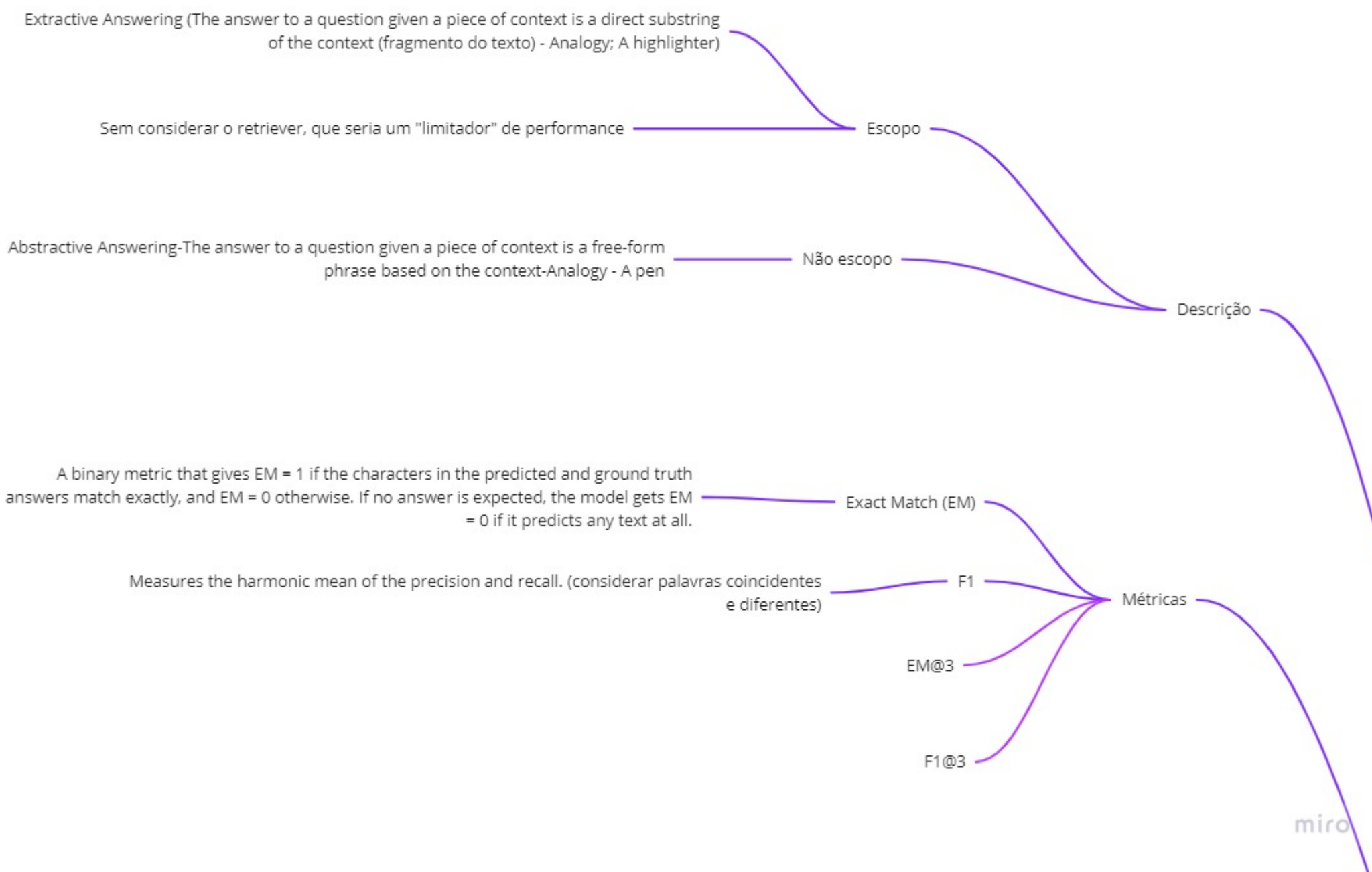
Extractive Q&A

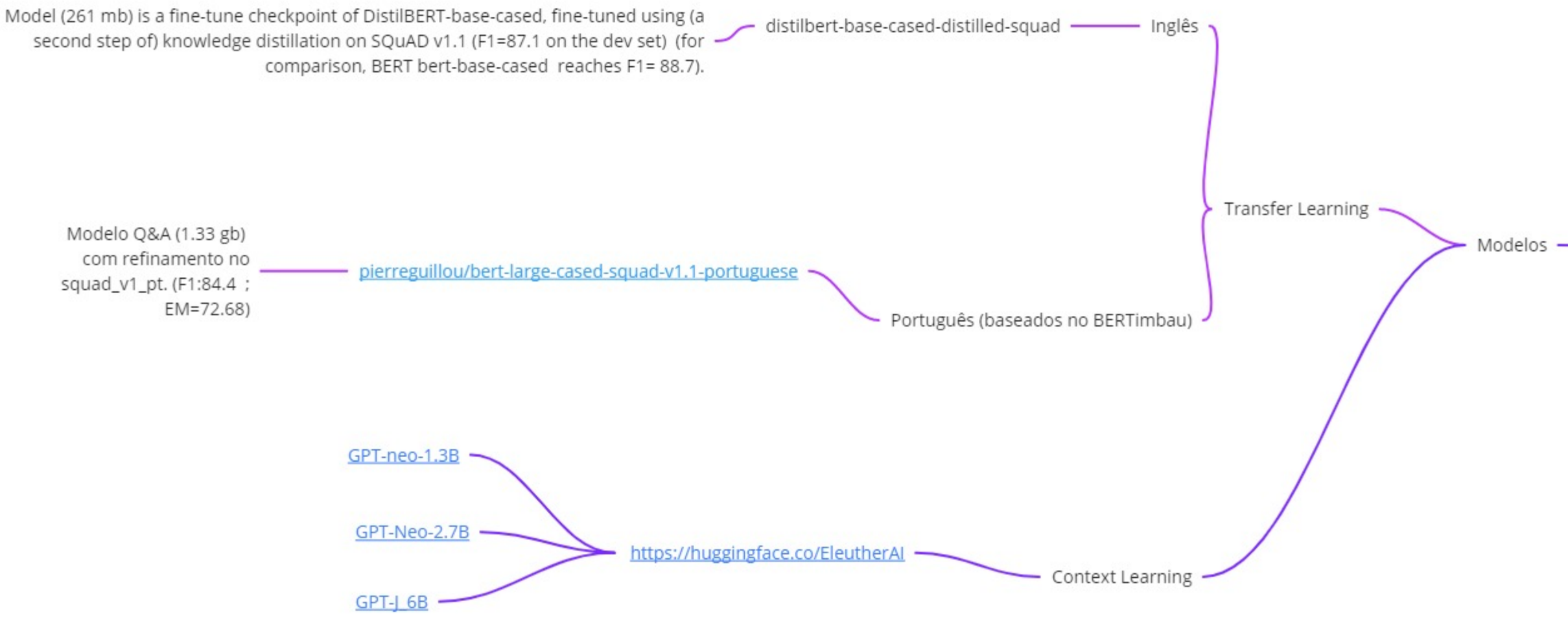
Performance Comparison between Learning Methods Context and Transfer Learning

Leonardo Augusto da Silva Pacheco
Marcus Vinícius Borela de Castro

Final Project at Discipline IA025 Introdução ao Aprendizado Profundo (2022/01)

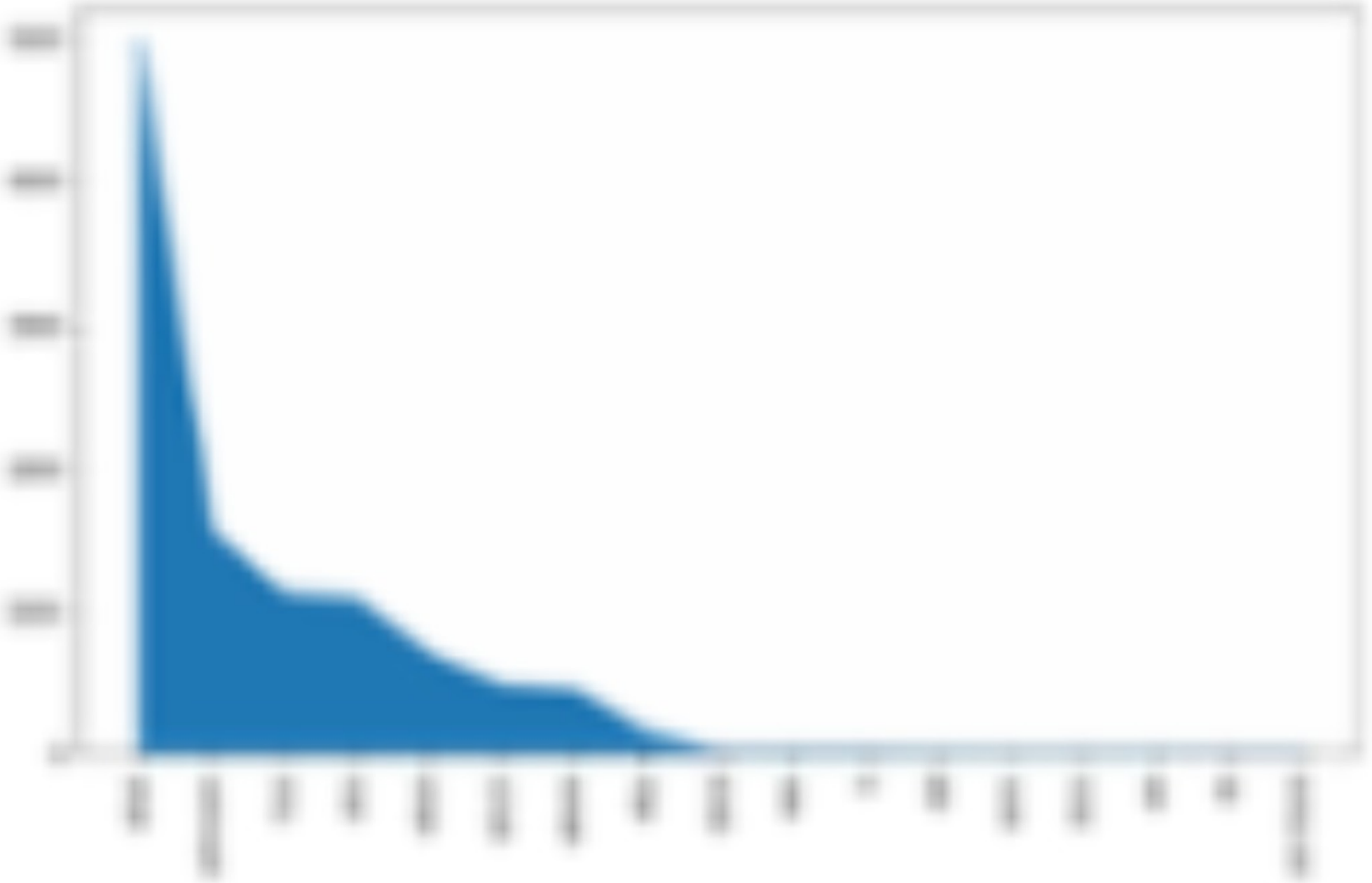
Link para github do projeto


Extractive Answering (The answer to a question given a piece of context is a direct substring of the context (fragmento do texto) - Analogy; A highlighter)
Sem considerar o retriever, que seria um "limitador" de performance
Escopo
Abstractive Answering-The answer to a question given a piece of context is a free-form phrase based on the context-Analogy - A pen
Não escopo
Descrição
A binary metric that gives EM = 1 if the characters in the predicted and ground truth answers match exactly, and EM = 0 otherwise. If no answer is expected, the model gets EM = 0 if it predicts any text at all.
Exact Match (EM)
Measures the harmonic mean of the precision and recall. (considerar palavras coincidentes e diferentes)
F1
EM@3
F1@3
Métricas


Modelos
Transfer Learning
Inglês
distilbert-base-cased-distilled-squad
Model (261 mb) is a fine-tune checkpoint of DistilBERT-base-cased, fine-tuned using (a second step of) knowledge distillation on SQuAD v1.1 (F1=87.1 on the dev set) (for comparison, BERT bert-base-cased reaches F1= 88.7).
Português (baseados no BERTimbau)
pierreguillou/bert-large-cased-squad-v1.1-portuguese
Modelo Q&A (1.33 gb) com refinamento no squad_v1_pt. (F1:84.4 ; EM=72.68)
Context Learning
https://huggingface.co/EleutherAI
GPT-neo-1.3B
GPT-Neo-2.7B
GPT-J 6B

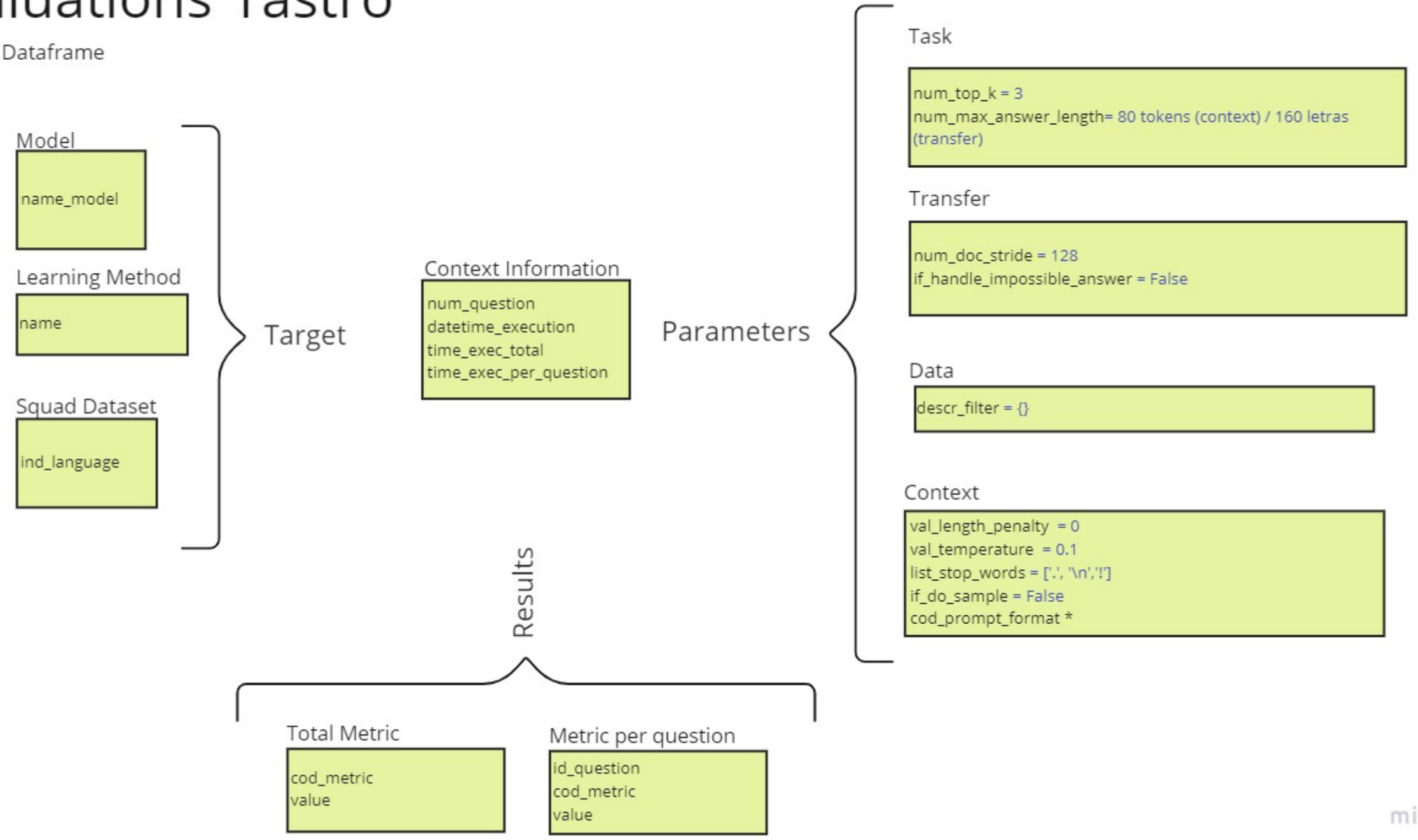

Exemplo
Inglês SQuAD 1.1
Português (SQuAD_v1_1_pt)
Totais artigos:48 paragrafos:2067 perguntas:10570 respostas:34726
Stanford Question Answering Dataset (base de teste)
Dataset
Dataset foi traduzido de forma automática e houve processo de revisão
Totais artigos:48 paragrafos:5644 perguntas:13075 respostas:34726
Erros perguntas: repetidas 2505 excluídas 2505
Responsável por sua criação foi avisado sobre ajuste dataset e revisão do desemenho publicado
Exclusão de perguntas repetidas dataset português

Distribuição por tipo de pergunta



Evaluations' rastro

Pandas Dataframe



Prompt format

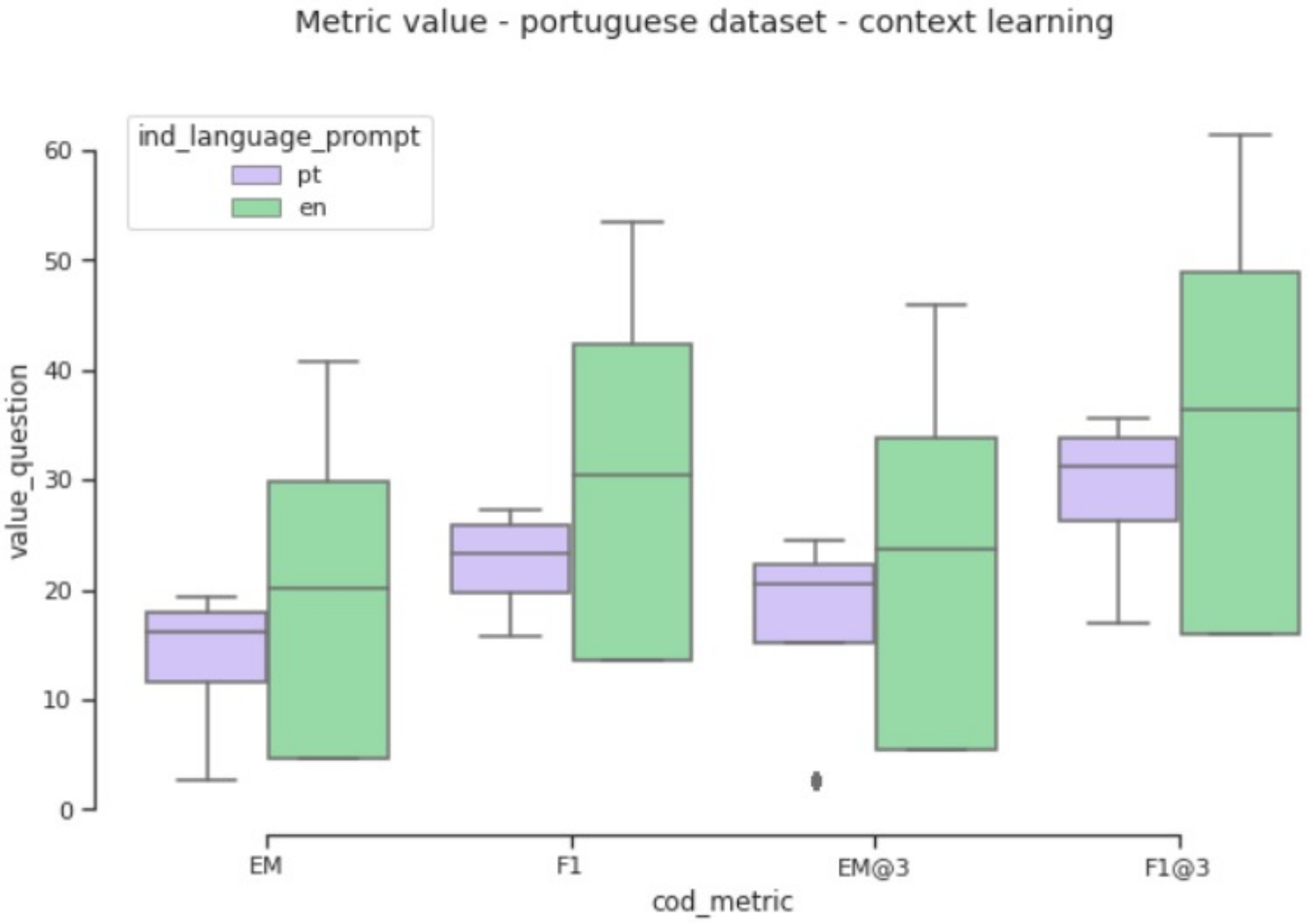
instrucao_pt = 'Instrução: Com base no texto abaixo, responda de forma sucinta à pergunta, evitando repetir palavras da pergunta:\n\n'
instrucao_en = 'Instruction: Based on the text below, answer the question succinctly, avoiding repeating words from the question:\n\n'

texto_questao_resposta_pt = 'Texto:{context}\n\nPergunta:{question}\nResposta:'
texto_questao_resposta_en = 'Text:{context}\n\nQuestion:{question}\nAnswer:'

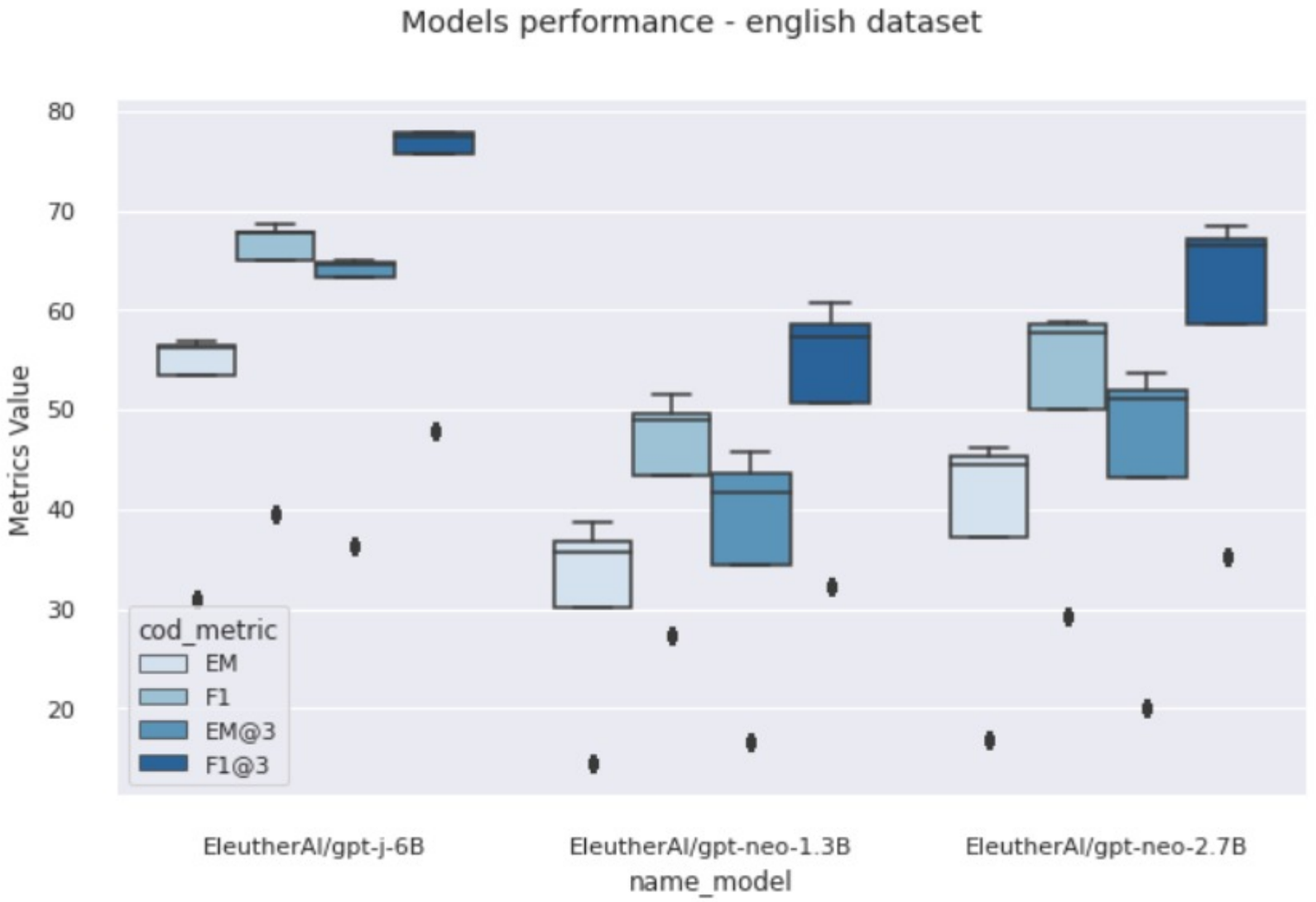
exemplo1_pt = 'Exemplo:\n\nTexto: Marcus nasceu em 1980 e trabalha no TCU desde 2005.\n\nPergunta: Quem nasceu em 1980?\nResposta: Marcus\n\n'
exemplo1_en = 'Example:\n\nText: Marcus was born in 1980 and has worked at TCU since 2005.\n\nQuestion: Who was born in 1980?\nAnswer: Marcus\n\n'
(...)

dict_prompt_format ={
1: {"prompt": instrucao_pt + texto_questao_resposta_pt, "num_shot":0, "ind_language":'pt'},
101: {"prompt": instrucao_en + texto_questao_resposta_en, "num_shot":0,"ind_language":'en'},
(...)
4: {"prompt": instrucao_pt + exemplo2_pt_tpp + texto_questao_resposta_pt, "num_shot":2, "format_example": "tpp","ind_language":'pt'},
104: {"prompt": instrucao_en + exemplo2_en_tpp + texto_questao_resposta_en, "num_shot":2, "format_example": "tpp","ind_language":'en'},
}

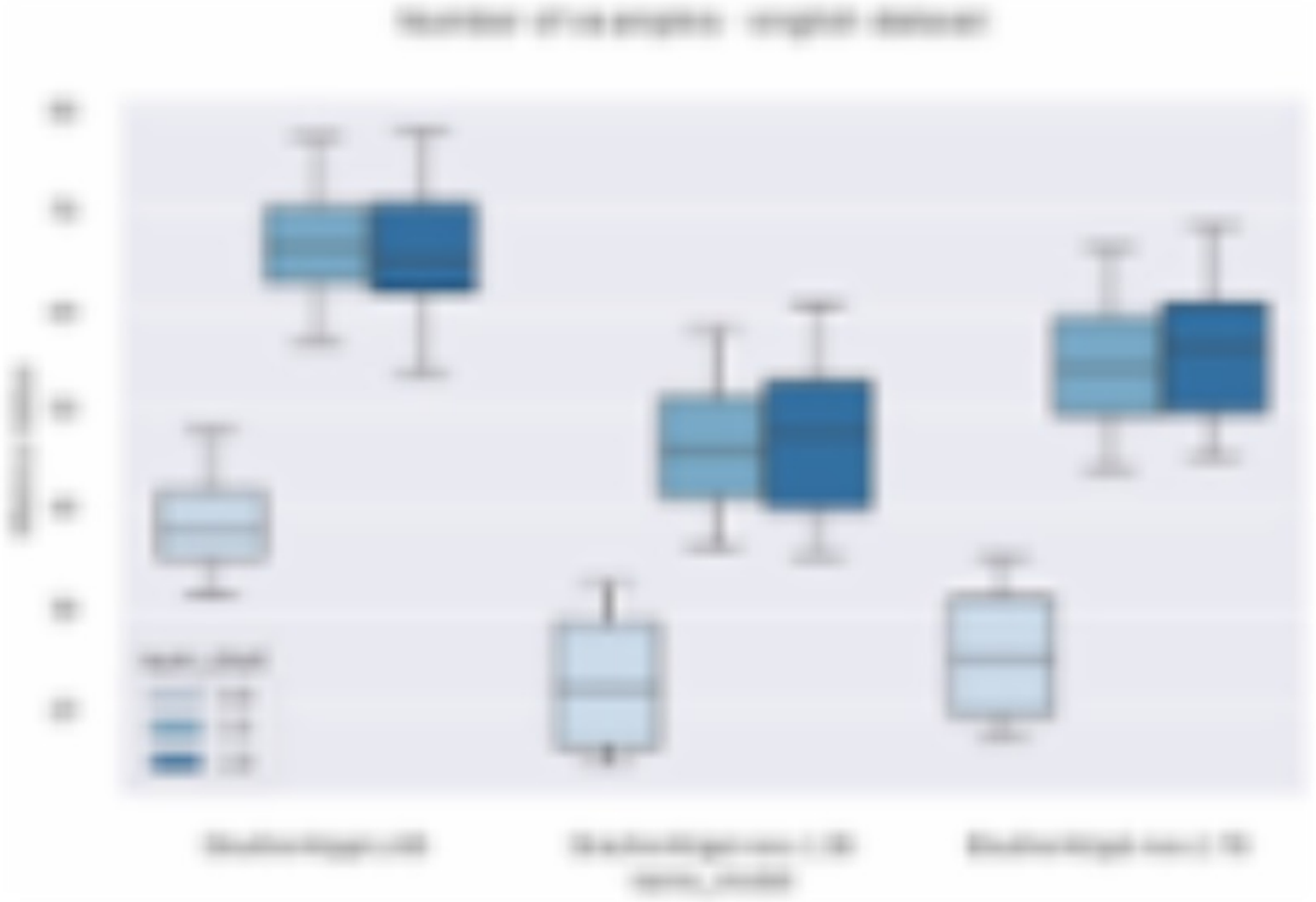
Alguns resultados alcançados - Context Learning



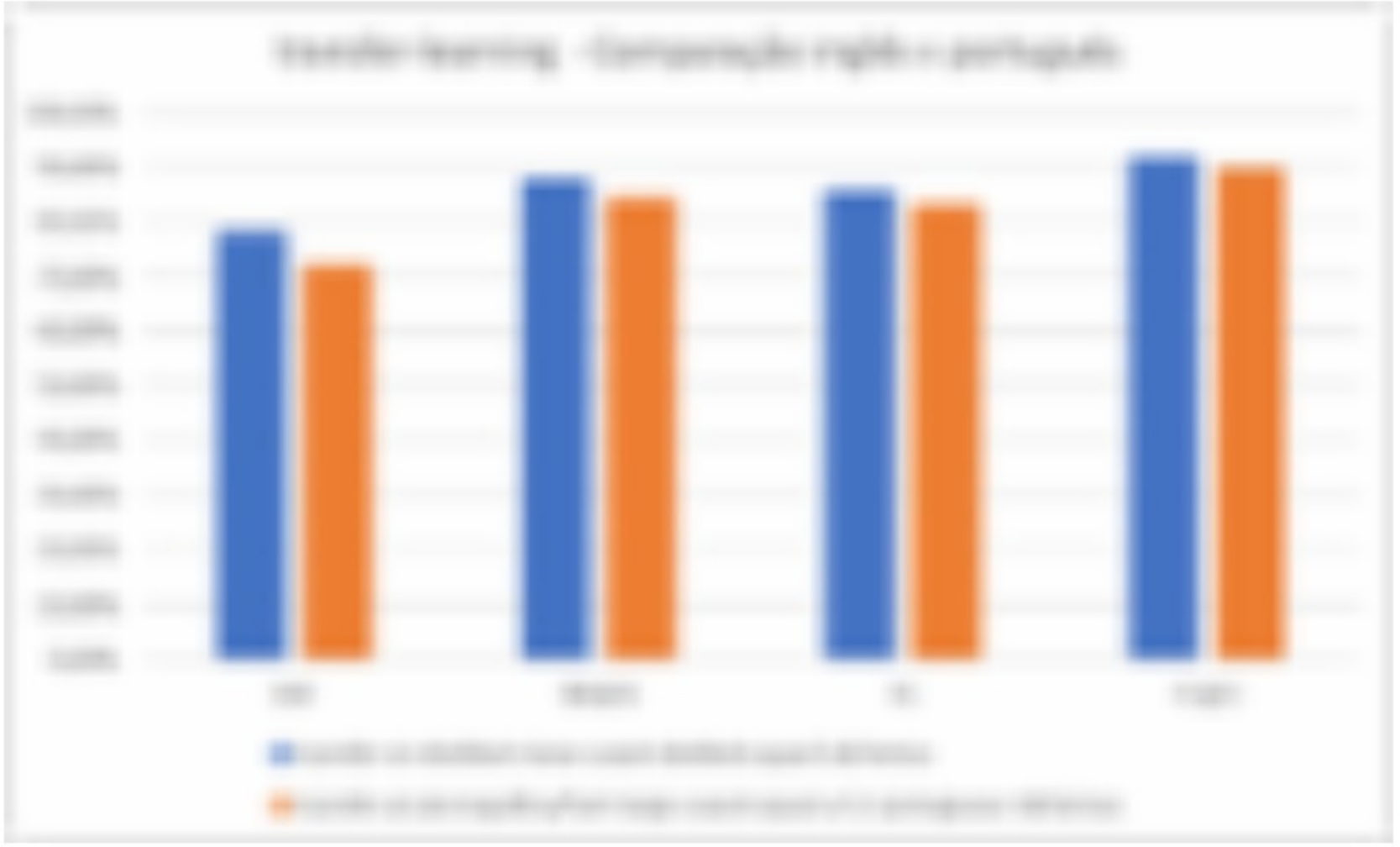
Context Learning



Context Learning



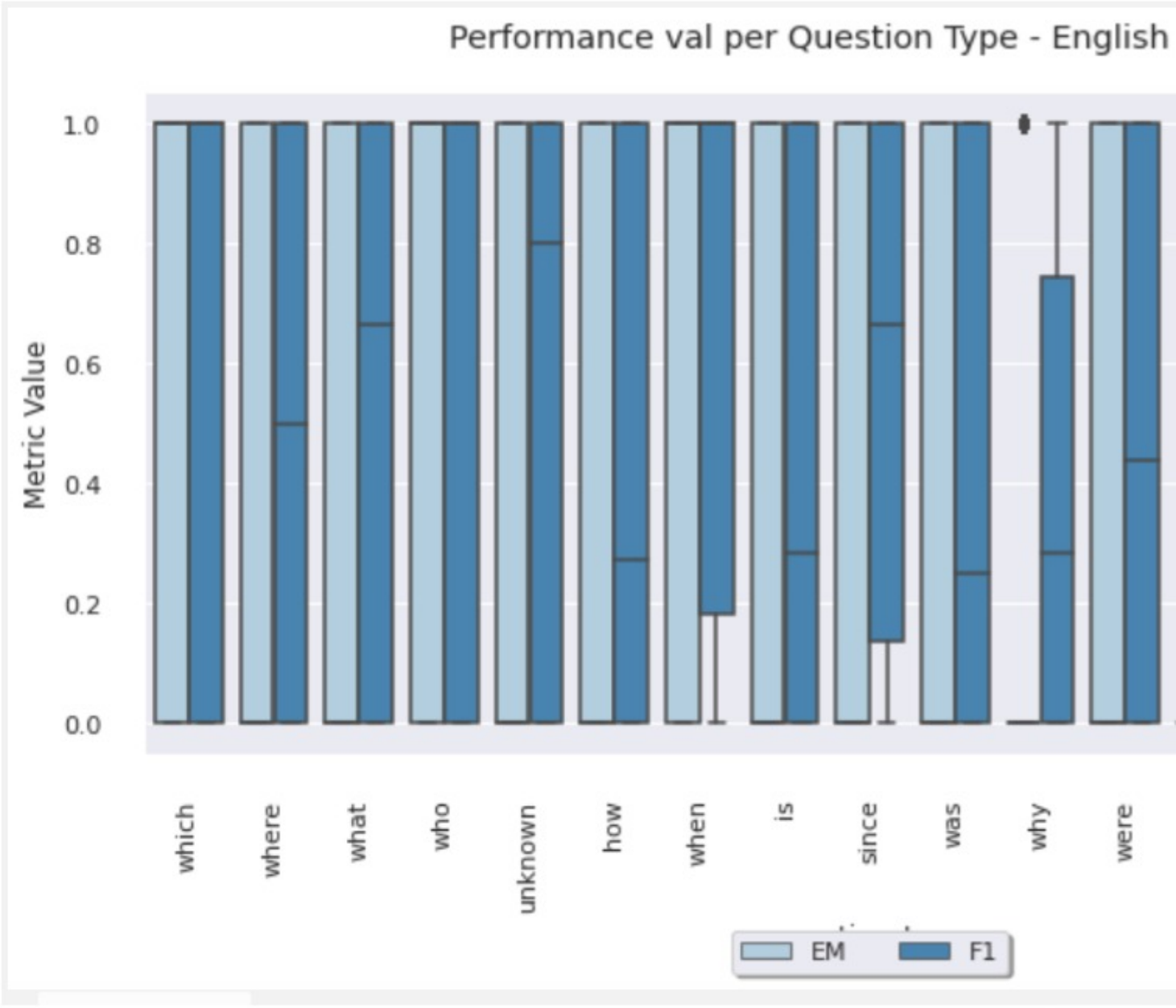
Transfer Learning



Comparação geral...

What?

How?



Comparação geral: tempo de execução

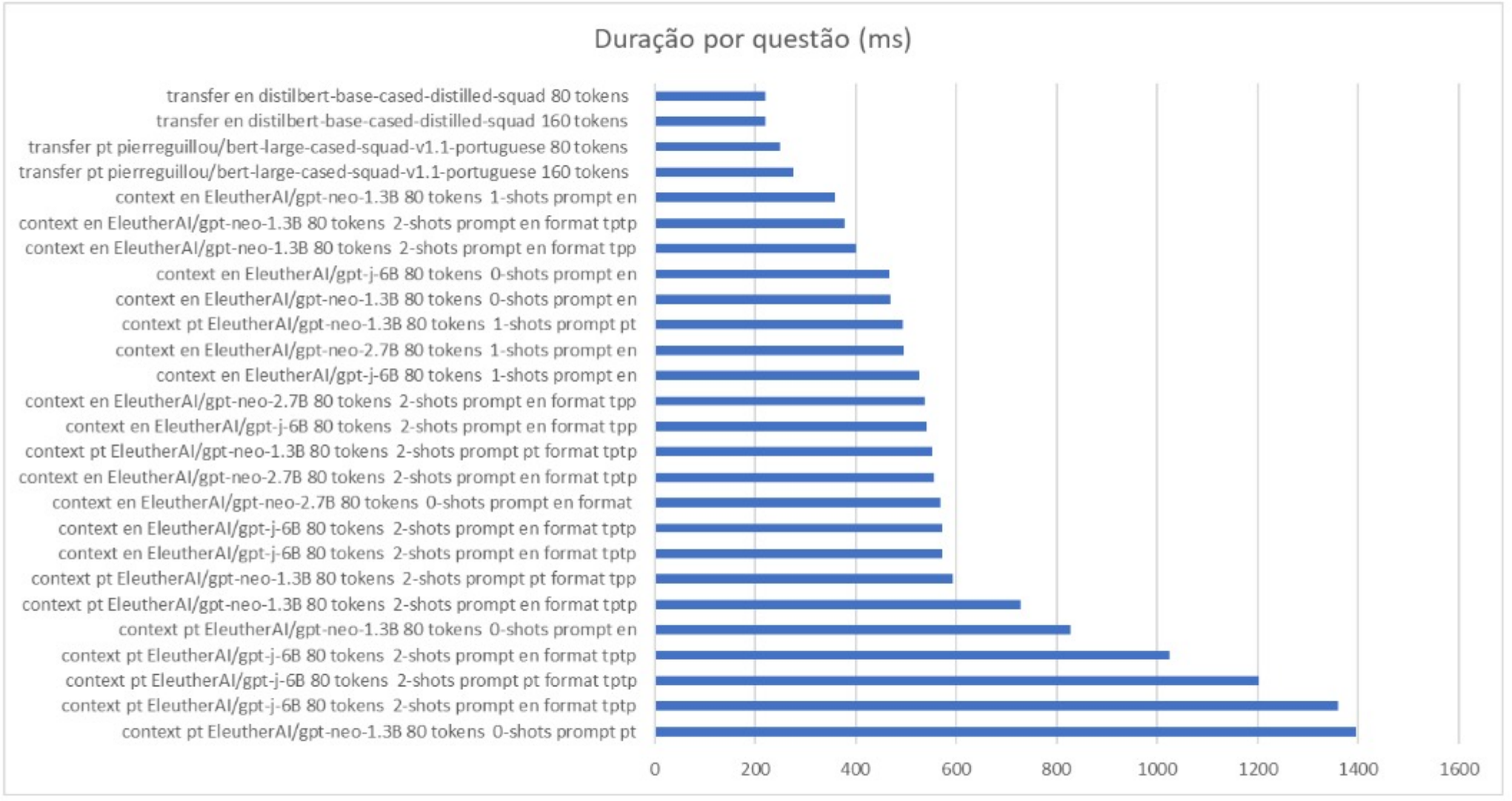
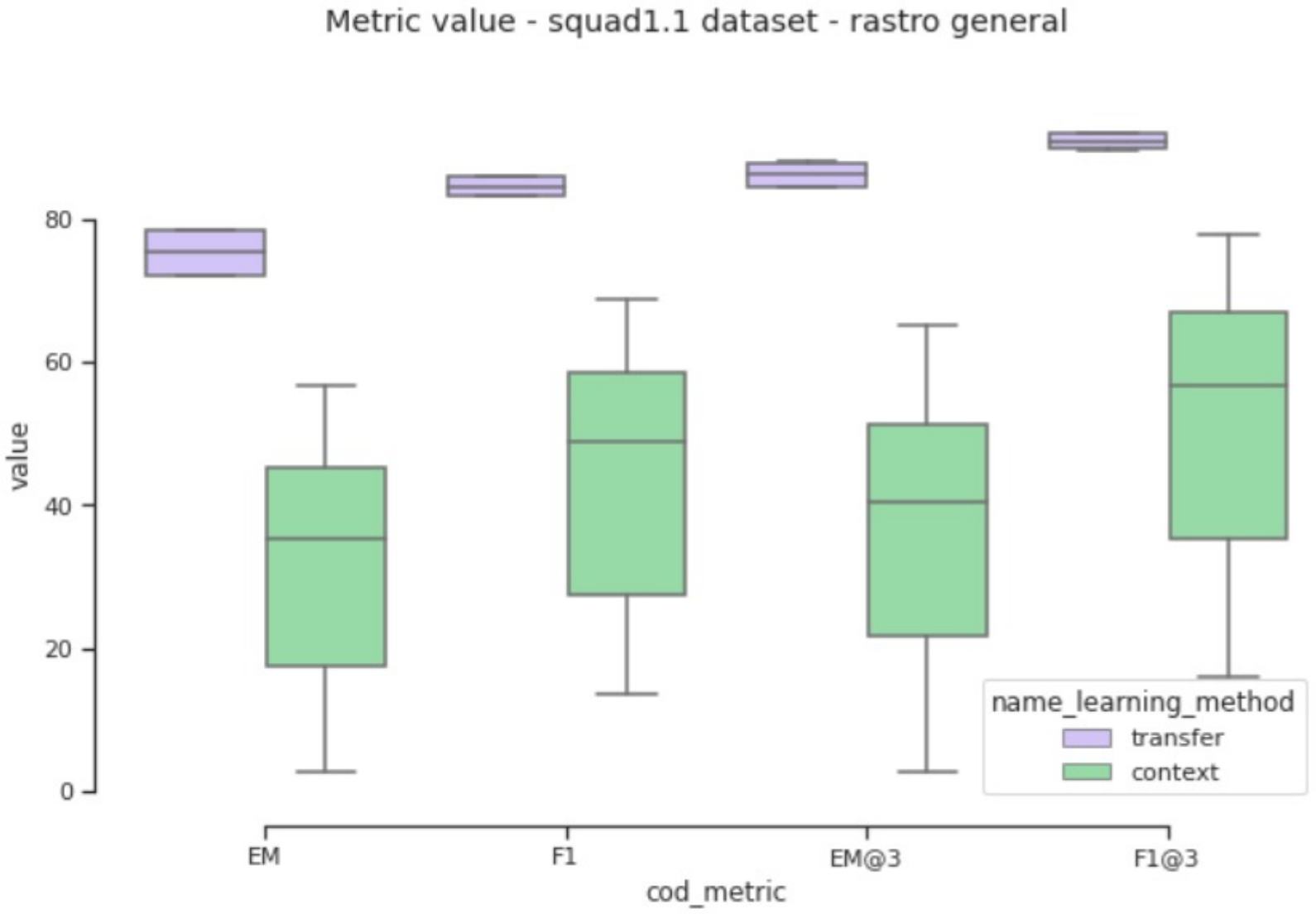


Figura 5 - Duração das execuções por questões, em milissegundos

Dá para comparar?



Trabalhos futuros

Repetir com GPT-3 (175B) e GPT-Neo-X (20B)

Avaliar dataset Squad 2.0 (possibilidade de não haver resposta)

Para finalizar...

Comparing Performance of Learning Methods, which is better: Context Learning and Transfer Learning?

Perguntando a quem entende...

Depende de condições de contexto
(inclusive de temperatura que
simula inspiração)
e de linguagem!