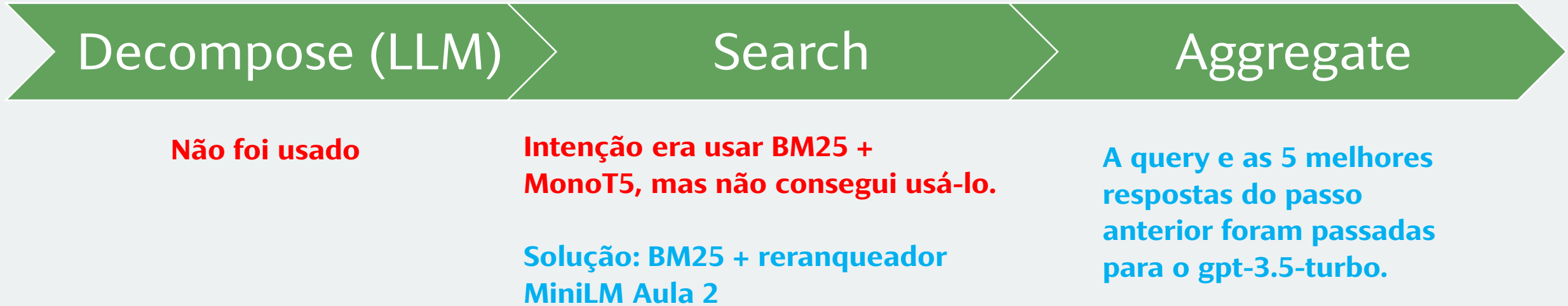


# Notebook – Visconde

Leandro Carísio

# Problemas e soluções no desenvolvimento



# Considerações feitas

- Usei as 100 primeiras perguntas do arquivo `iirc_test.json`.
- Usei apenas os artigos de `context_articles.json` que eram referenciados em `iirc_test.json` (~ 7000 artigos, ~500.000 janelas de sentenças).
- Few-shot: 2 exemplos fixos para cada chamada. Em cada exemplo, 4 documentos retornados.
- O resultado final foi processado para pegar a string depois de "Answer: ". Além disso, resultados como "unknown", "it is not possible...", "no information..." foram uniformizados para "none".

# Prompt usado com problema

Pedi para usar docs como evidência e dar apenas respostas diretas.

*You are a helpful reader that reads a list of documents and use them as evidence to answer a question. **You only give direct answers.***

Em vez de usar a role assistant, usei assitant (com typo). **Na prática era um zero-shot?**

F1: 0.33 / EM: 0.30 (81 questões)

F1: 0.21 / EM: 0.17 (removendo respostas none, 58 questões)

# Prompt final

*You are a helpful reader that reads a list of documents and use them as evidence to answer a question. You provide detailed explanations and concise answers.*

Nesse caso o gpt-3.5-turbo já forneceu a explicação e a resposta.

F1: 0.377 / EM: 0.31 (100 questões)

F1: 0.257 / EM: 0.159 (removendo respostas none, 69 questões)

# Caso curioso

Um dos exemplos no few-shot foi justamente o último caso de teste.

On what date did the Loma Prieta earthquake occur?

Evidence: According to [Document 1] and [Document 4], the Loma Prieta earthquake occur on **October 17, 1989**

Answer: **1989-10-18**

A evidência apresentada foi diferente da resposta e essa diferença tem a ver com o fuso horário usado nos documentos.

Quando se deparada com a mesma pergunta e os mesmos documentos, o modelo respondeu com “**october 17 1989**”, diferente da resposta passada no few-shot.

Não significa nada, mas achei curioso.

# Dúvida básica

Nesse exercício usamos um subconjunto de `context_articles`, considerando apenas os documentos apontados no `test_set`.

Não deveríamos ter usado todo o `context_article` em vez de menos de 15% dele?

# Obrigado

Leandro Carísio  
carisio@gmail.com