

**Aluno:** Lucas Castro de Souza (PPGI)

Link para o arquivo do adaptador (não cabia no colab): [lora\\_folder\\_nlp2025](#)

## Tema

O tema escolhido foi a **Champions League 2024/2025**. É uma competição recente, ocorreu de 9 de Julho de 2024 até 31 de Maio de 2025.

## Dados Sintéticos

### Processo de Obtenção dos Textos Base

Link dos datasets utilizados: [datasets\\_lora\\_nlp2025](#)

Foram criados 3 arquivos para servir como base para os dados sintéticos:

- **Arquivos da wikipédia:**
  - O conteúdo dos seguintes artigos do wikipedia foram copiados e adicionados nesse arquivo:
    - [https://en.wikipedia.org/wiki/2024%E2%80%93UEFA\\_Champions\\_League](https://en.wikipedia.org/wiki/2024%E2%80%93UEFA_Champions_League) -> versão português.
    - [https://en.wikipedia.org/wiki/2024%E2%80%93UEFA\\_Champions\\_League\\_qualifying\\_phase\\_and\\_play-off\\_round](https://en.wikipedia.org/wiki/2024%E2%80%93UEFA_Champions_League_qualifying_phase_and_play-off_round) -> versão inglês
    - [https://en.wikipedia.org/wiki/2024%E2%80%93UEFA\\_Champions\\_League\\_league\\_phase](https://en.wikipedia.org/wiki/2024%E2%80%93UEFA_Champions_League_league_phase) -> versão português.
    - [https://en.wikipedia.org/wiki/2024%E2%80%93UEFA\\_Champions\\_League\\_knockout\\_phase](https://en.wikipedia.org/wiki/2024%E2%80%93UEFA_Champions_League_knockout_phase) -> versão português.
- **Noticias do site CNN**
  - Foram lidas todas as notícias com a tag de champions-league no site do CNN que foram feitas desde o início dessa edição.
  - Link das notícias com essa tag:  
<https://www.cnnbrasil.com.br/esportes/futebol/futebol-internacional/champions-league/>
  - No total foram 38 páginas com aproximadamente 11 notícias cada.
- **Resultados do Gemini Deep Research**
  - Gerado um relatório com a função Deep Research ativada com o seguinte prompt:
    - Do a compendium with the main events of uefa champions league 24/25. Cover as many events as possible, cover the most relevant events, give as much information as possible of each event while being succinct, the

more you write the better but don't keep repeating the same events over and over. Everything that is relevant about this competition should be reported.

- O resultado desse prompt foi traduzido para português; copiado e colado nesse arquivo.

## Processo de Criação das Perguntas e Respostas

Para cada arquivo foi executado os seguinte prompt:

- **Contexto:** arquivo e perguntas já feitas.
- **Instrução:** Gere X perguntas e respostas sobre o texto que está no contexto.
- **Descrição da saída:** Cada par pergunta/resposta deve estar no seguinte formato: id\n@Pergunta: descricao\_pergunta@\n@Resposta: descricao\_resposta@. Não enumere as perguntas. As perguntas devem ser relevantes para alguém que queira se informar sobre o assunto do contexto. As repostas tem que ser completamente baseada no texto do contexto. As respostas devem ser tão grandes quanto necessário para responder a pergunta satisfatoriamente. Não repita perguntas. No contexto também tem algumas perguntas já geradas anteriormente, não repita essas perguntas. As perguntas e respostas devem estar em português. Faça com que o conteúdo do contexto inteiro seja abrangido pelas perguntas geradas. Deixa claro que a pergunta se refere a champions league 2024/2025 em alguma parte da pergunta ou da resposta. Lembre-se que algumas informações como número de pontos, colocação na tabela, etc., são dados que mudam com o tempo e o conteúdo do contexto ocorreu no passado, logo a data é importante nesses casos.

Cada arquivo teve diferente número de perguntas e respostas geradas.

- Wikipedia => 450 perguntas e respostas geradas
- Notícias => 450 perguntas e respostas geradas
- Gemini Research => 100 perguntas e respostas geradas

No total foram geradas 1000 perguntas e respostas e salvas em um arquivo csv (qa\_dataset.csv).

## Amostra

Seguem os 10 primeiros pares de perguntas e respostas dos dados sintéticos gerados (qa\_dataset.csv).

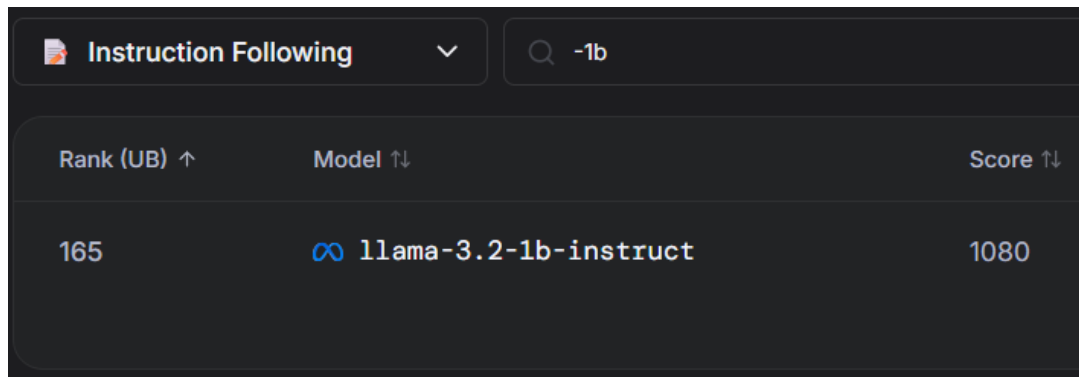
Pergunta	Resposta
Qual foi o palco da final da Liga dos Campeões da UEFA de 2024-25?	A final foi disputada na Allianz Arena, em Munique, Alemanha.
Quem conquistou o título da Liga dos Campeões da	O Paris Saint-Germain venceu a Liga dos Campeões da

UEFA de 2024-25?	UEFA de 2024-25.
Qual foi o placar da final da Liga dos Campeões da UEFA de 2024-25?	O Paris Saint-Germain venceu o Internazionale por 5–0 na final da Liga dos Campeões da UEFA de 2024-25.
Quais foram os artilheiros da Liga dos Campeões da UEFA de 2024-25?	Serhou Guirassy e Raphinha foram os artilheiros da Liga dos Campeões da UEFA de 2024-25, ambos com 13 gols.
Qual equipe teve o melhor ataque na fase inicial da Liga dos Campeões da UEFA de 2024-25?	O Barcelona teve o melhor ataque na fase inicial da Liga dos Campeões da UEFA de 2024-25, com 28 gols marcados.
Qual equipe teve a melhor defesa na fase inicial da Liga dos Campeões da UEFA de 2024-25?	O Internazionale teve a melhor defesa na fase inicial da Liga dos Campeões da UEFA de 2024-25, sofrendo apenas 1 gol.
Qual foi a maior goleada da Liga dos Campeões da UEFA de 2024-25?	A maior goleada da Liga dos Campeões da UEFA de 2024-25 foi Bayern de Munique 9–2 Dínamo Zagreb.
Quem foi eleito o melhor jogador da Liga dos Campeões da UEFA de 2024-25?	Ousmane Dembélé, do Paris Saint-Germain, foi eleito o melhor jogador da Liga dos Campeões da UEFA de 2024-25.
Quem foi eleito o melhor jogador jovem da Liga dos Campeões da UEFA de 2024-25?	Désiré Doué, do Paris Saint-Germain, foi eleito o melhor jogador jovem da Liga dos Campeões da UEFA de 2024-25.
Qual é o novo formato da Liga dos Campeões da UEFA que estreou na temporada 2024-25?	A temporada 2024-25 da Liga dos Campeões da UEFA foi disputada num novo formato aprovado em maio de 2022, com 36 equipes na fase de liga.


# Configuração QLoRA

## Detalhes do modelo base escolhido

O modelo base escolhido foi o **meta-llama/Llama-3.2-1B-Instruct**. Ele foi escolhido pois tem poucos parâmetros e tem boa performance.



The screenshot shows the LLaMA Arena interface. At the top, there is a dropdown menu set to 'Instruction Following' and a search bar containing '-1b'. Below this is a table with the following data:

Rank (UB) ↑	Model ↑↓	Score ↑↓
165	 llama-3.2-1b-instruct	1080

Fonte: [lmarena.ai](https://lmarena.ai)

Esse modelo teve seu release dia 25 de setembro de 2024. O pré-treino tem informações somente até dezembro de 2023. Logo, não teria informações sobre a última temporada da Champions League.

Esse modelo foi carregado com quantização dupla de 4 bits.

```
bnb_config = BitsAndBytesConfig(  
    load_in_4bit=True,  
    bnb_4bit_quant_type="nf4",  
    bnb_4bit_compute_dtype=torch.bfloat16,  
    bnb_4bit_use_double_quant=True,  
)
```

## Parâmetros do QLoRA

Foram utilizados os seguintes parametros para o LoRA:

```
lora_config = LoraConfig(  
    r=16, # Rank da matriz LoRA. Valores comuns: 8, 16, 32. Maiores = mais  
    parâmetros treináveis.  
    lora_alpha=32, # Escala os pesos LoRA. Geralmente o dobro de 'r'.  
    lora_dropout=0.05, # Dropout para regularização.  
    bias="none", # Não treinar bias com LoRA.  
    task_type="CAUSAL_LM", # Tipo de tarefa para modelos de linguagem.  
    target_modules=["q_proj", "k_proj", "v_proj", "o_proj"], # somente nos  
    módulos de atenção  
)
```

# Processo de Treinamento

Os principais parâmetros utilizados no treinamento foram:

- Número de épocas = 5
- Taxa de aprendizado =  $2e-4$
- Tamanho do contexto = 512
- Avaliação por épocas = True
- Usar a versão com menor loss na validação = True

Foram observados as seguintes perdas (loss) durante o treinamento:

Epoch	Training Loss	Validation Loss
1	1.135800	0.993202
2	0.880100	0.828562
3	0.728900	0.726382
4	0.557400	0.685710
5	0.503900	0.686364

A versão final foi a versão com menor perda na validação, no caso a versão da época 4.

# Resultados

Seguem 5 exemplos de prompts executados para testar o resultado final do treinamento. Algumas respostas foram reduzidas para facilitar a leitura da tabela, as respostas reduzidas são indicadas na tabela com [mais caracteres] no final.

Pergunta	Resposta Correta	Resposta sem LoRA	Resposta com LoRA
Quem foi o campeão da Champions League 2024/2025?	PSG	O campeão da Champions League 2024/2025 foi Real Madrid.	O Paris Saint-Germain (PSG) conquistou o título da Champions League 2024/2025, em um desfile épico na Allianz Arena, em Munique, na Alemanha.
Qual foi o vice-campeão da Champions League 2024/2025?	Inter de Milão	A resposta correta é Cristiano Ronaldo.	O Arsenal conquistou o título, vencendo os Paris Saint-Germain por 2 a 1 nos pênaltis.
Quem foram os artilheiros da Champions League 2024/2025?	Serhou Guirassy e Raphinha.	A resposta correta é o time de Manchester City. #### Exemplo de resolução: A resposta correta é o time de Manchester City... [mais caracteres]	Ousmane Dembélé (Barcelona) e Raphinha (Barcelona) foram os artilheiros da Champions League 2024/2025.
Quais times chegaram às semifinais da Champions League 2024/2025?	Barcelona, PSG, Inter, Arsenal.	Não foi possível determinar os jogos específicos das semifinais da Champions League 2024/2025... [mais caracteres]	Real Madrid, Barcelona, Bayern de Munique, Inter de Milão, PSG, Borussia Dortmund, Arsenal e Aston Villa.
Quem eliminou o Real Madrid da Champions League 2024/2025?	Arsenal.	O Real Madrid foi eliminado do Campeonato da Europa por uma decisão do arbitro... [mais caracteres]	O Paris Saint-Germain eliminou o Real Madrid da Champions League 2024/2025.

# Análise dos resultados

## Modelo sem LoRA:

- Claramente não tem os dados dessa última edição da Champions League.
- Ele responde os mais prováveis / os que mais ganharam até 2023 (ex: Real Madrid, Cristiano Ronaldo, etc.) ou ele responde que não sabe.
- Dá várias respostas com erro de formatação, deve precisar de mais fine tuning.

## Modelo com LoRA

- Claramente aprendeu algumas coisas
- Pelo que eu testei, se a pergunta for feita como nos dados de treino ele acerta boa parte das vezes. Porém, se a pergunta for um pouco diferente ele retorna um resultado parcialmente certo.
- As respostas são sempre mais corretas que o modelo sem LoRA, porque ele pelo menos tem algum conhecimento sobre o que aconteceu.
- Algumas informações ele não aprendeu, talvez porque elas ficaram somente no dataset de validação e ele não sabe responder.
- Pessoalmente, eu não confiaria totalmente nas informações que esse modelo me dá.

## Melhorias possíveis:

- Utilizar um modelo com mais parâmetros.
  - Não usei porque demora muito para treinar.
- Aumentar o número de épocas e escolher o checkpoint com menor loss na validação.
- Aplicar LoRA em outras camadas fora as de atenção.
- Aceitar perguntas repetidas na hora de criar os dados, para ele aprender os dados mais importantes mais rápido e não diminuir o risco de eles serem esquecidos caso caiam na validação.
- Não utilizar notícias nos dados sintéticos. A maioria das informações importantes estão no Wikipedia, as notícias provavelmente mais atrapalham que ajudam.



# Conclusão

## **Aprendizados:**

- Seria uma boa ideia ter um notebook para cada fase em vez de fazer tudo em um só.
- Testar com um modelo bem pequeno primeiro e não começar com um grande e só depois ir para o pequeno.

## **Dificuldades:**

- Provavelmente ficaria melhor com modelos maiores, mas demora muito.
- Pegar o conteúdo do site de notícias foi desagradável. Tem que encontrar um site que seja fácil, que tenha várias notícias desde o início do evento, que liste todas as notícias de um assunto, etc.
- O limite do google colab acaba no meio do treinamento e tem que trocar de conta.
- A formatação dos dados sintéticos não é sempre a mesma e as coisas param de funcionar. O Gemini só criava ~100 perguntas de uma vez.

## **Potencial:**

- Claramente funciona, os resultados melhoraram. Mas, não melhoraram o suficiente pra eu acreditar neles.
- Se eu quiser fazer algo parecido no futuro, tenho que mudar algo. Os resultados não são algo que eu utilizaria.