



Tech Challenge 3

IA para Devs

Daniela Cruz de Malta - RM353365

Gabriela Maciel Godoi - RM355125

Lucas Sutelo - RM353721

Sumário

Caso de Estudo.....	3
Metodologia.....	3
1. Preparação do Dataset.....	3
2. Fine-tuning com a OpenAI.....	3
3. Instalação de Biblioteca Necessárias para LLAMA.....	4
2. Configuração do Ambiente para LLAMA.....	4
4. Chamada do Foundation Model LLAMA.....	4
5. Processo de Fine-Tuning do LLAMA.....	4
Resultados.....	5
1. OpenAI.....	5
2. LLAMA.....	5
Conclusão.....	6

Caso de Estudo

O objetivo deste Tech Challenge é realizar o fine-tuning de um foundation model utilizando a base de dados "The AmazonTitles-1.3MM" para receber perguntas dentro do contexto da base no Prompt de usuário e a resposta conter os aprendizados do fine-tuning.

Este documento detalha o processo de seleção e preparação da base de dados e descreve o processo de fine-tuning do modelo, com os parâmetros e ajustes utilizados.

O código fonte do processo de fine-tuning está no repositório do GitHub: <https://github.com/Danizinh/Fine-Tuning>.

Metodologia

A equipe decidiu realizar o Fine-Tuning utilizando dois métodos, usando a biblioteca da OpenAI e importando o LLM LLAMA. A seguir está o procedimento realizado para os dois casos.

1. Preparação do Dataset

O dataset utilizado para o treinamento é o **AmazonTitles-1.3MM**, carregado a partir do arquivo **trn.json**, contendo títulos e descrições de produtos.

As colunas "title" e "content" foram preparadas para formar os *prompts* de entrada no modelo em um arquivo com extensão JSONL, que é necessária para realizar o Fine-Tuning tanto no LLAMA, como na OpenAI.

Foi necessário realizar uma quebra no dataset nos dois casos devido a limitação de recursos.

2. Fine-tuning com a OpenAI

A equipe gerou uma API key para utilizar a biblioteca OpenAI. O treinamento foi feito com uma base de dados limitada (50 linhas),

devido ao alto custo para um treinamento com mais linhas. O dataset em questão é o arquivo `./data/dataset2.jsonl`. Os testes foram realizados usando produtos disponíveis neste dataset.

O prompt de sistema informa claramente que o modelo é um assistente que deve fornecer a descrição de um produto inserido pelo usuário.

3. Instalação de Biblioteca Necessárias para LLAMA

- As bibliotecas *accelerate*, *peft*, *bitsandbytes*, *transformers* e *trl* são instaladas para fornecer as ferramentas necessárias para realizar o *fine-tuning* e a quantização de modelos.
-

2. Configuração do Ambiente para LLAMA

- O código importa PyTorch e outros módulos da biblioteca **Hugging Face** para manipulação de modelos e datasets.2. Processo de fine-tuning
- As classes **AutoModelForCausalLM**, **Auto Tokenizer**, e **SFTTrainer** são utilizadas para carregar o modelo base e realizar o *fine-tuning*.

4. Chamada do Foundation Model LLAMA

- O modelo base escolhido é o **Llama-2-7b-chat-hf**, carregado da plataforma **Hugging Face**.
- Um teste inicial é executado para medir a performance do modelo antes do *fine-tuning*.

5. Processo de Fine-Tuning do LLAMA

- O modelo é configurado para executar o *fine-tuning* utilizando a quantização em 4 bits através da configuração **BitsAndBytesConfig**.
- O processo de ajuste fino utiliza a técnica **LoRA** (*Low-Rank Adaptation*) para treinar o modelo em menor escala, economizando recursos computacionais.
- Parâmetros de treinamento: 15 épocas, tamanho de lote de 4, taxa de aprendizagem 0.001 e otimizador *Paged AdamW*.

Resultados

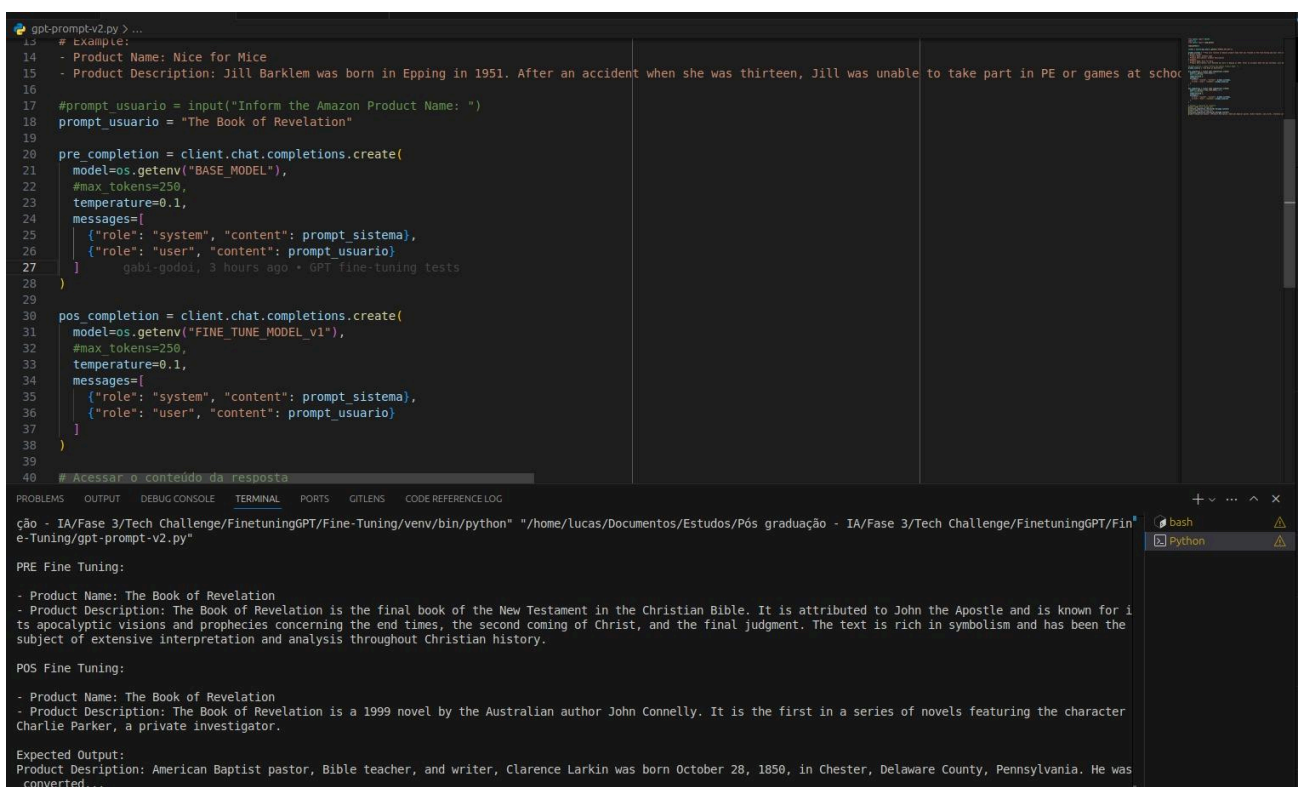
1. OpenAI

O algoritmo está atingindo o objetivo proposto de atuar como um assistente virtual que informa a descrição de produtos inseridos no prompt de usuário.

A descrição está de acordo com os produtos, porém, ela não é exatamente o conteúdo da base de dados treinada.

Na imagem a seguir, trazemos o exemplo do “The Book of Revelation”, com a resposta do modelo base (Pre Fine Tuning) , resposta do modelo treinado (Pos Fine Tuning) e a descrição fornecida na base de dados.

Veja que na base de dados a descrição do produto inicia com a descrição do autor, enquanto que, no modelo tanto pré como pós treino, a resposta traz apenas a descrição do produto.



```
gpt-prompt-v2.py > ...
13 # Example:
14 - Product Name: Nice for Mice
15 - Product Description: Jill Barklem was born in Epping in 1951. After an accident when she was thirteen, Jill was unable to take part in PE or games at school.
16
17 #prompt_usuario = input("Inform the Amazon Product Name: ")
18 prompt_usuario = "The Book of Revelation"
19
20 pre_completion = client.chat.completions.create(
21     model=os.getenv("BASE_MODEL"),
22     #max_tokens=250,
23     temperature=0.1,
24     messages=[
25         {"role": "system", "content": prompt_sistema},
26         {"role": "user", "content": prompt_usuario}
27     ]
28 )
29
30 pos_completion = client.chat.completions.create(
31     model=os.getenv("FINE_TUNE_MODEL_v1"),
32     #max_tokens=250,
33     temperature=0.1,
34     messages=[
35         {"role": "system", "content": prompt_sistema},
36         {"role": "user", "content": prompt_usuario}
37     ]
38 )
39
40 # Acessar o conteúdo da resposta

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS GIT LENS CODE REFERENCE LOG

ção - IA/Fase 3/Tech Challenge/FinetuningGPT/Fine-Tuning/venv/bin/python" "/home/lucas/Documents/Estudos/Pós graduação - IA/Fase 3/Tech Challenge/FinetuningGPT/Fin
e-Tuning/gpt-prompt-v2.py"

PRE Fine Tuning:
- Product Name: The Book of Revelation
- Product Description: The Book of Revelation is the final book of the New Testament in the Christian Bible. It is attributed to John the Apostle and is known for i
ts apocalyptic visions and prophecies concerning the end times, the second coming of Christ, and the final judgment. The text is rich in symbolism and has been the
subject of extensive interpretation and analysis throughout Christian history.

POS Fine Tuning:
- Product Name: The Book of Revelation
- Product Description: The Book of Revelation is a 1999 novel by the Australian author John Connolly. It is the first in a series of novels featuring the character
Charlie Parker, a private investigator.

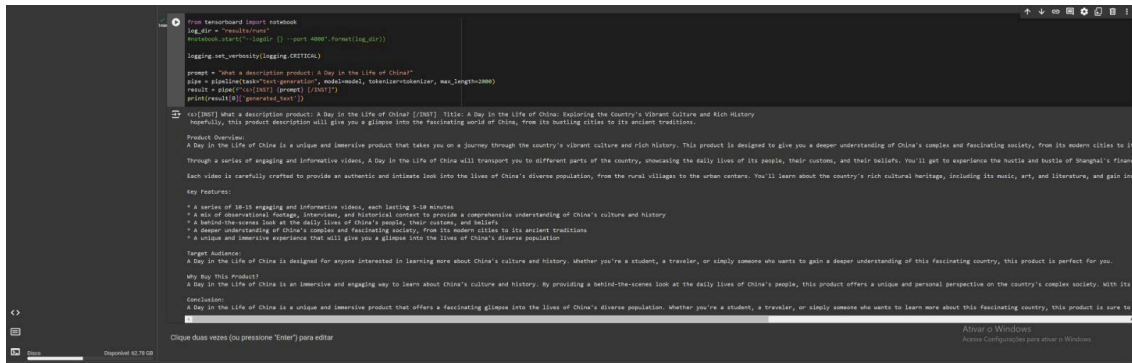
Expected Output:
Product Description: American Baptist pastor, Bible teacher, and writer, Clarence Larkin was born October 28, 1850, in Chester, Delaware County, Pennsylvania. He was
converted...
```

2. LLAMA

O resultado foi muito similar ao caso da OpenAI, em que o algoritmo está atingindo o objetivo proposto de atuar como um

assistente virtual que informa a descrição de produtos inseridos no prompt de usuário, porém, ela não é exatamente o conteúdo da base de dados treinada.

Na imagem abaixo é possível verificar o teste feito com a série “A day in the life of China”, em que é retornada a descrição detalhada do produto.



```
from IPython.display import Image
img_url = "https://i.imgur.com/1234567.jpg"
!python3 -c "import sys; sys.stdout.write('A Day in the Life of China\n'); sys.stdout.flush()"
logging.info('Logging CRITICAL')

prompt = "What a description product: A Day in the Life of China"
pipe = pipeline(task="text-generation", model_name="gpt-4", tokenizer="gpt2", max_length=2000)
result = pipe(prompt)
print(result[0]['generated_text'])
```

Product Overview:
A Day in the Life of China is a unique and immersive product that takes you on a Journey through the country's vibrant culture and rich history. This product is designed to give you a deeper understanding of China's complex and fascinating society, from its modern cities to its ancient traditions.

Through a series of engaging and informative videos, A Day in the Life of China will transport you to different parts of the country, showcasing the daily lives of its people, their customs, and their beliefs. You'll get to experience the hustle and bustle of Shanghai's financial district, the serene beauty of the Great Wall, and the traditional tea ceremony in a rural village.

Each video is carefully crafted to provide an authentic and intimate look into the lives of China's diverse population, from the rural villages to the urban centers. You'll learn about the country's rich cultural heritage, including its music, art, and literature, and gain insights into the challenges and opportunities it faces in the modern world.

Key Features:

- A series of 10-15 engaging and informative videos, each lasting 5-10 minutes.
- A mix of observational footage, interviews, and historical context to provide a comprehensive understanding of China's culture and history.
- A behind-the-scenes look at the daily lives of China's people, their customs, and beliefs.
- A deeper understanding of China's complex and fascinating society, from its modern cities to its ancient traditions.
- A unique and immersive experience that will give you a glimpse into the lives of China's diverse population.

Target Audience:

A Day in the Life of China is designed for anyone interested in learning more about China's culture and history. Whether you're a student, a traveler, or simply someone who wants to gain a deeper understanding of this fascinating country, this product is perfect for you.

Why Buy This Product?

A Day in the Life of China is an immersive and engaging way to learn about China's culture and history. By providing a behind-the-scenes look at the daily lives of China's people, this product offers a unique and personal perspective on the country's complex society. With its high-quality production and informative content, it's a must-have for anyone looking to explore the wonders of China.

Conclusion:

A Day in the Life of China is a unique and immersive product that offers a fascinating glimpse into the lives of China's diverse population. Whether you're a student, a traveler, or simply someone who wants to learn more about this fascinating country, this product is sure to provide you with a wealth of knowledge and a deeper understanding of China's rich culture and history.

Conclusão

O código cumpre o objetivo do **Tech Challenge** de realizar o *Fine-Tuning* de um modelo de linguagem grande utilizando o dataset **AmazonTitles-1.3MM**.

O fluxo de trabalho documentado garante que o modelo esteja preparado para responder perguntas baseadas em descrições de produtos, oferecendo uma base sólida para aplicações em processamento de linguagem natural no contexto de comércio eletrônico.

A equipe verificou, que mesmo após o Fine-Tuning ser realizado, o assistente retorna um resumo geral das descrições dos produtos contidos no dataset, e não a descrição exata. Por isso, para análise futura, o grupo pretende aplicar o modelo de **RAG** para testar se esse método gera descrições mais parecidas com o input direto do dataset.