

1. Documento do Processo de Seleção e Preparação do Dataset

Dataset Utilizado: Amazon Products (trn.json)

Processo de Preparação:

- Carregamento de 200 produtos do arquivo JSON comprimido
- Formatação expandida criando 3 variações de pergunta por produto
- Total de 294 exemplos de treinamento gerados
- Estrutura de dados no formato Alpaca (instruction, input, output)

Transformações Aplicadas:

python

Variações de perguntas criadas:

```
questions = [  
    f"Tell me about this product: {title}",  
    f"What are the characteristics of: {title}",  
    f"Describe in detail: {title}",  
    f"What can you tell me about: {title}",  
    f"I need information about: {title}"  
]
```

2. Processo de Fine-tuning e Código-fonte

Modelo Base: Llama-3-8B (unsloth/llama-3-8b-bnb-4bit)

Configuração LoRA:

```
model = FastLanguageModel.get_peft_model(  
    model,  
    r=16,                # Rank  
    target_modules=["q_proj", "k_proj", "v_proj", "o_proj",  
                   "gate_proj", "up_proj", "down_proj"],  
    lora_alpha=16,  
    lora_dropout=0,  
    bias="none",  
    use_gradient_checkpointing="unsloth"  
)
```

Parâmetros de Treinamento:

- Batch size: 2 por device
- Gradient accumulation: 4 steps

- Learning rate: 2e-4
- Max steps: 100
- Warmup steps: 15
- Otimizador: adamw_8bit
- Scheduler: linear

Resultados:

- Loss inicial: 3.189
- Loss final: 0.449
- Parâmetros treináveis: 41,943,040 (0.52% do modelo total)

3. Demonstração do Modelo Treinado

Exemplo de Resposta Pós Fine-tuning:

Input: "Describe Sony WH-1000XM4 headphones"

Output: "The WH-1000XM4 Wireless Noise Cancelling Headphones improve on already excellent predecessors. They're lighter, and offer improved sound, better voice-calling performance and longer battery life. They're also packed with features, including multipoint Bluetooth pairing, Amazon's Alexa and Google Assistant voice control, and an app that lets you personalize your experience."

Melhorias Observadas:

- Respostas mais específicas e detalhadas sobre produtos
- Foco em características técnicas relevantes
- Linguagem comercial apropriada para descrições de produtos
- Capacidade de generalizar para produtos não vistos no treinamento

Código para Teste:

```
test_prompt = alpaca_prompt.format(
    "DESCRIBE THIS AMAZON PRODUCT",
    "Tell me about: [PRODUCT_NAME]",
    ""
)
```

O modelo demonstra sucesso no fine-tuning específico para descrições de produtos Amazon, mantendo qualidade e especificidade nas respostas.

Link Github: <https://github.com/RogerPhilippe/TechChallenge-Fase3>

Link Youtube: https://youtu.be/Tp25_H6O5Oo