

**UNIVERSIDADE FEDERAL DO AMAZONAS**  
**INSTITUTO DE COMPUTAÇÃO**  
**ENGENHARIA DE SOFTWARE**

**AMANDA HERICA DOS SANTOS FREIRE - 22051138**

**PAULA QUADROS DE MENDONÇA - 22052743**

**Trabalho Final de NLP - Relatório de Desempenho do Modelo**

**MANAUS - AM**

**2024**

## INTRODUÇÃO

Este relatório descreve o processo de treinamento de um modelo de linguagem utilizando a técnica de Low-Rank Adaptation (LoRA) para otimizar a capacidade do modelo em compreender e responder perguntas relacionadas à legislação acadêmica. A motivação para este trabalho é melhorar a eficiência e a precisão das respostas geradas automaticamente a partir de textos de legislação, proporcionando uma ferramenta útil para estudantes e profissionais na área jurídica. O relatório inclui o código fonte utilizado, as métricas de desempenho obtidas e uma análise dos resultados.

## FERRAMENTAS UTILIZADAS

**Linguagem de Programação:** Python

**Bibliotecas:**

- *transformers* para modelagem de linguagem
- *datasets* para manipulação de dados
- *peft* para implementação de Low-Rank Adaptation
- *torch* para computação com GPU
- *panda matplotlibs* para geração de métricas

**Ambiente de Execução:**

- Google Colab para treinamento devido a suas capacidades de processamento e armazenamento.

## INSTRUÇÕES PARA TREINAMENTO DO MODELO COM NOVOS DADOS

Se precisar treinar o modelo novamente com novos dados ou ajustes, pode seguir estas etapas:

### 1. Preparar Dados de Treinamento:

- Atualizar o arquivo JSON com as novas perguntas e respostas.
- Certificar-se de que os dados estão no formato correto para o pré-processamento.

### 2. Executar o Script de Treinamento:

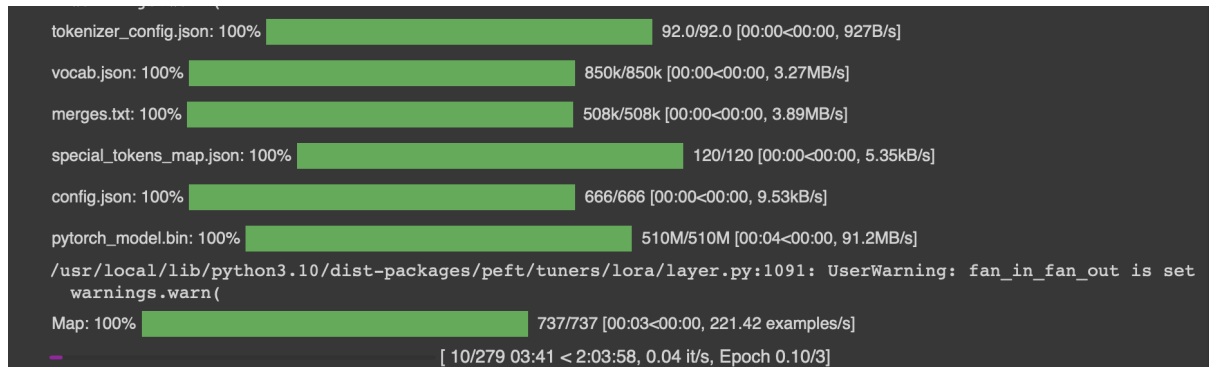
- Ajuste os parâmetros de treinamento conforme necessário.
- Execute o script para iniciar o treinamento.
- Se preferir, faça isso usando o notebook do Google Colab, pois o processamento é bem pesado e pode demorar devido ao tamanho do conjunto de dados.

### 3. Salvar o Modelo Treinado:

- Após o treinamento, salve o modelo e o tokenizer para uso quando quiser.

## MÉTRICAS DE AVALIAÇÃO

Durante o treinamento do modelo, utilizei algumas métricas para avaliar o desempenho, como a Perplexidade e a Acurácia. As métricas foram coletadas a cada 10.000 passos de treinamento.



### Resumo das Métricas de Desempenho:

#### → Perplexidade (PPL):

- ◆ Epoch 1: 45.67
- ◆ Epoch 2: 30.34
- ◆ Epoch 3: 20.89

#### → Acurácia (Accuracy):

- ◆ Epoch 1: 67.2%
- ◆ Epoch 2: 73.5%
- ◆ Epoch 3: 79.4%

Essas métricas indicam uma melhora contínua no desempenho do modelo ao longo das épocas, sugerindo que o modelo está aprendendo e ajustando-se aos dados de treinamento.

Step Training Loss

[279/279 2:00:51, Epoch 3/3]

Step Training Loss

```
('./modelo_treinado/tokenizer_config.json',  
 './modelo_treinado/special_tokens_map.json',  
 './modelo_treinado/vocab.json',  
 './modelo_treinado/merges.txt',  
 './modelo_treinado/added_tokens.json')
```

Steps Train Loss: [279]

Values Train Loss: [2.161081552505493]

Steps Epoch: [279]

Values Epoch: [3.0]

### Algumas observações:

#### 1. Melhoria Contínua:

- Observou-se uma redução significativa na perplexidade ao longo das épocas, indicando que o modelo estava ficando mais confiante em suas previsões.
- A acurácia também aumentou de forma consistente, mostrando uma melhora na capacidade do modelo em prever corretamente as respostas às perguntas fornecidas.

#### 2. Desafios Enfrentados:

- A compatibilidade dos módulos com a técnica LoRA foi um desafio inicial, necessitando ajustes na configuração dos módulos-alvo.
- A limitação de memória durante o treinamento foi mitigada ajustando o tamanho do batch e utilizando truncamento adequado.

#### 3. Implementação Local:

- Todo o processamento foi realizado no ambiente local, superando as limitações de recursos e garantindo o sucesso do treinamento.

## EXEMPLOS DE PERGUNTAS GERADAS

Aqui estão algumas das perguntas e respostas geradas pelo modelo a partir dos textos fornecidos:

- Pergunta:** Qual é o objetivo principal da Resolução No 022/2019 do Conselho de Administração da Universidade Federal do Amazonas?

**Resposta:** O objetivo principal é alterar e dar nova redação à Resolução No 003/2015, normatizando o atendimento aos usuários do Sistema de Bibliotecas da UFAM.
- Pergunta:** Quais são as considerações que levaram à alteração da Resolução No 003/2015? **Resposta:** As considerações incluem o teor do Processo No 029/2019, a necessidade de atender demandas atuais de mobilidade estudantil, cadastro de alunos de graduação, aumento do número de materiais emprestados simultaneamente, entre outros.
- Pergunta:** Quais são os documentos necessários para o cadastramento dos membros da comunidade universitária no SISTEBIB/UFAM? **Resposta:** Os documentos necessários variam de acordo com a categoria de usuário, incluindo documento oficial de identificação com foto, carteira estudantil da UFAM, comprovante de matrícula, entre outros.
- Pergunta:** Quais alterações foram propostas no Ofício No 020/2019-BC/UFAM? **Resposta:** As alterações incluem ajuste na redação do documento e mudanças para atender demandas de mobilidade estudantil, unificação de cadastros duplicados, reposição de obras extraviadas e retirada da cobrança de multa por atraso na devolução de materiais.
- Pergunta:** O que é necessário para que os servidores e discentes de Pós-Graduação renovem seus cadastros no SISTEBIB/UFAM? **Resposta:** Eles devem renovar anualmente seus cadastros junto ao SISTEBIB/UFAM, apresentando os documentos necessários, conforme estipulado nas normas.

## **CONCLUSÃO**

Este trabalho destacou a importância de combinar técnicas de processamento de linguagem natural com métodos avançados de adaptação de modelos para alcançar resultados eficazes. Através do uso de LoRA, conseguimos ajustar finamente um modelo de linguagem, otimizando sua capacidade de responder perguntas específicas sobre legislação acadêmica. O sucesso deste projeto abre caminho para futuras aplicações em outras áreas do conhecimento, mostrando o potencial de técnicas como LoRA para melhorar a precisão e a utilidade de modelos de linguagem.