



UNIVERSIDADE FEDERAL DA BAHIA - UFBA
Instituto de Humanidades, Artes e Ciências
Professor Milton Santos - IHAC

Alberto Pereira de Queiroz Junior
Ícaro Mello Brandão
Lucas Rocha Marques
Manoel Ribeiro de Jesus
Pedro Henrique França Bonfim

Benchmark de Modelos LLM acerca do Código de Defesa do Consumidor

SALVADOR/BA
2025

Artigo apresentado como requisito parcial para obtenção da aprovação na disciplina HACA46 - TÓPICOS ESPECIAIS EM TECNOLOGIA I, na Universidade Federal da Bahia.

Prof. Genaro Fernandes

SALVADOR/BA

2025

RESUMO

Palavras - Chaves: llm, inteligência artificial, modelos de linguagem, robos,

1. INTRODUÇÃO

Nos últimos anos, a inteligência artificial IA se consolidou como uma ferramenta essencial para o mundo com sua forma de simplificar o aprendizado e oferecer acesso a informações de forma mais ágil e eficiente. No campo jurídico, seu uso vem se expandindo, começando com aplicações básicas em serviços administrativos, e atendimentos básicos e evoluindo para soluções mais avançadas. Entre essas, destaca-se a possibilidade de facilitar o entendimento de legislações que estão presentes no cotidiano dos cidadãos, como o Código de Defesa e Proteção do Consumidor (CDC), regulamentado pela Lei nº 8.078 de 1990.

Este projeto visa apresentar o desenvolvimento de uma ferramenta baseada em IA, projetada para responder perguntas relacionadas ao CDC, utilizando como base dúvidas, problemas, práticas vivenciadas por consumidores e empresas. Para isso, foram empregados modelos de IA, como o Chat GPT e o LLAMA 3.2, que foram apresentados a 69 questões organizadas em três formatos principais: perguntas diretas, múltipla escolha e questões objetivas (verdadeiro ou falso) com o objetivo de avaliar seu entendimento do CDC e avaliar suas respostas, tornando viável como uma ferramenta de consulta com o entendimento do CDC mais acessível, permitindo que as pessoas conheçam seus direitos de forma prática e inclusiva.

Os testes realizados evidenciaram grandes diferenças no desempenho dos modelos. O Chat GPT alcançou 86,95% de acertos, demonstrando grande potencial para ser utilizado como suporte na educação e esclarecimento de dúvidas sobre o CDC. Por outro lado, o LLAMA 3.2

obteve um baixo índice de acertos, com apenas 17,40%, evidenciando a necessidade de maior treinamento para atender às demandas. Esses resultados demonstraram a capacidade do Chat GPT para aplicações práticas no auxílio das dificuldades e entendimento do CDC, sobretudo em tópicos com bastante importância e procura de consumidores sobre o CDC como responsabilidade de fornecedores, formas de pagamento e direitos básicos do consumidor.

2. DESENVOLVIMENTO

Para a criação desta ferramenta uma série de passos foram desenhados a fim de otimizar tempo e garantir o bom funcionamento da IA de acordo com o esperado. A princípio com pesquisas de temas importantes para a sociedade e que tenha uma grande dificuldade para o cidadão, em seguida a criação de questionamentos e dúvidas possíveis embasadas em pesquisas e conhecimentos próprios de cada membro.

Diante de todo o entendimento já adquirido foi necessário a realização de uma avaliação com ferramentas diferentes de IA para um melhor entendimento da capacidade de resposta quanto ao assunto levado.

2.1 - APLICAÇÃO BRUTE FORCE DO QUESTIONÁRIO

2.1.1 - Criação do questionário

As perguntas foram elaboradas através dos conhecimentos prévios de cada membro da equipe. Foi decidido que seriam avaliadas três tipos de respostas: discursivas, objetivas de ‘certo ou errado’ e objetivas de múltipla escolha (com 4 opções de resposta). Assim, foi estabelecido o número de 25 questões por membro da equipe para serem formuladas. Após, foi feita uma filtragem nas perguntas similares e chegamos ao número de 69 perguntas.

3- Estabelecimentos devem ser responsáveis pelos veículos estacionados em seus estacionamentos?

Os estabelecimentos são responsáveis pelos veículos em seu estacionamento, mesmo que o mesmo seja apenas cortesia. De acordo com a súmula 130 do STJ, a empresa responde, perante o cliente, pela reparação de dano ou furto de veículo ocorrido em seu estacionamento.

Imagem 1 - Pergunta de resposta discursiva

63 O Código de Defesa do Consumidor (CDC) é uma importante legislação voltada para a proteção das relações de consumo no Brasil. No entanto, algumas áreas do Direito não são abrangidas por essa legislação. Assinale a alternativa que corretamente exemplifica um conflito de interesses que **não** é regido pelo CDC:

- a) A cobrança de taxas abusivas em contratos de financiamento celebrados entre consumidores e instituições financeiras;
- b) A recusa injustificada de fornecimento de garantia em produtos adquiridos por consumidores;
- c) A disputa entre empregador e empregado sobre o pagamento de verbas rescisórias após a demissão;**
- d) A entrega de produtos com defeito, gerando a necessidade de substituição ou reparo;
- e) A negativa de cobertura de um procedimento médico contratado em um plano de saúde;

Resposta: C

Imagem 2 - Pergunta de resposta objetiva

26 Os preços de produtos de um mesmo estabelecimento podem variar dependendo da plataforma de venda?

Sim. Não há obrigatoriedade de que o preço dos produtos seja o mesmo para os diversos canais de venda de um mesmo fornecedor.

Imagem 3 - Pergunta de resposta objetiva

2.1.2 - Geração do prompt

O prompt utilizado foi elaborado através de um esboço inicial e aprimoramento por parte dos membros da equipe em sala de aula. Tendo acesso a um computador, o prompt utilizado foi utilizado no modelo Chat GPT

Eu vou mandar pra você uma lista de questões, você deve responder elas de forma curta.

Da pergunta 1 até a 39, você deve responder a pergunta com um pequeno texto de no máximo 300 caracteres. Da pergunta 40 até 49 você deve dizer se a afirmação está correta ou errada. E nas perguntas objetivas, apenas dizer qual alternativa correta.

Você deve numerar as respostas de acordo com o número da pergunta.

1- Estabelecimentos podem cobrar valores diferentes para cartão de

Imagem 4 - Prompt

2.1.3 - Resultados do treinamento (o que a IA respondeu, os scores, quantidade de erros e acertos etc.)

Foi realizado teste da com as ferramentas Chat GPT e Llama 3.2, onde foram lançadas uma a uma as **69 questões produzidas e verificadas posteriormente as respostas**, onde computamos os erros e acertos e classificamos quais os tipos de questões mais tivemos erros. **A margem de acertos da IA Chat GPT foi de 86,95%**, o que foi considerado algo muito bom, pois a ferramenta possui grande quantidade de informações e acesso a resolver dúvidas. O resultado na Llama 3.2 foi bem contrário ao esperado, houve um baixo rendimento da IA quanto a responder às mesmas questões enviadas, **onde de acertos obteve apenas um total**

de 17,40%, demonstrando a necessidade de treinamento intensivo para a melhoria das respostas.

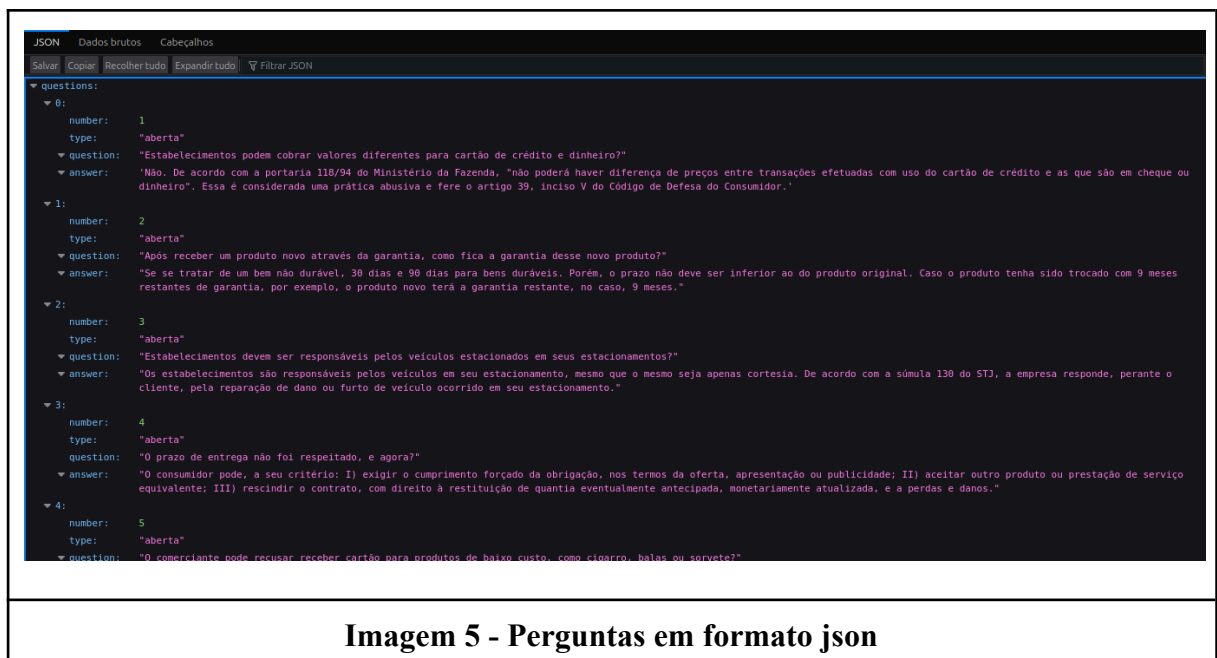
2.1.4 - Alterações realizadas com a finalidade de melhoramento das respostas do modelo (mudamos o prompt?, inserimos uma nova etapa de classificação das respostas fornecidas?)

Etapa do projeto não foi elaborada devido ao não desenvolvimento correto da etapa de automação da obtenção de respostas do questionário aplicado.

2.2 - APLICAÇÃO AUTOMATIZADA

2.2.1 - Geração do arquivo json:

O questionário inicial foi adaptado ao padrão json através da ferramenta Chat GPT, que auxiliou diversas vezes a equipe.



2.2.2 - Construção do código:

Para a execução desta etapa foi utilizada a ferramenta Python devido a seu amplo uso, principalmente pela plataforma de modelos utilizada como exemplo durante as aulas (<https://huggingface.co/>).

O código utilizado foi resultado da adaptação do disponível na página do modelo utilizado (<https://huggingface.co/meta-llama/Llama-3.2-1B>), com pequenas adaptações para atender a demanda específica, como a importação de bibliotecas específicas e blocos de código que realizam a própria tarefa de inquirir o modelo e outras essenciais.

```
1  import torch, json
2  from transformers import pipeline
3
4  file = open('perguntas_sem_resposta.json', 'r')
5  output = "respostas.txt"
6
7  data = json.load(file)
8
9  model_id = "meta-llama/Llama-3.2-1B"
10
11  pipe = pipeline("text-generation", model=model_id, torch_dtype=torch.bfloat16, device_map="auto")
12
13  with open(output, "w", encoding="utf-8") as output:
14      for i, item in enumerate(data["questions"], start=1):
15          # Obtendo a pergunta
16          question = item.get("question", "")
17
18          # Gerando a resposta para a pergunta
19          prompt = f"Responda a questão: {question}"
20          generated = pipe(prompt, max_length=100, num_return_sequences=1)
21          response = generated[0]["generated_text"].strip()
22
23          # Escrevendo os dados no arquivo de saída
24          output.write(f"Item {i}:\n")
25          output.write(f"Pergunta: {question}\n")
26          output.write(f"Resposta: {response}\n")
27          output.write("-" * 50 + "\n")
28
29          # Opcional: também imprimir no console
30          print(f"Processado Item {i}")
31
```

Imagem 6 - Código de execução principal

2.2.3 - Resultados:

Não foi possível coletar resultados devido ao não desenvolvimento do processo de automação necessário para coleta das respostas do questionário aplicado.

2.2.4 - Conclusão comparando o procedimento realizado sem a automatização e com a automatização;

A etapa do projeto foi suspensa devido ao não desenvolvimento do processo de automação necessário para coleta das respostas do questionário aplicado.

3. CONSIDERAÇÕES FINAIS

O desenvolvimento desta ferramenta baseada em inteligência artificial para responder questões relacionadas ao Código de Defesa do Consumidor destaca o potencial das tecnologias de IA no campo jurídico e educacional. Os testes realizados comprovaram a eficácia do modelo Chat GPT, que alcançou 86,95% de acertos, demonstrando sua capacidade de atuar como uma ferramenta prática para esclarecer dúvidas sobre o CDC. Por outro lado, o desempenho limitado do modelo LLAMA 3.2, com apenas 17,40% de acertos, evidenciou a necessidade de ajustes e treinamentos mais específicos para melhorar sua aplicação em cenários reais.

Apesar de algumas etapas, como a automação da coleta de respostas, não terem sido concluídas, o projeto gerou aprendizados valiosos para o desenvolvimento de metodologias e a aplicação de IA em problemas práticos. A criação de questionários e a análise detalhada das respostas mostraram como ferramentas de IA podem ser ajustadas para atender a demandas específicas, principalmente em temas que envolvem direitos dos consumidores, como responsabilidade de fornecedores e formas de pagamento.

Os resultados obtidos reforçam a viabilidade de ferramentas de IA para democratizar o acesso à informação de forma ágil e eficiente. Mesmo com os desafios encontrados, o projeto aponta para oportunidades de aprimoramento e incentiva a continuidade de iniciativas que busquem tornar a consulta de legislações mais acessível para consumidores e empresas, promovendo maior inclusão e conhecimento de direitos básicos no cotidiano.

4. REFERÊNCIAS

Hugging face: <https://huggingface.co/meta-llama/Llama-3.2-1B>

Chat GPT: <https://chatgpt.com/>

Git Hub: <https://github.com/marrocs/haca47-repo/tree/main>