



PLATAFORMA DE AVALIAÇÃO DE
MODELOS DE IA

**DOM
ROCK**



PLATAFORMA DE AVALIAÇÃO DE MODELOS DE IA

MANUAL DO USUÁRIO

FACULDADE DE TECNOLOGIA DE SÃO JOSÉ DOS CAMPOS – PROF. JESSEN
VIDAL

São José dos Campos – 1º semestre 2025

Sumário

• Introdução	3
• 1 Plataforma	4
• 2 Público-Alvo	4
• 3 Página Inicial	5
• 4 Loading	5
• 5 Página com as Duas Respostas	6
• 6 Avaliação por Critérios	7
• 7 Respostas com uma Avaliação Concluída	8
• 8 Ambas Respostas Avaliadas	9
• 9 Dashboard de Desempenho das LLMs	10
• 10 Glossário	11
• Passo a passo.....	12

Introdução



Este manual tem como objetivo orientar o usuário na utilização da Plataforma de Avaliação de Modelos de IA, desenvolvida para comparar respostas de diferentes modelos de linguagem. A plataforma foi projetada para oferecer uma experiência simples, rápida e eficiente, para que usuários avaliem qual modelo apresenta melhor desempenho em termos de clareza, empatia, relevância e veracidade de texto.

1. Plataforma

A Plataforma de Avaliação de Modelos de IA é acessada via navegador web.

Ela é responsiva, compatível com dispositivos móveis e utiliza tecnologias modernas para garantir uma experiência fluida.

Usuários podem interagir diretamente com modelos LLMs, comparar respostas e gerar avaliações.

2. Público-Alvo

Este manual destina-se a:

- Estudantes e pesquisadores de IA e PLN.
- Desenvolvedores e profissionais de tecnologia.
- Usuários interessados em avaliar e melhorar modelos de linguagem.
- Equipes de curadoria de conteúdo e treinadores de LLMs.

3. Página Inicial

A página inicial da plataforma exibe o campo onde o usuário pode digitar uma pergunta.

Após o envio, duas respostas são geradas, uma por cada modelo LLM.

4. Loading

Ao enviar a pergunta, a plataforma apresenta uma animação de carregamento enquanto as LLMs processam e retornam suas respostas.

5. Página com as Duas Respostas

A tela é dividida em dois blocos paralelos contendo as respostas de dois modelos distintos.

Cada bloco apresenta:

- O texto da resposta
- As fontes utilizadas (se houver)
- Um botão para o usuário selecionar a melhor resposta

6. Avaliação por Critérios

Cada resposta pode ser avaliada por critérios específicos:

Clareza – A resposta está bem escrita e fácil de entender?

Relevância – A resposta responde à pergunta de forma objetiva?

Veracidade – A resposta apresenta fatos corretos?

Empatia – A resposta demonstra linguagem respeitosa e adequada?

O usuário atribui notas ou marcações para cada critério e pode incluir um comentário explicando sua decisão.

7. Respostas com uma Avaliação Concluída

Após avaliar uma resposta, ela é marcada como "avaliada".

A outra permanece ativa até ser avaliada também.

O usuário pode retornar e reavaliar se desejar.

8. Ambas Respostas Avaliadas

Quando ambas as respostas são avaliadas, a plataforma solicita:

- Um feedback qualitativo no campo "Nos conte o que achou"
- A seleção do modelo que teve melhor desempenho geral
- A opção de acessar o dashboard de desempenho comparativo

9. Dashboard de Desempenho das LLMs

O painel interativo exibe:

- A comparação visual entre os modelos
- O número total de avaliações de cada modelo
- A taxa de respostas mais bem avaliadas

- A performance em cada critério (clareza, veracidade, etc.)
- Um botão “Finalizar” para concluir o processo e retornar à tela inicial

10. Glossário

LLM (Large Language Model):

Modelo de linguagem treinado com grandes volumes de dados, capaz de compreender e gerar textos de forma autônoma com base em prompts fornecidos pelo usuário.

Avaliação por Critérios:

Processo estruturado em que o usuário atribui pontuações ou marcações a respostas com base em parâmetros como clareza, empatia, veracidade e relevância.

Dashboard:

Painel visual interativo que apresenta métricas, gráficos e indicadores de desempenho das LLMs avaliadas.

Empatia:

Capacidade de um modelo de IA utilizar linguagem acolhedora, compreensiva e respeitosa nas respostas ao usuário.

Feedback Qualitativo:

Comentário textual fornecido pelo usuário após avaliar ambas as respostas, com impressões pessoais ou sugestões de melhoria.

Prompt:

Texto inserido pelo usuário para gerar uma resposta dos modelos de IA. É a pergunta ou comando inicial da interação.

Fontes:

Materiais ou referências utilizados pelas LLMs para justificar ou embasar a resposta fornecida.

Veracidade:

Critério que avalia se a resposta apresenta dados corretos e não gera desinformação.

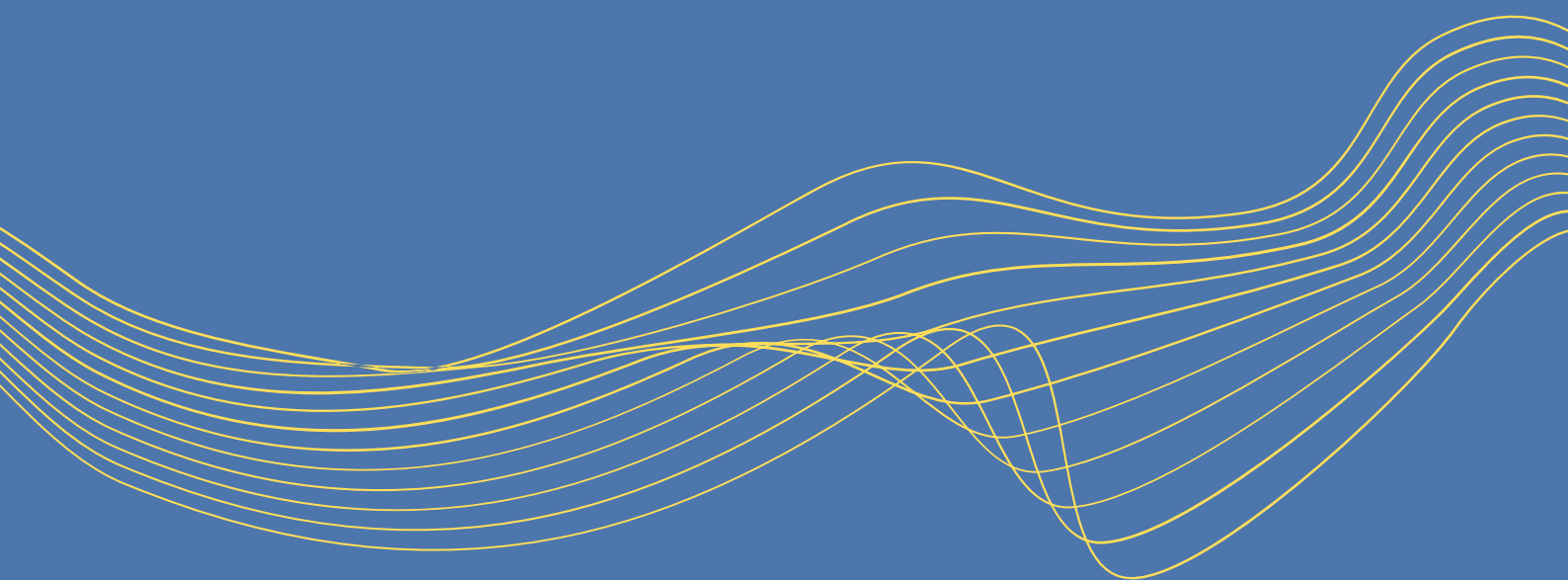
Relevância:

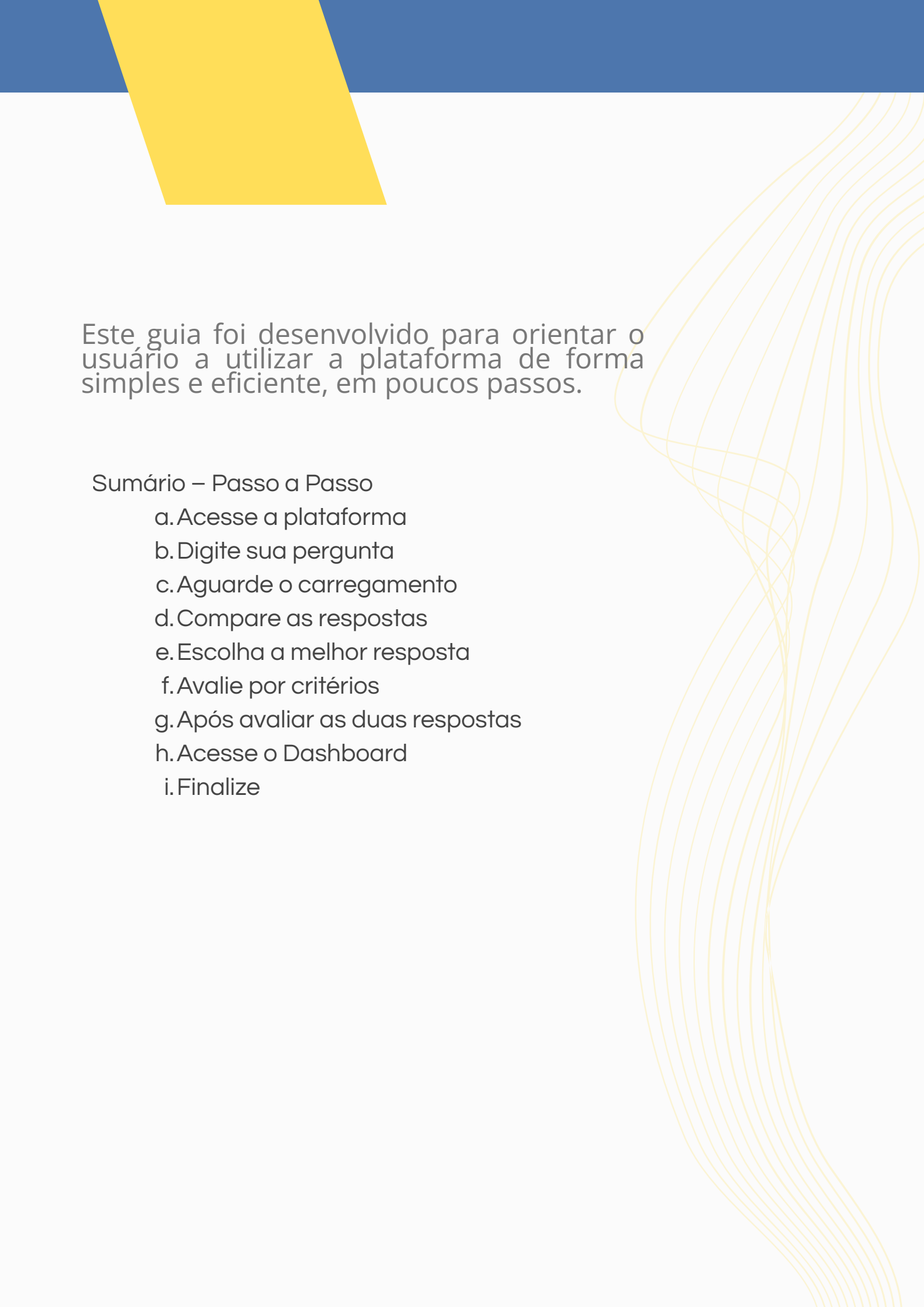
Critério que verifica se a resposta está de fato alinhada com o que foi perguntado, evitando dispersões ou respostas vagas.

Clareza:

Critério que julga se a linguagem da resposta é direta, compreensível, bem escrita e estruturada.

PASSO A PASSO – COMO UTILIZAR A PLATAFORMA DE AVALIAÇÃO DE MODELOS DE IA





Este guia foi desenvolvido para orientar o usuário a utilizar a plataforma de forma simples e eficiente, em poucos passos.

Sumário – Passo a Passo

- a. Acesse a plataforma
- b. Digite sua pergunta
- c. Aguarde o carregamento
- d. Compare as respostas
- e. Escolha a melhor resposta
- f. Avalie por critérios
- g. Após avaliar as duas respostas
- h. Acesse o Dashboard
- i. Finalize

1. Acesse a plataforma

- Abra seu navegador preferido.
- Digite o endereço da plataforma na barra de endereços.
- Aguarde o carregamento automático da interface.

2. Digite sua pergunta

- No campo inferior com o texto "Digite aqui sua pergunta", escreva sua dúvida ou comando.
- Certifique-se de usar linguagem clara e objetiva.
- Clique no botão de envio (ícone de seta azul) à direita do campo.

3. Aguarde o carregamento

- Um ícone animado aparecerá indicando que os modelos estão processando sua pergunta.
- Isso pode levar alguns segundos, dependendo da complexidade da pergunta.

4. Compare as respostas

- Duas respostas aparecerão lado a lado: uma do Modelo 1 e outra do Modelo 2.
- Leia ambas com atenção e analise o conteúdo.

5. Escolha a melhor resposta

- Clique no botão "Selecionar como melhor" no cartão da resposta que você considera superior.

6. Avalie por critérios

- A plataforma pode solicitar que você avalie a resposta com base em critérios como clareza, veracidade, relevância e empatia.
- Marque as opções disponíveis ou mova os sliders conforme sua percepção.
- Você pode escrever um comentário justificando sua escolha.
- Clique em "Salvar Resposta".

7. Após avaliar as duas respostas

- Quando ambas forem avaliadas, aparecerá um campo "Nos conte o que achou" para um feedback textual.
- Logo abaixo, selecione qual modelo teve melhor desempenho geral na pergunta respondida.
- Clique em "Visualizar Desempenho".

8. Acesse o Dashboard

- Veja gráficos que comparam o desempenho entre os modelos.
- Consulte dados como número de respostas avaliadas, taxa de melhor desempenho e critérios destacados.

9. Finalize

- Clique em "Finalizar" para concluir o processo e retornar à página inicial.
- Você pode iniciar uma nova avaliação a qualquer momento.