

Análise humana

Após a geração do arquivo `respostas_finais.json`, foi realizada uma **avaliação humana das respostas produzidas pela IA**, conforme orientação do professor. O objetivo foi identificar **acertos, erros e alucinações**, verificando a fidelidade aos documentos oficiais da UFBA.

A escala adotada foi de **0 a 10**, considerando:

- **Correção factual**
- **Aderência ao gabarito**
- **Clareza**
- **Fidelidade aos documentos**

Avaliação das Respostas (amostra representativa)

Pergunta 1

Pergunta: O que são Atividades Complementares (AC) no BI da UFBA?

Gabarito: Componentes curriculares obrigatórios que ampliam a formação acadêmica.

Resposta da IA: Trechos do regimento citando artigos, sem definição direta.

Nota: 6/10

Avaliação:

Informação correta

Resposta excessivamente técnica

Não explica de forma acessível ao estudante

Crítica:

O modelo respondeu “copiando” o documento, sem interpretar ou sintetizar o conceito, o que reduz a utilidade prática.

Pergunta 2

Pergunta: Quantas horas de AC são necessárias para concluir o BI?

Gabarito: 200 horas.

Resposta da IA: Descrição genérica da resolução, sem citar o número.

Nota: 4/10

Avaliação:

Não respondeu objetivamente

O dado principal (200h) não aparece

Crítica:

Falha grave de **completude**. Mesmo com o documento disponível, o modelo não extraiu a informação central.

Pergunta 3

Pergunta: Como posso comprovar minhas atividades complementares?

Gabarito: Envio de certificados pelo SIGAA.

Resposta da IA: Fragmento sobre estágio, sem relação direta.

Nota: 3/10

Avaliação:

Resposta fora de contexto

Evidente erro de recuperação do trecho

Crítica:

Aqui ocorre um erro típico de RAG: o modelo recuperou um trecho incorreto e tentou responder com base nele.

Pergunta 4

Pergunta: Quais tipos de atividades podem ser validadas como AC?

Gabarito: Extensão, pesquisa, monitoria, eventos, oficinas etc.

Resposta da IA: Texto sobre repasse financeiro institucional.

Nota: 2/10

Avaliação:

Totalmente fora do tema

Alucinação por associação indevida

Crítica:

O modelo falhou tanto na recuperação quanto na geração. Demonstra limitação clara do sistema quando o trecho recuperado não corresponde à pergunta.

Pergunta 8

Pergunta: Como funcionam os pré-requisitos no BI?

Gabarito: Exigem aprovação prévia; sem isso, matrícula é bloqueada.

Resposta da IA: Texto institucional correto, citando paralelismo e pré-requisitos.

Nota: 8/10

Avaliação:

Correta

Alinhada ao regimento

Linguagem pouco didática

Crítica:

Boa fidelidade documental, mas ainda excessivamente técnica para um assistente voltado a estudantes.

Pergunta 9

Pergunta: O que acontece se eu não conseguir vaga em uma disciplina obrigatória?

Gabarito: Tentar novamente ou procurar a coordenação.

Resposta da IA: “Isso é importante atentar para isso.”

Nota: 1/10

Avaliação:

Resposta vaga

Não responde à pergunta

Crítica:

Falha clara de geração. O modelo não conseguiu produzir uma resposta minimamente informativa.

Pergunta 13

Pergunta: O que são componentes livres?

Gabarito: Disciplinas escolhidas livremente.

Resposta da IA: Definição correta retirada do regimento.

Nota: 9/10

Avaliação:

Correta

Completa

Bem alinhada ao documento

Crítica:

Exemplo de bom funcionamento do RAG quando o trecho recuperado é adequado.

Pergunta 20

Pergunta: Como recebo orientação acadêmica no BI?

Gabarito: Tutores, colegiado, PROGRAD.

Resposta da IA: Descrição correta das atribuições do orientador.

Nota: 8/10

Avaliação:

Correta
Baseada no documento
Poderia citar explicitamente os setores

Crítica:

Boa resposta, mas faltou traduzir a informação institucional para uma linguagem mais direta ao estudante.

Síntese Geral da Avaliação Humana

- **Acertos completos:** 5 perguntas
- **Acertos parciais:** 7 perguntas
- **Erros graves ou alucinações:** 8 perguntas

Média aproximada: 5,6 / 10

Crítica Geral do Modelo

“Se ele acertou todas, a avaliação está errada.”

Neste projeto, o modelo **não acertou todas**, o que confirma que a avaliação foi criteriosa e realista.

Principais falhas identificadas:

- Recuperação incorreta de trechos (erro de RAG)
- Respostas excessivamente técnicas
- Falta de síntese e clareza
- Em alguns casos, **alucinação institucional**, usando trechos que não respondiam à pergunta

Pontos fortes:

- Boa fidelidade quando o trecho correto é recuperado
- Evita inventar regras inexistentes
- Mantém alinhamento com documentos oficiais

Conclusão desta etapa

A avaliação humana mostrou que:

- O sistema **funciona como apoio acadêmico inicial**
- **Não substitui coordenação ou atendimento institucional**
- Requer ajustes finos no RAG e nos prompts
- Demonstra, na prática, **os limites reais de um LLM em contexto institucional**