

Desafio Java Spring Boot com LLM – Alura

Laisla Stephanie Gomes Barbosa

Conclusão

Nos testes iniciais, o modelo apresentou dificuldades para identificar corretamente as funcionalidades do aplicativo mencionadas, possivelmente devido ao viés causado pelo sistema de memória. No entanto, após reinicializá-lo e realizar uma nova bateria de testes, o modelo desempenhou adequadamente suas funções de detectar e classificar intents, identificar sentimentos, determinar se o comentário era ofensivo ou inofensivo e gerar respostas para o usuário com base nas funcionalidades destacadas.

No segundo teste, as classificações de intents foram mais precisas, embora as respostas tenham sido geradas em inglês, um ponto que pode ser aprimorado com o uso de um modelo mais adequado ao português (apesar de o **gemma2** já mostrar bons resultados). Sinta-se à vontade para testar a funcionalidade de "Harmless", indo além dos cenários mais simples, em que todos os comentários são inofensivos. É crucial testar todos os fluxos possíveis que programamos no modelo LLM.

Quanto às respostas do assistente, elas foram sempre simpáticas e adotaram um tom próximo, respondendo de acordo com as intents dos usuários. No entanto, o desempenho poderia ser aprimorado com a implementação de um sistema **RAG** (Retrieval-Augmented Generation), que proporcionaria uma base mais sólida para as respostas, tornando-as mais úteis para o usuário. Isso poderia ser feito com a utilização de um sistema de **embedding**, empregando modelos específicos e uma base de dados vetorial para calcular a proximidade semântica e gerar respostas mais realistas e alinhadas às regras de negócio.

Outra funcionalidade interessante seria a integração com um **dashboard analítico** impulsionado por IA, que forneceria insights valiosos a partir da base de dados, identificando quais áreas do aplicativo recebem mais ou menos feedback, as reclamações mais recorrentes, o risco de churn, além das plataformas de comentário mais utilizadas. Isso auxiliaria os analistas a tomarem decisões mais informadas sobre o aplicativo.