

Uso de Inteligência Artificial no Desenvolvimento de um Chatbot para Auxílio Acadêmico e Administrativo no IFPI

Larissa Souza do Nascimento¹, Ricardo Moura Sekeff Budaruiche¹

¹Instituto Federal de Educação Ciência e Tecnologia do Piauí (IFPI)
Avenida Rio dos Matos, S/N, Bairro Germano – 64.260-000 – Piripiri – PI – Brazil

larissasouza.nascimento01@gmail.com ricardo.budaruiche@ifpi.edu.br

Abstract. This paper presents the proposed development of an institutional chatbot for IFPI, leveraging advanced artificial intelligence technologies. The solution integrates the LLaMA 3 language model, the LangChain framework, and the ChromaDB vector database within an AI orchestration architecture. The system aims to deliver precise and contextualized responses to user queries, optimizing communication and institutional information management. The methodology includes fine-tuning strategies and Retrieval-Augmented Generation (RAG), enabling continuous chatbot personalization based on user feedback and normative updates.

Resumo. Este artigo apresenta a proposta de desenvolvimento de um chatbot institucional para o IFPI, utilizando tecnologias avançadas de inteligência artificial. A solução integra o modelo de linguagem LLaMA 3, o framework LangChain e o banco de dados vetorial ChromaDB, dentro de uma arquitetura baseada na orquestração de IA. O sistema visa oferecer respostas precisas e contextualizadas às consultas dos usuários, otimizando a comunicação e a gestão de informações institucionais. A metodologia contempla estratégias de fine-tuning e Retrieval-Augmented Generation (RAG), permitindo a personalização contínua do chatbot com base em feedbacks e atualizações normativas.

1. Introdução

A administração e a gestão acadêmica em instituições de ensino superior enfrentam desafios significativos decorrentes do elevado volume de dados e documentos regulatórios que demandam constante gerenciamento. No contexto do Instituto Federal do Piauí (IFPI), esses desafios se intensificam devido à diversidade de normativas, resoluções e manuais operacionais que norteiam o funcionamento institucional. O IFPI registra inúmeros atendimentos mensais relacionados a demandas acadêmicas e administrativas, enfrentando como principais dificuldades a demora no tempo médio de resposta e a sobrecarga de servidores para esclarecer dúvidas recorrentes. A gestão eficiente desses documentos é essencial para assegurar a transparência e a agilidade dos processos acadêmicos e administrativos; contudo, tal atividade é frequentemente complexa e suscetível a erros, em razão da grande quantidade e da complexidade das informações envolvidas.

Com o avanço das tecnologias de Inteligência Artificial (IA), novas possibilidades têm emergido para otimizar a gestão e o acesso a esses dados. APIs de IA generativa, como GPT e Gemini, viabilizam o processamento e a interpretação de grandes volumes

de informações de maneira eficiente. Nesse cenário, este trabalho propõe o desenvolvimento de um *chatbot* inteligente, voltado a auxiliar servidores e discentes na busca e interpretação de normativas institucionais, resoluções acadêmicas e manuais administrativos.

A adoção de um *chatbot* baseado em IA representa uma solução inovadora frente aos desafios enfrentados pelo IFPI, ao permitir a oferta de respostas automatizadas, reduzir a sobrecarga decorrente de consultas repetitivas e melhorar a comunicação interna. Adicionalmente, tal tecnologia apresenta potencial para identificar inconsistências e ambiguidades presentes nos documentos institucionais, contribuindo para o aprimoramento contínuo dos processos administrativos e acadêmicos. A possibilidade de treinar o modelo com dados específicos do IFPI confere à solução um elevado grau de personalização, ampliando sua relevância e eficácia no contexto institucional.

Estudos recentes evidenciam que *chatbots* têm sido amplamente implementados em instituições de ensino com o objetivo de otimizar o atendimento a estudantes e facilitar o acesso à informação. Essas iniciativas demonstram que a IA pode transformar a experiência acadêmica, tornando os processos mais ágeis, acessíveis e eficientes. Inspirado por tais aplicações, este trabalho apresenta o IFPIA, um *chatbot* desenvolvido para atender às demandas acadêmicas e administrativas do IFPI, com o objetivo de reduzir tarefas manuais repetitivas e proporcionar uma experiência mais fluida aos usuários.

Este artigo está organizado da seguinte forma: a Seção 2 apresenta trabalhos relacionados com foco em classificação de textos e uso de *chatbots* no contexto educacional. A Seção 3 descreve a metodologia adotada para o desenvolvimento do *chatbot*. Na Seção 4, são discutidos os resultados obtidos até o momento. Por fim, a Seção 5 apresenta as conclusões e propõe direcionamentos para trabalhos futuros.

2. Referencial Teórico

A Inteligência Artificial (IA) generativa corresponde a um subconjunto de técnicas de IA voltadas para a criação e compreensão de textos, imagens, áudios e outros tipos de dados, com base em padrões aprendidos durante a fase de treinamento, que envolve o uso de grandes volumes de dados. Entre as abordagens mais proeminentes nesse campo, destacam-se os modelos baseados na arquitetura de Transformadores, como o *Generative Pre-trained Transformer* (GPT), desenvolvido pela OpenAI, e o Gemini, desenvolvido pela Google. Esses modelos têm revolucionado o campo da IA, demonstrando capacidades avançadas na geração de conteúdo coerente e contextualizado, no processamento de linguagem natural (NLP) e na resolução de tarefas complexas a partir de entradas textuais.

Os modelos fundamentados em transformadores, conforme descrito por [Vaswani et al. 2017], utilizam mecanismos de atenção que permitem considerar todas as partes de uma sequência de dados ao gerar cada elemento subsequente. Tal abordagem possibilita a captura de relacionamentos de longo alcance nos textos. O GPT, em particular, tem sido amplamente estudado e aprimorado, sendo o GPT-4 um dos modelos mais avançados atualmente, com treinamento baseado em 175 bilhões de parâmetros [et al. 2024]. Esse volume expressivo de parâmetros confere ao modelo uma notável capacidade de gerar textos em diversos estilos, executar tarefas relacionadas à NLP e responder com precisão e coerência a perguntas formuladas em NLP.

No contexto institucional, a aplicação de IA generativa configura-se como uma

solução promissora para a análise e interpretação de documentos regulatórios e normativos. De acordo com [Susskind 2019], o uso de IA na interpretação de textos legais pode automatizar a análise de grandes volumes documentais, facilitando a compreensão de normas e procedimentos e permitindo a identificação de eventuais ambiguidades. Esse tipo de aplicação contribui não apenas para a aceleração dos processos de revisão e auditoria, mas também para o aumento da precisão interpretativa, favorecendo a consistência e a clareza nas práticas institucionais.

Adicionalmente, conforme discutido por [de Brito 2021], a IA pode desempenhar papel relevante na geração automatizada de documentos, otimizando tarefas que tradicionalmente demandariam significativa intervenção humana. Tal capacidade mostra-se especialmente pertinente em ambientes acadêmicos, nos quais a elaboração de planos de disciplina e demais documentos institucionais pode ser aprimorada por meio do uso de ferramentas baseadas em IA. Essas ferramentas não apenas aceleram a produção de conteúdo, mas também asseguram maior aderência às normativas vigentes, minimizando inconsistências e erros.

Diversos *chatbots* e assistentes virtuais já são empregados no ensino superior, como o **IBM Watson Assistant** e o **Microsoft Azure Bot Service**. Embora ofereçam integração com múltiplos canais e suporte multilíngue, essas soluções muitas vezes dependem de configurações genéricas, limitando sua adaptação a contextos institucionais específicos. Frequentemente, não incorporam dados normativos da instituição, resultando em respostas insuficientes para demandas administrativas ou acadêmicas complexas.

Neste trabalho, propõe-se o desenvolvimento de um *chatbot* customizado para o IFPI, treinado com resoluções, normativas e manuais operacionais da instituição. Essa abordagem permite oferecer respostas contextualizadas e automatizar atendimentos recorrentes, aliviando a carga sobre os servidores. Os principais desafios incluem a manutenção da atualização da base de dados e a validação constante do modelo, a fim de evitar respostas incorretas que comprometam decisões institucionais.

O uso de IA generativa na análise de textos acadêmicos e legais permite integrar tecnologias avançadas aos processos institucionais, promovendo eficiência, transparência e conformidade. Modelos como GPT e Gemini não apenas representam avanços técnicos, mas configuram-se como ferramentas estratégicas para a gestão acadêmica. Contudo, sua personalização para normativas internas, como a proposta neste trabalho, ainda é pouco explorada na literatura, o que evidencia a originalidade e relevância da presente iniciativa.

3. Metodologia

Nesta seção, são apresentadas as abordagens adotadas no projeto IFPIA, com o objetivo de descrever detalhadamente a arquitetura da ferramenta e os processos envolvidos para a sua futura implementação.

3.1. Visão Geral da Arquitetura

A arquitetura proposta para o *chatbot* do IFPI adota uma abordagem baseada em orquestração de IA, com o objetivo de assegurar o processamento eficiente e a adequada contextualização das interações. No campo da IA, a orquestração refere-se à coordenação inteligente entre diferentes modelos e componentes, com vistas à otimização da geração

de respostas. Essa estrutura permite que o sistema direcione cada entrada do usuário ao modelo mais apropriado, considerando o contexto e a natureza da solicitação realizada.

A Figura 1 apresenta uma representação esquemática dessa arquitetura, na qual o sistema é constituído por múltiplos modelos de linguagem (LLM^1 , LLM^2 , ..., LLM), cada um treinado com dados específicos e especializado em um domínio particular (D^1 , D^2 , ..., D). Quando uma consulta é recebida, o orquestrador atua como um intermediário inteligente, responsável por analisar semanticamente a entrada e selecionar o modelo mais relevante para a construção de uma resposta precisa. Esse mecanismo garante que o *chatbot* forneça respostas alinhadas ao domínio correspondente, minimizando ambiguidades e elevando a qualidade da experiência do usuário. Além disso, a modularidade inerente à abordagem adotada viabiliza a aplicação de técnicas de *fine-tuning* em modelos específicos, sem comprometer o funcionamento do sistema como um todo.

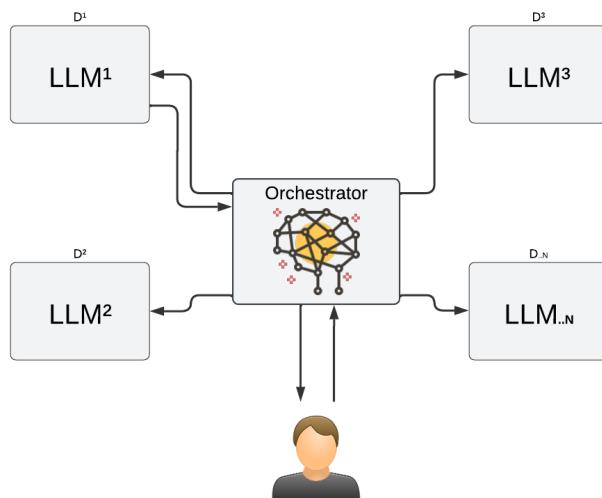


Figura 1. Arquitetura de Orquestração de IA

A estratégia de orquestração de IA no *chatbot* do IFPI oferece diversas vantagens operacionais e estruturais, tais como:

- **Precisão e contextualização:** A análise automatizada da entrada do usuário assegura que a resposta seja gerada pelo modelo mais apropriado, resultando em maior precisão e relevância das informações fornecidas.
- **Eficiência e escalabilidade:** O sistema pode ser expandido com a adição de novos modelos especializados, acompanhando o crescimento da base de conhecimento institucional sem prejuízo ao desempenho global.
- **Modularidade e flexibilidade:** A separação entre o orquestrador e os modelos individuais permite atualizações e ajustes localizados em cada domínio, sem impactar a operação geral do *chatbot*.

Na implementação da solução institucional, emprega-se o modelo de linguagem *LLaMA 3*, orquestrado por meio do framework *LangChain*. Essa integração combina o processamento de NLP com mecanismos de recuperação de informações baseados em normativas, resoluções e demais documentos institucionais. O principal objetivo é fornecer respostas contextualizadas e precisas a alunos e servidores, contribuindo para a otimização da gestão acadêmica e administrativa.

O armazenamento das representações vetoriais é realizado por meio do banco de dados *ChromaDB*. Esse sistema distingue-se, em relação a outros bancos vetoriais, por sua eficiência no armazenamento de dados, gratuidade e adequação para a prototipagem em conjuntos de dados de pequena e média escala, características que o tornam particularmente apropriado para as etapas iniciais de desenvolvimento do *chatbot* institucional [Mathur and Chhabra 2024]. O *ChromaDB* é especialmente eficaz na armazenagem e consulta de vetores extraídos de documentos normativos, apresentando desempenho elevado nos processos de recuperação semântica da informação.

3.2. Treinamento e Personalização

O processo de treinamento e personalização do *chatbot* é fundamentado em duas abordagens complementares: *Fine-tuning* e *Retrieval-Augmented Generation* (RAG).

Fine-tuning: Esta técnica consiste na adaptação de modelos pré-treinados por meio do ajuste de seus pesos, permitindo que o modelo seja refinado de maneira eficiente para atender às necessidades específicas do IFPI. Ao explorar o conhecimento previamente incorporado ao modelo, o *fine-tuning* reduz de forma significativa o tempo e os custos associados ao treinamento, além de aprimorar o desempenho do sistema na resolução de questões particulares [Saad and Qawaqneh 2024].

Retrieval-Augmented Generation (RAG): Esta abordagem integra a recuperação de documentos relevantes com a geração textual baseada em modelos de linguagem. No contexto do IFPI, o RAG permite a incorporação de informações específicas de domínio — extraídas de normativas, diretrizes e editais institucionais — ao processo de geração de respostas pelo modelo *LLAMA 3*. Tal integração promove a produção de respostas contextualizadas e atualizadas, contribuindo para maior explicabilidade e precisão nas interações com os usuários. Ressalta-se, entretanto, que o desempenho dessa abordagem depende diretamente da robustez do sistema de recuperação de informações subjacente.

A base de dados utilizada para o *fine-tuning* contempla um amplo conjunto de documentos institucionais, como normativas, diretrizes e publicações oficiais do IFPI, o que assegura ao modelo um conhecimento sólido e contextualizado do ambiente institucional. Além disso, está prevista a atualização contínua do modelo com base no *feedback* dos usuários e nas alterações periódicas nos documentos oficiais. Esse ciclo de aprimoramento incremental visa garantir a progressiva melhoria da precisão, relevância e personalização das respostas fornecidas, promovendo uma experiência mais fluida e eficaz para os usuários do sistema.

4. Resultados Esperados e Discussões

Até o momento, os principais resultados concentram-se na definição e implementação das tecnologias base do *chatbot*. A adoção do *LangChain* permitiu construir fluxos estruturados para orquestração das consultas, enquanto o uso do *ChromaDB* viabilizou a indexação eficiente de documentos institucionais por meio de *embeddings* semânticos, melhorando a precisão das respostas. A integração do modelo *LLAMA 3* mostrou potencial na interpretação contextualizada, embora o projeto ainda esteja em desenvolvimento e não configurado como produto final. Testes iniciais já são possíveis, mas são necessárias otimizações para desempenho e escalabilidade.

O IFPI registra mensalmente centenas de atendimentos relacionados a normativas e processos acadêmicos, com tempo médio de resposta manual entre 24 e 72 horas, evidenciando a necessidade de soluções automatizadas para agilizar o suporte. Para avançar, destacam-se as seguintes ações futuras: refinamento contínuo do modelo com avaliação via métricas objetivas (precisão, *recall*, F1), aprimoramento da infraestrutura para suportar alta demanda simultânea, desenvolvimento de interface amigável, incorporação sistemática de *feedback* dos usuários, atualização constante da base de conhecimento e foco na experiência do usuário.

Embora as tecnologias empregadas não sejam inéditas, a personalização para o contexto do IFPI é diferencial, mas o projeto carece ainda de validação experimental robusta e métricas consolidadas para comprovar sua efetividade e justificar adoção ampla. Em resumo, os resultados iniciais são promissores, porém o trabalho está em fase inicial e requer validações e aprimoramentos para consolidar uma solução escalável e relevante para a gestão acadêmica e administrativa do IFPI.

5. Conclusão

O desenvolvimento do *chatbot* institucional para o IFPI representa um avanço significativo na modernização dos processos de comunicação e no acesso a informações acadêmicas e administrativas. A adoção de tecnologias como *LangChain*, *ChromaDB* e o modelo *LLaMA 3* estabeleceu uma base robusta para a recuperação eficiente de documentos institucionais e para a geração de respostas contextuais e precisas. A abordagem baseada em RAG mostrou-se promissora ao promover maior aderência às normativas da instituição, por meio da contextualização dinâmica das respostas. Embora ainda em fase inicial, os resultados preliminares indicam potencial para facilitar o acesso à informação e otimizar processos internos, reduzindo a carga administrativa dos servidores e aumentando a autonomia dos usuários finais. No entanto, desafios como escalabilidade, falta de validação empírica robusta, métricas de desempenho e a necessidade de uma interface acessível permanecem.

Referências

- de Brito, L. (2021). Zuboff, shoshana. the age of surveillance capitalism: The fight for a human future at the new frontier of power. new york: Publicaffairs, 2019, 691p. isbn 9781610395694. *Mural Internacional*, 12:e55150.
- et al., O. (2024). Gpt-4 technical report.
- Mathur, S. and Chhabra, A. (2024). Vector search algorithms: A brief survey. In *2024 4th International Conference on Ubiquitous Computing and Intelligent Information Systems (ICUIS)*, pages 365–371.
- Saad, M. and Qawaqneh, Z. (2024). Closed domain question-answering techniques in an institutional chatbot. In *2024 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCMEE)*, pages 1–8.
- Susskind, R. (2019). *Online Courts and the Future of Justice*. Oxford University Press.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need.