

# Integração entre modelos de linguagem e chatbot para o Departamento de Engenharia Elétrica

Ian A. Mendes e Rodolpho V. A. Neves

Departamento de Engenharia Elétrica,

Universidade Federal de Viçosa, Viçosa - MG Email: {ian.mendes, rodolpho.neves}@ufv.br

**Abstract**—Este trabalho aborda a integração entre modelos de linguagem e chatbots para Departamento de Engenharia Elétrica (DEL), explorando como a inteligência artificial (IA) está impulsionando o mercado de informações. A utilização de chatbots inteligentes capacitados por IA permite uma interação eficiente e personalizada com os usuários, melhorando a qualidade do atendimento e a satisfação do cliente. Este trabalho tem como principal objetivo criar uma interface funcional entre chatbot e usuário, capaz de fornecer respostas sobre documentos fornecidos, através do uso de IA.

**Index Terms**—Modelos de linguagem, Processamento de Linguagem Natural (PNL), Inteligência Artificial, Chatbot, Comunicação Automatizada, Departamento de Engenharia Elétrica.

## I. INTRODUÇÃO

Com base na compreensão do mercado e do setor de informações, é evidente que o aumento do uso de novas tecnologias, como a inteligência artificial (IA), está impulsionando aprimoramentos significativos no desempenho e na velocidade do atendimento às necessidades do público-alvo. Neste sentido, são necessários alguns esclarecimentos conceituais que estão presentes nesta introdução.

Deve-se entender, inicialmente, o que é mercado. Segundo [1], mercado pode ser definido como o conjunto de todos os compradores, efetivos e potenciais, de uma oferta de mercado e detalhou ao definir mercado potencial como o conjunto de consumidores que possuem renda, interesse suficiente por uma oferta ao mercado e acrescentou que o mercado disponível é o conjunto de consumidores que possuem renda, interesse, acesso e uma determinada oferta. Logo, percebe-se que a ideia de mercado engloba tanto o cenário quanto os atores, como os processos de determinado setor.

No que diz respeito ao mercado da informação, sua natureza está em evolução contínua, e todas as organizações da indústria da informação precisam formar novas alianças estratégicas, identificar novos segmentos de mercado, desenvolver novos produtos e novos relacionamentos com seus públicos, empregando vasto arsenal de estratégias, táticas, técnicas e instrumentos de marketing para obter sucesso no ambiente de mudança de paradigmas na sociedade [2].

O crescimento exponencial da tecnologia e da digitalização tem impulsionado o mercado da informação, permitindo o acesso e a troca de informações de forma rápida e global. Com a crescente disponibilidade de dados

e informações, empresas e indivíduos buscam formas de capturar, organizar, analisar e utilizar essas informações para tomar decisões mais informadas e obter vantagens competitivas.

A informação é uma mercadoria simbólica, de características muito especiais. É uma mercadoria que não se esgota com o consumo, como uma maçã ou um copo de leite. Ao ser consumida, a mercadoria ainda permanecerá consumível por um tempo e espaço, que será determinado por sua qualidade e validade [3]. Podemos extrair que a informação, como mercadoria, é abundante, mas por ser abundante, só possui valor para o consumidor que a considere potencialmente útil. A precificação da informação é imprecisa por não estar associada a uma base fixa e por não ser homogênea. Seu preço, quando o possui, pouco tem a ver com o seu custo, e os dois, preço e custo, não se relacionam com o valor como uma mercadoria tradicional, e muito menos com as condições de oferta e demanda que determinariam o mercado de informação [3].

De acordo com [4], o mercado de informação possui características peculiares, pois é a oferta que determina a demanda por informação, ao afirmar que a posição parece indicar que o homem da informação é substancialmente diferente do homem econômico. Sem dúvida, ele vive em um mundo onde oferta pode criar demanda.

No meio deste cenário, percebe-se que os consumidores estão cada vez mais exigentes, seletivos e têm um alto grau de expectativa em relação ao atendimento. Atualmente, com a globalização, os clientes sabem exatamente o que querem e buscam cada vez mais serviços mais rápidos e atendimento de alta qualidade.

Conforme a Figura 1, podemos observar o aumento crescente no faturamento do e-commerce no Brasil ao longo dos anos. "E-commerce" ou comércio eletrônico é o termo utilizado para englobar todas as atividades comerciais realizadas por meios eletrônicos, principalmente pela internet, incluindo a compra e venda de produtos e serviços, transações financeiras, marketing, atendimento ao cliente e outras atividades relacionadas.

Temos que o atendimento ao cliente é um setor fundamental em diversas indústrias e segmentos, pois desempenha um papel essencial na interação entre as empresas e seus clientes. Esse mercado engloba uma ampla variedade de serviços e soluções que visam garantir uma experiência positiva ao cliente, atendendo suas necessidades, solucio-



Fig. 1. Faturamento e-commerce no Brasil. Fonte: (EBIT, 2017)

nando problemas e fornecendo suporte.

O tamanho do mercado de atendimento ao cliente é significativo e continua a crescer à medida que as empresas reconhecem a importância de investir em um atendimento eficiente e de qualidade. De acordo com relatórios e estatísticas, o mercado global de serviços de atendimento ao cliente foi avaliado em vários bilhões de dólares em 2020 e é projetado para crescer a uma taxa significativa nos próximos anos.

A Figura 2 ilustra a evolução do faturamento do e-commerce no Brasil ao longo do tempo e revela uma tendência ascendente, refletindo o contínuo crescimento desse setor. Esse gráfico destaca um aumento consistente nas receitas do comércio eletrônico, indicando a preferência crescente dos consumidores por compras online, tanto de produtos quanto de informações. É importante ressaltar que esse mercado ainda é relativamente inexplorado, por isso as projeções são altas para o futuro.

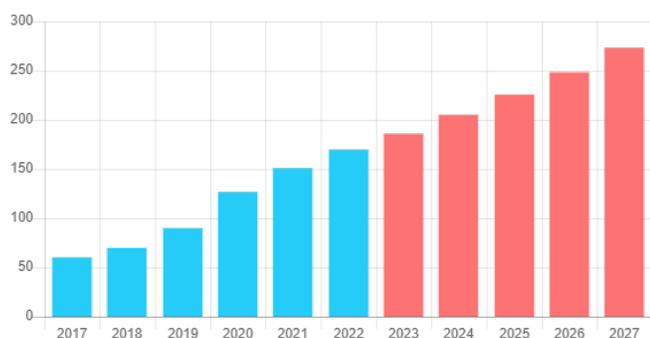


Fig. 2. Previsão de vendas no E-commerce para os Próximos 5 anos com faturamento em bilhões. Fonte: (ABCOMM, 2022)

O avanço na aplicação de novas tecnologias no dia a dia revolucionou a comunicação e a interação entre humanos e computadores, assim como a comunicação entre pessoas ao redor do mundo, devido ao fácil acesso a aplicativos de tradução. A forma pela qual os seres humanos se comunicam pode ser chamada de linguagem natural, que inclui fala, texto e emoções [8].

Os seres humanos estão mais envolvidos em falar como meio de comunicação do que em escrever (texto). O texto, em vez da fala, é importante para o desenvolvimento de qualquer aplicativo de IA que facilite o processo de

comunicação, uma vez que as máquinas, até o momento, não são tão eficientes em aprender idiomas quanto os humanos, e, como resultado, elas têm que depender do texto/dados [9].

O estudo da linguagem, no que diz respeito à sua gramática, regras, semântica e fonética, é chamado de linguística. Criar regras de linguagem, métodos de sintaxe e semântica estão incluídos como parte do estudo da linguística. Os métodos e regras propostos pelos linguistas teóricos podem ser processados por sistemas computacionais, que geram linguagem natural que lida com questões como gramática, normas semânticas, etc. Essa abordagem é conhecida como linguística computacional, na qual abordagens estatísticas são utilizadas no processo de análise de texto/dados [10]. Contudo, a abordagem para o processamento da linguagem natural pode não se limitar à estatística, pois também pode envolver a aplicação de métodos avançados de inferência, como machine learning (ML) e aprendizado profundo, que são técnicas populares de inteligência artificial (IA).

O aprendizado profundo tem impulsionado a atividade em diferentes subcampos da IA, como Visão Computacional ou processamento de linguagem natural (PLN), resultando em produtos comerciais sofisticados para aplicações como reconhecimento facial ou geração automatizada de linguagem.

Tecnicamente, um modelo de linguagem é uma representação estatística de uma linguagem, que nos diz a probabilidade de uma determinada sequência (uma palavra, frase ou sentença) ocorrer nessa linguagem. Devido a essa capacidade, modelos de linguagem podem ser usados para fazer previsões sobre como uma sentença pode continuar e, conseqüentemente, gerar texto. Modelos de linguagem sofisticados, frequentemente baseados em redes neurais e grandes corpora de texto, são muito poderosos, pois podem ser usados em uma ampla variedade de aplicações diferentes, como tradução ou reconhecimento de texto [11].

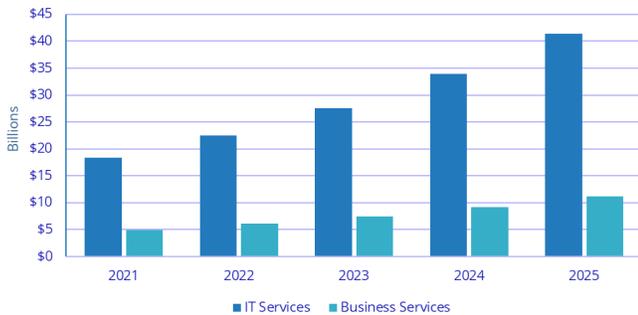
Um chatbot é um exemplo típico de um sistema de IA e um dos exemplos mais básicos e difundidos de Interação Humano-Computador (IHC) inteligente [12]. O chatbot permite a interação com o usuário de maneira inteligente e personalizada, através de técnica de PLN. O chatbot compreende e interpreta as mensagens fornecidas pelo usuário, envolvendo a análise do texto fornecido para identificar palavras-chave, entidades, intenções e contexto. De acordo com o texto do usuário, o chatbot pode recuperar informações para encontrar respostas relevantes ou ações a serem executadas. Os chatbots também permitem integrações com sistemas externos, por exemplo, banco de dados, concedendo acesso a grandes quantidades de dados ao chatbot, proporcionando respostas mais relevantes e coerentes aos usuários.

Os dados apresentados na Figura 3 revelam um crescimento significativo nos gastos globais com serviços de Tecnologia da Informação (TI) relacionados à IA, assim como no desenvolvimento de modelos de negócios baseados

em IA. Esses números demonstram que a IA continua sendo um impulsionador chave dos investimentos em TI.



Worldwide AI Services Forecast by Technology  
Category Detail, 2021-2025



Source: IDC 2022

Fig. 3. Previsão mundial de serviços de IA por categoria de tecnologia. Fonte: (IDC, 2021).

Existem várias empresas que fazem uso atualmente de chatbots para melhorar seus processos de atendimento ao cliente, fornecer suporte automatizado e facilitar interações com os usuários. Temos como exemplo a Amazon, que utiliza chatbots em sua plataforma de atendimento ao cliente para responder perguntas sobre pedidos, fornecer informações sobre produtos, processar devoluções e fornecer suporte técnico. A Sephora, uma empresa de cosméticos, também utiliza chatbots em seu aplicativo para ajudar os usuários a encontrar produtos, obter dicas de maquiagem, fazer agendamentos para serviços em lojas físicas e responder perguntas sobre produtos.

Contudo, pode-se perceber uma defasagem no uso de chatbots em algumas áreas que possuem grande potencial de evolução através dessa tecnologia. Na área da educação, por exemplo, chatbots podem ser empregados para o suporte ao aluno, respondendo perguntas sobre matérias e fornecendo informações relevantes, auxiliando o aluno na busca por recursos educacionais, oferecendo dicas de estudo e fornecendo feedback sobre o desempenho acadêmico. Da mesma forma, na orientação para a escolha de carreira, podem fornecer informações sobre diferentes áreas profissionais, habilidades necessárias, oportunidades de trabalho e dicas para o desenvolvimento de carreira.

O objetivo geral deste trabalho é o desenvolvimento de uma integração entre um chatbot e um banco de dados, através do uso de LLM's, que melhore a comunicação e suporte ao aluno, permitindo a este uma maior facilidade de acesso a informações sobre o Departamento de Engenharia Elétrica (DEL) da Universidade Federal de Viçosa (UFV). Como objetivos específicos, será feita a implementação da interface do chatbot, integração entre modelos de linguagem e o chatbot, criação de um banco

de dados abrangente e atualizado, para gerar respostas mais precisas e coerentes, configuração das respostas do chatbot, teste e otimização do sistema e documentação do processo de desenvolvimento.

## II. REVISÃO DE LITERATURA

De acordo com [13], o processamento de linguagem natural (PLN ou NLP, na sigla em inglês) é um conjunto de técnicas computacionais teoricamente motivadas para analisar e representar textos de ocorrência natural em um ou mais níveis de análise linguística, com o objetivo de alcançar um processamento de linguagem semelhante ao humano para uma variedade de tarefas ou aplicações. Em suma, temos que o campo de PLN está preocupado em criar algoritmos e sistemas computacionais para processar e analisar dados de linguagem natural, sendo seus pilares os pipelines de processamento de texto que realizam análises linguísticas de rotina em grandes quantidades de texto.

A modelagem estatística da linguagem (LM), em sua forma mais genérica, é a tarefa de capturar regularidades da linguagem natural em um modelo probabilístico [14]. Na maioria dos casos, no entanto, a modelagem da linguagem é entendida como a tarefa de estimar a probabilidade de uma sequência de palavras. A história da LM remonta a Markov [15] e Shannon [16], que usaram n-gram para prever o próximo caractere ou palavra, respectivamente, em textos em linguagem natural. N-gram é um dos modelos mais simples que atribui probabilidades a frases e sequências de palavras, geralmente a palavra modelo é omitida e somente o termo n-gram é utilizado para significar tanto a sequência de palavras em si quanto o modelo preditivo que lhe atribui uma probabilidade.

A partir dos conceitos de PLN, LM e n-gram é possível explicar o que são Grandes Modelos de Linguagem (GML ou LLM, na sigla em inglês). LLMs são projetados para entender e gerar linguagem humana, utilizando técnicas como aprendizado profundo (deep learning), especificamente um tipo de rede neural chamada de transformer. Eles são treinados em grandes quantidades de dados de texto provenientes de diversas fontes, incluindo livros, artigos, sites e outros materiais escritos. Esse treinamento extensivo permite que os LLMs aprendam os padrões estatísticos e estruturas da linguagem, possibilitando a geração de textos coerentes e contextualmente relevantes com base em uma entrada ou instrução fornecida.

Para a execução deste trabalho, foi utilizado o LangChain, sendo este um framework de desenvolvimento de aplicativos baseados em modelos de linguagem, oferecendo a capacidade de criar aplicativos que são conscientes dos dados, conectando modelos de linguagem a outras fontes de informação. Além disso, a LangChain permite que os modelos de linguagem ajam como agentes, interagindo com o ambiente em que estão inseridos. Os principais benefícios desse framework incluem o uso de componentes modulares, que fornecem abstrações para trabalhar

com modelos de linguagem, e a disponibilidade de cadeias prontas para uso, que são estruturas pré-configuradas de componentes para realizar tarefas específicas de alto nível. Essas cadeias prontas para uso simplificam o desenvolvimento inicial, enquanto os componentes permitem a personalização das cadeias existentes ou a criação de novas para casos de uso mais complexos e específicos.

Os embeddings de texto representam entradas discretas de texto (frases, documentos e código) como vetores de tamanho fixo que podem ser utilizados em várias tarefas subsequentes. Essas tarefas incluem similaridade textual semântica [17], recuperação de informações [18], avaliação automática de texto [19], recuperação de prompts para aprendizado em contexto [20], entre outros.

### III. MATERIAIS E MÉTODOS

Inicialmente, empregaram-se quatro modelos de linguagem no estudo. O primeiro, denominado LLaMA2 (Large Language Model Meta AI), é um modelo desenvolvido pelo Meta. Duas versões desse modelo foram utilizadas, treinadas com sete e 13 bilhões de parâmetros, visando analisar a influência do número de parâmetros na qualidade das respostas obtidas. Além disso, incorporou-se o BLOOM LLM, um modelo autoregressivo projetado para continuar o texto a partir de um prompt, sendo treinado extensivamente com vastas quantidades de dados de texto e fazendo uso de recursos computacionais em escala industrial [20]. Outra contribuição foi proveniente do WizardLM, um modelo de linguagem pré-treinado com sete bilhões de parâmetros. O código do sistema foi desenvolvido utilizando as linguagens Python, JavaScript e HTML.

Para aprimorar a integração entre o Python e os modelos de linguagem, foi utilizado o framework LangChain. Essa ferramenta permitiu a criação de abstrações e implementações modulares para trabalhar com os modelos de linguagem selecionados. Além disso, os embedders da OpenAI e um embedder chamado Instructor foram empregados para a geração de representações vetoriais dos textos.

O FAISS (Facebook AI Similarity Search), caracterizado como uma biblioteca que permite buscar rapidamente embeddings de documentos multimídia que são semelhantes entre si, foi escolhido para armazenar os embeddings. Além disso, outros vector stores, como o Chroma e o Pinecone, também foram considerados durante a seleção dos métodos de armazenamento e busca semântica dos dados.

Utilizou-se o servidor fornecido pelo Google Colab para executar o sistema desenvolvido em um ambiente de execução com uma GPU V100 de alta capacidade de RAM, utilizando o Ngrok como host para hospedar este trabalho. Essas plataformas forneceram os recursos necessários para processar eficientemente as consultas dos usuários.

Inicialmente, os documentos de texto fornecidos pelo administrador são divididos em chunks, segmentos menores de texto. Esses chunks foram divididos em 1000 caracteres

por chunk com uma sobreposição de 500 caracteres a fim de evitar perda de dados. Este método de divisão é utilizado para facilitar a análise e o armazenamento dos embeddings. Posteriormente, os embeddings foram gerados utilizando os embedders selecionados e armazenados no vector store mencionado.

Para uma melhor interação entre o usuário e o sistema, foi feito um código em HTML e JavaScript para gerar um site com aspecto semelhante ao site do departamento, mas com um botão no canto inferior direito, onde o usuário pode interagir com o chatbot.

Quando o usuário faz uma pergunta ao chatbot, o sistema realiza o embedding da pergunta e efetua uma busca semântica entre a pergunta e os dados armazenados no vector store. Os resultados são classificados e enviados para o modelo de linguagem selecionado, que gera a resposta mais adequada a ser retornada. Também é oferecida ao usuário a opção de fornecer um feedback; ao reagir à resposta do chatbot, tanto a pergunta quanto a resposta são armazenadas em um arquivo no Excel. Esse arquivo foi utilizado para medir a qualidade das respostas dos modelos de linguagem e poderá ser utilizado em trabalhos futuros para o fine-tuning dos modelos.

Esses métodos estão representados na Figura 4, na qual as setas de cor preta representam os processos executados uma única vez, e as setas de cor azul são executadas de acordo com o número de perguntas do usuário.

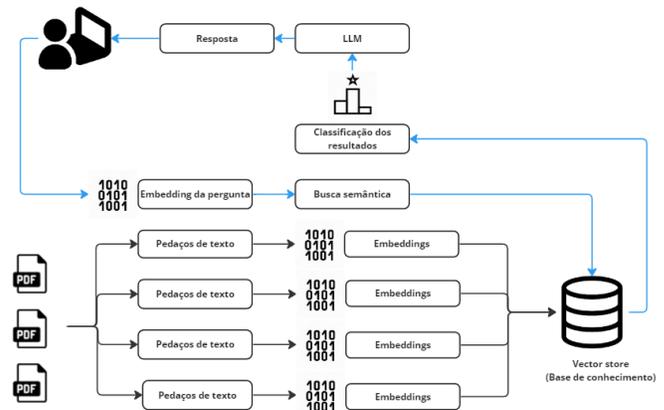


Fig. 4. Diagrama de fluxo dos processos realizados.

Com o software em pleno funcionamento, usuários tiveram a oportunidade de fazer perguntas e fornecer o feedback necessário para comparar os modelos de linguagem em relação à sua precisão e qualidade nas respostas.

### IV. RESULTADOS

Foi produzida uma interface funcional entre chatbot e usuário, representada na Figura 5. Nessa interface, foi utilizada uma réplica do site do departamento, com a adição de um botão no canto direito inferior para acessar o chatbot. O chatbot respondeu às perguntas de acordo

com os documentos fornecidos, cujo conteúdo contém informações sobre o Departamento de Engenharia Elétrica (DEL) e informações sobre seus docentes.

Abaixo, seguem as diferentes respostas obtidas utilizando os modelos de linguagem fornecidos.

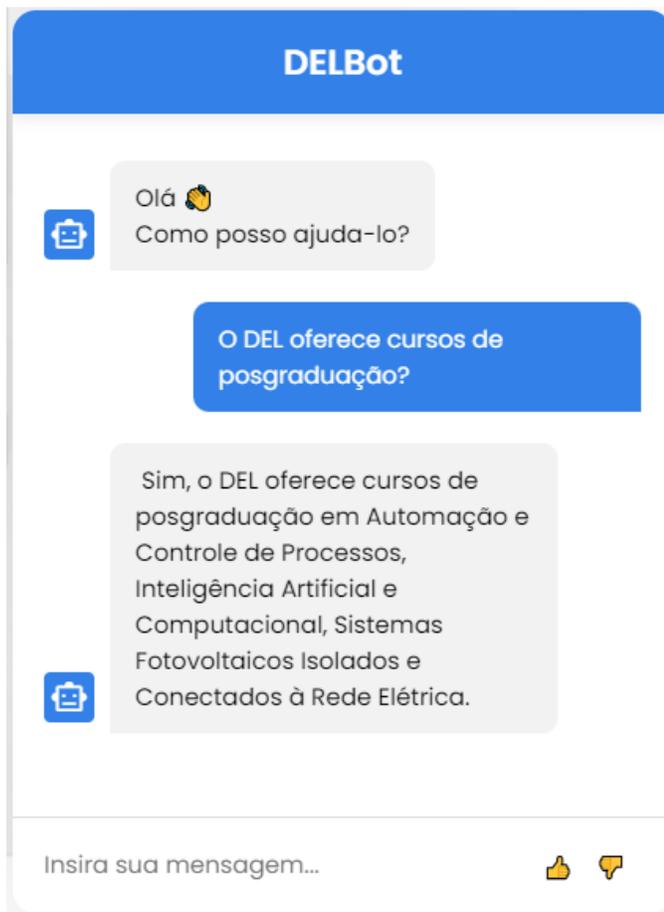


Fig. 5. Resposta obtida utilizando o modelo LLaMA-7b.

Na Figura 5, é apresentada a resposta obtida pelo modelo LLaMA-7b. Ao compará-la com as respostas dos outros modelos que possuem a mesma quantidade de parâmetros, observa-se que esta resposta se destacou positivamente. Nos outros modelos, ocorreram alucinações ou divisões incorretas dos cursos. Por exemplo, o curso de Automação e Controle foi erroneamente dividido em dois cursos distintos, um de Automação e outro de Controle.

Na Figura 6, observamos a resposta produzida pelo modelo WizardLM, a qual foi semelhante à resposta do modelo LLaMA-7b. No entanto, a resposta do modelo BLOOM indicou que a informação era tratada como privada pela universidade e sugeriu entrar em contato para obter mais informações, sendo essa avaliada como uma resposta inadequada.

A resposta obtida, conforme apresentado na Figura 7, foi semelhante em todos os modelos, resultando em avaliações positivas para cada uma delas.

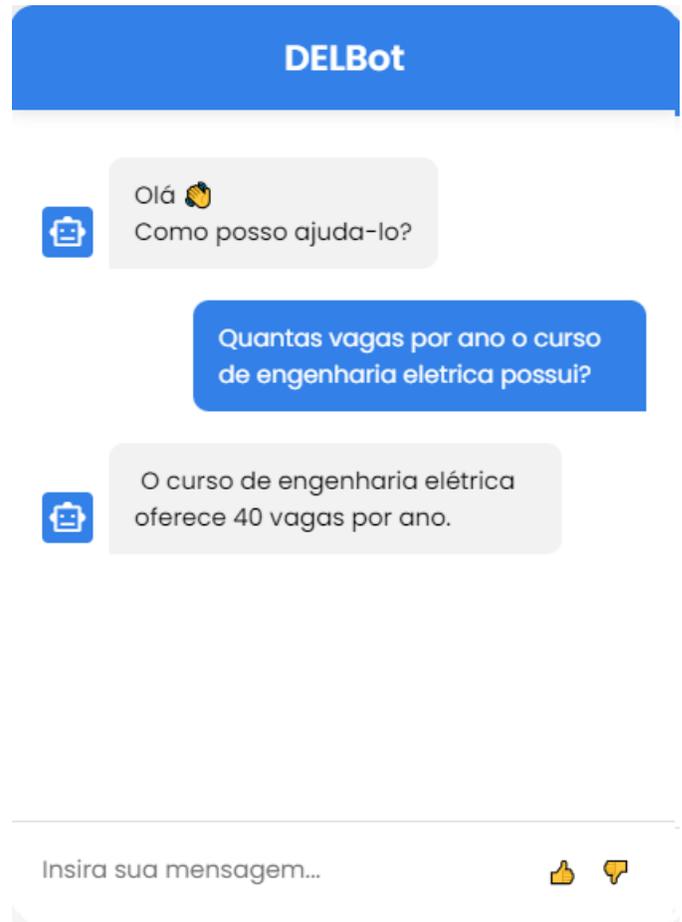


Fig. 6. Resposta obtida utilizando o modelo WizardLM.

É possível observar, na Figura 8, que o aumento no número de parâmetros resultou em uma melhoria na qualidade das respostas. Em contrapartida, os modelos treinados com menos parâmetros não foram capazes de responder a essa pergunta, apresentando todos eles o fenômeno de alucinação.

É relevante destacar que, embora os modelos de sete bilhões de parâmetros nem sempre apresentem respostas semelhantes inicialmente, ao aprimorar as perguntas, foi possível obter respostas corretas na maioria dos casos.

A partir da análise do feedback fornecido usuários, obteve-se os dados presentes na Figura 9.

A análise das interações entre os usuários e o chatbot revelou discrepâncias notáveis. Algumas perguntas foram respondidas em inglês, apesar de estarem corretas, o que gerou divergências entre os estudantes. Houve tanto apoio quanto objeção a essa abordagem, indicando que a escolha de um modelo de representação multilíngue, em detrimento de um modelo exclusivamente em português, pode ter influenciado esse comportamento. Além disso, a observação de que o chatbot apresentou respostas em inglês em um modelo com menos parâmetros, enquanto respondeu em português em um modelo com mais parâmetros,

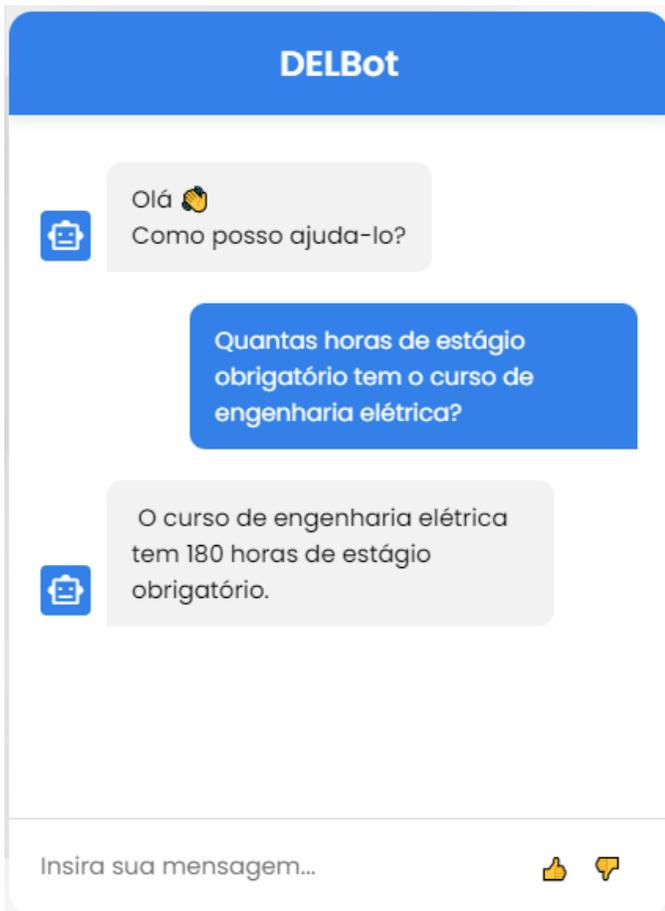


Fig. 7. Resposta obtida utilizando o modelo BLOOM.

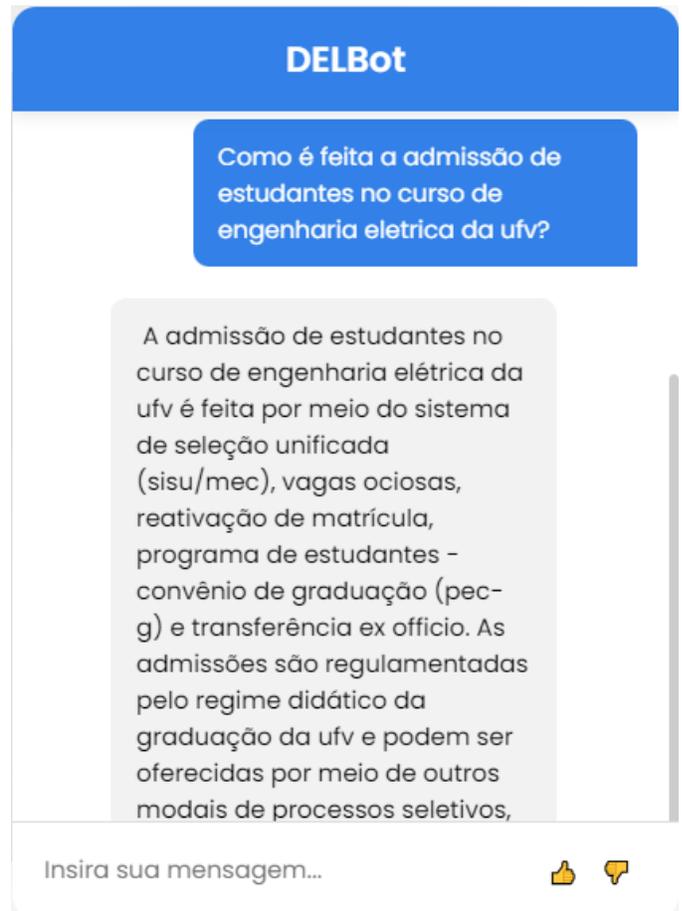


Fig. 8. Resposta obtida utilizando o modelo LLaMA-13b.

destaca a possível influência da complexidade do modelo nesse cenário.

Outro aspecto relevante para avaliar os resultados dos modelos de linguagem foi a qualidade das perguntas formuladas pelos usuários. Quando as perguntas eram amplas, o chatbot enfrentava dificuldades em gerar respostas satisfatórias. No entanto, ao reformular as perguntas de forma mais específica, observou-se uma melhora significativa na capacidade do chatbot em fornecer respostas mais precisas e adequadas. Essa constatação ressalta a importância de considerar não apenas o modelo em si, mas também a formulação precisa das perguntas ao avaliar o desempenho do chatbot.

É fundamental destacar algumas alucinações geradas pelos modelos de linguagem, um fenômeno caracterizado quando o modelo gera um texto que é incorreto, sem sentido ou irreal. Essas alucinações podem ser atribuídas a uma compreensão contextual limitada, uma vez que os modelos são compelidos a transformar o prompt e os dados de treinamento em uma abstração na qual algumas informações podem ser perdidas. Além disso, devido à utilização de modelos de linguagem pré-treinados, o ruído nos dados de treinamento pode introduzir um padrão es-

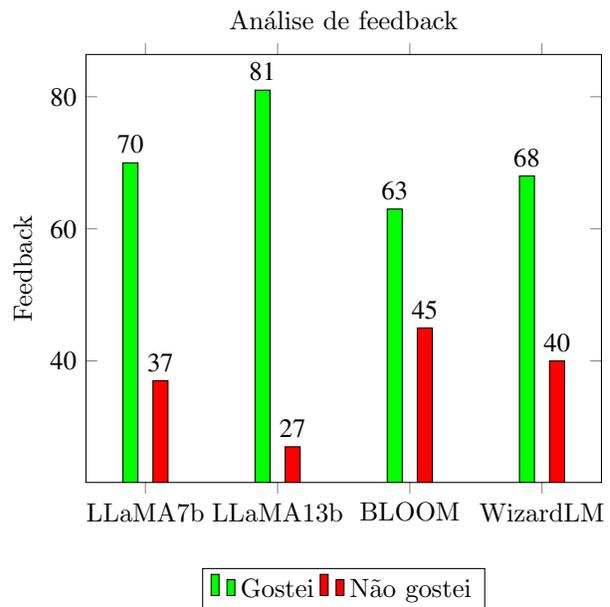


Fig. 9. Distribuição de avaliações entre os modelos de linguagem.

tatístico distorcido, levando o modelo a reagir de maneira

inesperada em algumas situações.

## V. CONCLUSÃO

A análise do feedback fornecido pelos usuários desempenhou um papel fundamental na classificação dos modelos. Diferentes perspectivas, opiniões e expectativas, foram observadas, permitindo identificar padrões e desafios específicos. Este processo forneceu insights valiosos sobre a eficácia, qualidade e a usabilidade dos modelos de linguagem integrados.

Com base na análise detalhada do feedback, foi possível classificar e comparar os modelos de linguagem utilizados. O modelo LLaMA-13b apresentou o melhor desempenho entre todos os modelos e o LLaMA-7b se mostrou o melhor modelo ao se comparar os modelos treinados com o mesmo número de parâmetros. Foi possível identificar pontos fortes e áreas de melhoria para cada modelo. Essa classificação proporciona uma base sólida para recomendações futuras e otimizações, visando aprimorar continuamente a experiência dos usuários do chatbot.

Este trabalho não apenas resultou na implementação prática em pequena escala de uma ferramenta para a comunidade acadêmica, mas também ofereceu contribuições significativas para o entendimento do impacto e da eficácia de diferentes modelos de linguagem em ambientes educacionais. Espera-se que os resultados e as lições aprendidas neste trabalho, forneçam uma base para pesquisas futuras e iniciativas de melhoria na integração de tecnologias de chatbot e modelos de linguagem em contextos educacionais específicos.

### A. Recomendações para trabalhos futuros

Com base nas conclusões obtidas neste trabalho, existem sugestões para orientar investigações futuras nesse campo. Essas recomendações buscam aprimorar ainda mais a eficácia do chatbot, considerando as nuances identificadas durante a pesquisa.

É recomendável explorar modelos de linguagem mais avançados, caracterizados por uma quantidade significativamente maior de parâmetros. Tal abordagem permite compreender como a complexidade do modelo influencia não apenas a capacidade de resposta, mas também a qualidade das interações.

Além disso, é aconselhável realizar o fine-tuning dos modelos como uma estratégia para adaptar os modelos às características específicas do ambiente acadêmico do Departamento de Engenharia Elétrica.

Por fim, sugere-se investigar o uso de modelos de representação exclusivamente com dados em português e provenientes de uma quantidade expressiva de informações. Essa abordagem pode contribuir para a melhoria da relevância e adequação das respostas geradas, alinhando-as de maneira mais precisa às necessidades do público-alvo.

## VI. AGRADECIMENTOS

Agradeço a Deus pela minha vida e por ser meu amparo em toda a minha jornada durante a graduação.

À minha família, meus pais Izabel e José Anilson, e meu irmão Dan, pelo incentivo nos momentos difíceis, pela confiança em mim e pela compreensão da minha ausência ao longo do curso.

À minha companheira, Jessica, por estar sempre ao meu lado e por acreditar constantemente no meu potencial.

Aos meus professores, em especial ao Rodolpho e ao Mauro, pelas correções, ensinamentos e por mostrarem que sempre há um caminho, e com determinação, é possível superar qualquer barreira.

## VII. REFERÊNCIAS

- [1] KOTLER, Philip. Administração de marketing São Paulo: Prentice Hall, 2000. 764p.
- [2] ROWLEY, J. Information marketing 2nd ed. Hants; Burlington: Ashgate Publishing Limited, 2006.
- [3] A. de A. Barreto, “O mercado de informação no Brasil”, *Inf.Inf.*, vol. 5, nº 1, p. 25–34, jul. 2000.
- [4] URQHART, D. J. Analysis of information services, *J. Doc.*, v. 32, n.2, p. 123-125
- [5] NIQ Ebit: Webshoppers 2017. Disponível em: <<https://www.ebit.com.br>>. Acesso em: 28 nov. 2023
- [6] ABComm: Previsão de vendas no e-Commerce para os Próximos 5 anos, 2022. Disponível em: <<https://dados.abcomm.org/previsao-de-vendas-online>>. Acesso em: 2 dez. 2023
- [7] IDC: Worldwide Artificial Intelligence Software Forecast, 2021. Disponível em: <<https://www.idc.com/>> Acesso em: 2 dez. 2023.
- [8] Eisenstein J. Introduction to Natural Language Processing. Cambridge, MA: The MIT Press; 2019
- [9] Ghosh S, Gunning D. Natural Language Processing Fundamentals. Birmingham: Packt Publishing Ltd; 2019
- [10] Grishman R. Computational Linguistics. 4th ed. New York: Cambridge University Press; 199
- [11] Luitse, D., Denkena, W. (2021). The great Transformer: Examining the role of large language models in the political economy of AI. *Big Data Society*, 8(2). <https://doi.org/10.1177/20539517211047734>
- [12] Bansal, H., Khan, R.: A review paper on human computer interaction. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.* 8, 53 (2018).
- [13] Liddy, E.D. 2001. Natural Language Processing. In *Encyclopedia of Library and Information Science*, 2nd Ed. NY. Marcel Decker, Inc.
- [14] Rosenfeld, Ronald (Aug. 2000). “Two decades of Statistical Language Modeling: Where Do We Go From Here?” In: *Proceedings of the IEEE* 88.8.
- [15] Markov, A. A. (1913). “Essai d’une recherche statistique sur le texte du roman “Eugene Onegin” illustrant la liaison des epreuve en chain (‘Example of a statistical investigation of the text of “Eugene Onegin” illustrating the dependence between samples in chain’)”. In: *Izvestia Imperatorskoi Akademii Nauk (Bulletin de l’Académie Impériale des Sciences de St.-Petersbourg)*. 6th ser. 7. English translation by Morris Halle, 1956., pp. 153–162.

[16] Shannon, Claude E. (1948). “A Mathematical Theory of Communication”. In: Bell System Technical Journal 27, pp. 379–423, 623–656

[17] SemEval-2012 Task 6: A Pilot on Semantic Textual Similarity](<https://aclanthology.org/S12-1051>) (Agirre et al., SemEval-\*SEM 2012

[18] Bhaskar Mitra, Fernando Diaz, and Nick Craswell. 2017. Learning to Match using Local and Distributed Representations of Text for Web Search. In Proceedings of the 26th International Conference on World Wide Web (WWW '17). International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 1291–1299. <https://doi.org/10.1145/3038912.3052579>

[19] Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. BERTScore: Evaluating text generation with BERT. In Proc. of ICLR.

[20] Jiachang Liu, Dinghan Shen, Yizhe Zhang, Bill Dolan, Lawrence Carin, and Weizhu Chen. 2022. What makes good in-context examples for GPT-3? In Proc. of DeeLIO 2022.

[21] Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.-A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., Lample, G. (2023). LLaMA: Open and Efficient Foundation Language Models. <https://arxiv.org/abs/2302.13971v1>.

[22] Workshop, B., :, Scao, T. Le, Fan, A., Akiki, C., Pavlick, E., Ilić, S., Hesslow, D., Castagné, R., Luccioni, A. S., Yvon, F., Gallé, M., Tow, J., Rush, A. M., Biderman, S., Webson, A., Ammanamanchi, P. S., Wang, T., Sagot, B., ... Wolf, T. (2022). BLOOM: A 176B-Parameter Open-Access Multilingual Language Model. <http://arxiv.org/abs/2211.05100>