

Doi: 10.17058/rzm.v13i1.19394

# A Inteligência artificial na contramão da desinformação: uma solução com *chatbot* para aprendizagem na área da saúde

Inteligencia artificial contra la desinformación: una solución de chatbot para el aprendizaje en el sector salud

Artificial Intelligence against disinformation: a chatbot solution for education in the health sector



**Rejane Frozza<sup>1</sup>**

**Gabriel Colbeich Lopes<sup>2</sup>**

**Rolf Fredi Molz<sup>3</sup>**

**Daniela Duarte da Silva Bagatini<sup>4</sup>**

**Resumo:** Com os avanços na digitalização e no livre acesso da produção acadêmica na área da saúde, os profissionais precisam se atualizar constantemente, apesar da sobrecarga de trabalho. Assim, a inteligência artificial (IA) pode ser aliada para auxiliar na busca de artigos científicos para qualificação do conhecimento e aprendizado. Questionários foram elaborados e aplicados com estudantes e profissionais da área da saúde, cujos resultados evidenciaram a busca por tecnologias baseadas em IA generativa, como o ChatGPT, como meio de informação. Dessa forma, o objetivo foi desenvolver um *chatbot* (agente conversacional) capaz de auxiliar os profissionais da saúde no processo de busca por informações com validação científica, apresentando uma alternativa baseada em IA

<sup>1</sup> Doutorado em Computação (UFRGS). Professora do Programa de Pós-Graduação em Sistemas e Processos Industriais da Universidade de Santa Cruz do Sul (UNISC).

<sup>2</sup> Bacharel em Ciência da Computação, Universidade de Santa Cruz do Sul (UNISC).

<sup>3</sup> Doutorado em Computação (UFRGS). Professor do Programa de Pós-Graduação em Sistemas e Processos Industriais da Universidade de Santa Cruz do Sul (UNISC).

<sup>4</sup> Doutorado em Informática na Educação (UFRGS). Professora do Departamento de Engenharias, Arquitetura e Computação da Universidade de Santa Cruz do Sul (UNISC).

extrativa. Além disso, o trabalho também buscou verificar se existem riscos ao utilizar aplicações apoiadas em técnicas de geração de texto como fonte de informação.

**Palavras-chave:** *Chatbot*. Inteligência Artificial. Conhecimento.

**Resumen:** Con los avances en la digitalización y el libre acceso a la producción académica en el área de la salud, surgen exigencias a los profesionales de este campo. Al mismo tiempo que necesitan estar al día con los materiales publicados, estos trabajadores sufren una sobrecarga de trabajo en Brasil. Además, los cuestionarios administrados a estudiantes y profesionales del campo mostraron que estos grupos están utilizando tecnologías basadas en IA generativa, como ChatGPT, como medio de información. Por tanto, el objetivo de este trabajo fue desarrollar un chatbot capaz de ayudar a los profesionales sanitarios en el proceso de búsqueda de información científicamente validada, presentando una alternativa basada en IA extractiva. Además, el trabajo también buscó verificar si existen riesgos al utilizar aplicaciones basadas en técnicas de generación de texto como fuente de información.

**Palabras clave:** Chatbot. Inteligencia Artificial. Conocimiento.

**Abstract:** With advances in digitalization and free access to academic production in the health area, demands arise on professionals in this field. At the same time as they need to be up to date with published materials, these workers suffer from work overload in Brazil. Furthermore, questionnaires administered to students and professionals in the field showed that these groups are using technologies based on generative AI, such as ChatGPT, as a means of information. Therefore, the objective of this work was to develop a chatbot capable of assisting healthcare professionals in the process of searching for scientifically validated information, presenting an alternative based on extractive AI. Furthermore, the work also sought to verify whether there are risks when using applications based on text generation techniques as a source of information.

**Keywords:** Chatbot. Artificial Intelligence. Knowledge.

## Introdução

O contexto da disponibilização de produção acadêmica no mundo se modifica conforme o avanço da tecnologia da informação. Cada vez mais trabalhos científicos estão sendo publicados no formato digital, ocasionando ampliação de alcance dessas produções e facilidade no seu acesso. Outro fator que impacta na acessibilidade desses materiais é seu livre acesso, sem a necessidade da compra do material ou de assinaturas. Os últimos anos foram marcados por um grande aumento na quantidade de artigos publicados com acesso aberto, e a tendência é esse número crescer ainda mais, especialmente, na área da clínica médica (JOACHIM JÄHNE, 2021).

Apesar da grande quantidade de dados textuais presentes nos trabalhos científicos disponíveis, utilizar esses dados em modelos de inteligência artificial e aprendizado de máquina não é uma tarefa simples. Dados presentes em textos escritos em linguagem natural não são catalogados ou tratados, isto é, as informações ali presentes não estão prontas para serem utilizadas por uma aplicação, por exemplo.

Nesse contexto, as áreas de mineração de textos e processamento de linguagem natural (PLN) possuem um papel fundamental: a extração das informações latentes de dentro do texto. Dentre as várias aplicações criadas a partir desses campos de estudo, estão a classificação de textos com base no seu conteúdo, identificação de pontos chave e respostas a perguntas (KOVALCHUK et al., 2022).

Na área da saúde, é importante o profissional ter conhecimento das informações e materiais científicos mais atualizados dentro do seu campo de atuação. Entretanto, o grande volume de publicações neste ramo, somado à sobrecarga de trabalho sofrida pelos profissionais da saúde no Brasil (NOVARETTI et al., 2014), são fatores que dificultam o acesso a essas informações.

Nesse sentido, é importante o desenvolvimento de ferramentas que combinem o uso de técnicas computacionais de aprendizado de máquina com a grande quantidade de dados textuais presentes nos artigos científicos de livre acesso. É importante que aplicações com foco em geração de conhecimento e educação aliem agilidade e acessibilidade na utilização e avaliação científica dos dados.

Assim, o objetivo deste trabalho é desenvolver um *chatbot* capaz de auxiliar os profissionais da saúde a encontrarem informações com validação científica de forma prática, utilizando-se de processamento de linguagem natural para extração de dados em

literatura biomédica. Além disso, também foi avaliado o papel da Inteligência Artificial no processo de aprendizagem, abordando, principalmente, as técnicas extrativas e abstrativas (generativas) do PLN.

Para apoiar e direcionar a construção da ferramenta proposta, foi realizado um levantamento bibliográfico sobre os principais temas e foram aplicados questionários com estudantes e profissionais da área da saúde.

O artigo está organizado nas seguintes seções: a seção 2 aborda a área da inteligência artificial no contexto na educação; a seção 3 apresenta a arquitetura do sistema desenvolvido com todas as suas etapas; a seção 4 descreve o processo de avaliação e os resultados atingidos.

## **Inteligência Artificial e Educação**

Para iniciar a discussão sobre os impactos da IA na área da educação, é preciso definir alguns temas essenciais.

Na Computação, a inteligência artificial (IA) é uma área que se dedica ao estudo da automação e do comportamento inteligente (LUGER, 2013). Sua definição é abrangente e engloba muitos campos de estudo, como sistemas baseados em conhecimento, aprendizado de máquina, agentes inteligentes, algoritmos genéticos e outros. Por isso, esta pesquisa tem como foco o processamento de linguagem natural (PLN), uma subárea da IA.

O PLN é um campo das áreas interação humano-computador e inteligência artificial, e pode ser definido pela interpretação e compreensão de linguagem natural humana por computadores. Apesar de ter surgido na metade do século XX, com algoritmos mais simples e lineares, o estado da arte atual do PLN são algoritmos mais complexos, que utilizam aprendizado de máquina e aprendizado profundo (RANI et al., 2017).

Para o desenvolvimento de uma aplicação que utiliza processamento de linguagem natural, existem alguns fatores a serem considerados. É importante definir a tarefa que será executada, e com base na tarefa, a escolha de um modelo de linguagem (*Language Model* - LM) existente ou o desenvolvimento de um novo. Como exemplo de

LM, pode ser citado o BioBERT (LEE et al., 2019), um modelo de linguagem criado para executar tarefas em textos da área da biomedicina, com destaque para seu desempenho na tarefa de perguntas e respostas (*Question Answering* - QA).

Uma das possíveis aplicações da área de PLN são sistemas de perguntas e respostas. Nesses sistemas, baseados na tarefa de QA, o usuário faz perguntas em linguagem natural e recebe uma resposta baseada em um texto, referido como contexto. Existem duas classificações para este tipo de tarefa: o QA extrativo e o QA generativo (abstrativo). A principal tarefa do QA extrativo é realizar um recorte de um trecho dentro de um determinado contexto. Já o método abstrativo utiliza um modelo de linguagem generativo para a construção da resposta, e tem como base a pergunta realizada e o contexto (DING et al., 2022).

Um exemplo de sistema baseado em perguntas e respostas que utiliza a abordagem generativa é o ChatGPT. Um dos principais atrativos dessa aplicação é a facilidade na utilização e velocidade de retorno das respostas. Entretanto, utilizá-lo como ferramenta de consulta pode apresentar alguns riscos. O ChatGPT utiliza um LLM (*Large Language Model*) treinado em textos de domínio geral, que faz com que respostas específicas possam ser imprecisas e enviesadas, como no caso de médicos ou estudantes de medicina realizando perguntas das suas áreas de atuação (WANG et al., 2023).

A IA generativa tornou-se popular há pouco tempo, após a disponibilização de aplicações para o público geral. Qadir (2023) explica que, apesar de ser importante a área da educação se atualizar e trazer novas tecnologias, é preciso avaliar os riscos e as limitações. O ChatGPT, ferramenta avaliada pelo autor, pode ser interessante como “instrumento para gerar engajamento durante a etapa de aprendizagem, devido ao seu estilo conversacional”. Entretanto, a presença de um professor ou tutor qualificado é necessária, levando em consideração que o aplicativo pode produzir informações falsas e enviesadas.

Qadir explica que o desempenho de um sistema baseado em inteligência artificial generativa está diretamente relacionado à qualidade da base de dados, cujo modelo foi submetido a treinamento. Contudo, mesmo considerando as técnicas e a base de dados do estado da arte, ainda não é possível garantir integralmente a veracidade dos dados gerados.

Por meio das entrevistas realizadas com estudantes e profissionais da área da saúde, foi constatado que o ChatGPT está incluso entre os métodos de busca de informação dos participantes. Nesse sentido, é evidente a necessidade de desenvolvimento de trabalhos e pesquisas que criem métodos não prejudiciais para introdução destas novas tecnologias no ramo da educação. Outros resultados obtidos foram que os principais objeto de busca são sintomas e tratamentos, e os motivadores dessas pesquisas são trabalhos solicitados pelo curso, escrita de artigos, discussões de caso e assistência clínica.

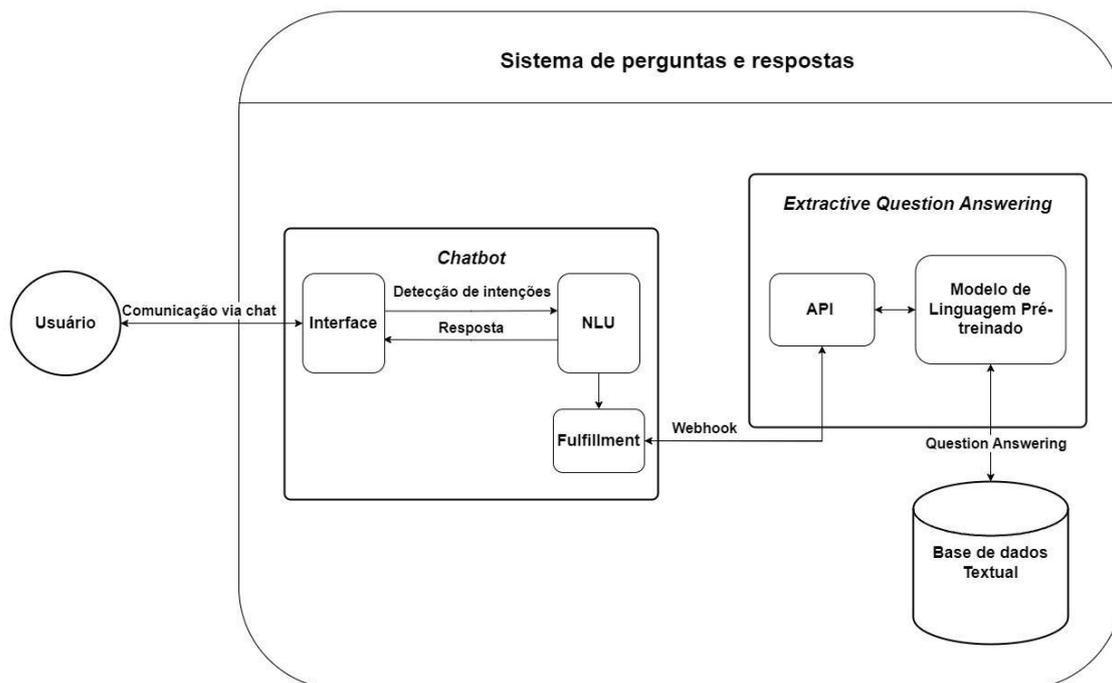
Levando em consideração o contexto definido para a proposta, que é desenvolver uma ferramenta que retorne respostas assertivas e com validação científica para apoiar profissionais na área da saúde, a técnica escolhida para compor o sistema é extrativa. Apoiada em um material acadêmico de qualidade, o papel da IA é de direcionar o usuário para encontrar a resposta de forma mais ágil, não a geração da resposta com base em um modelo de linguagem.

## **Arquitetura da ferramenta desenvolvida**

A partir das abordagens e técnicas de PLN analisadas, foi elaborada uma arquitetura para a ferramenta. A arquitetura tem como objetivo atingir dois principais requisitos: agilidade na utilização e validação científica nas informações retornadas.

A solução está dividida em 2 módulos principais: o *chatbot* e o sistema de QA extrativo. A Figura 1 apresenta a comunicação entre os módulos e elementos do sistema, com o fluxo da informação iniciando com uma solicitação realizada pelo usuário até o retorno de uma resposta por meio do *chatbot*.

Figura 1 - Arquitetura da Ferramenta



Fonte: (AUTORES, 2023)

A interação com a aplicação sempre inicia pelo lado do usuário, com o envio de uma mensagem para o *chatbot*. O *chatbot* possui um sistema de detecção de intenções, que tem como papel compreender o que o usuário deseja de acordo com a mensagem enviada (*Natural Language Understanding* – NLU). O *fulfillment* é uma forma de retornar ao usuário respostas dinâmicas, permitindo integração com outras ferramentas via *webhook*, por meio de API (*Application Programming Interface*) que retorne trechos de respostas retirados de um material científico. Caso a intenção detectada seja a de “dúvida sobre o tratamento e diagnóstico”, o sistema envia a pergunta para o módulo de QA extrativo.

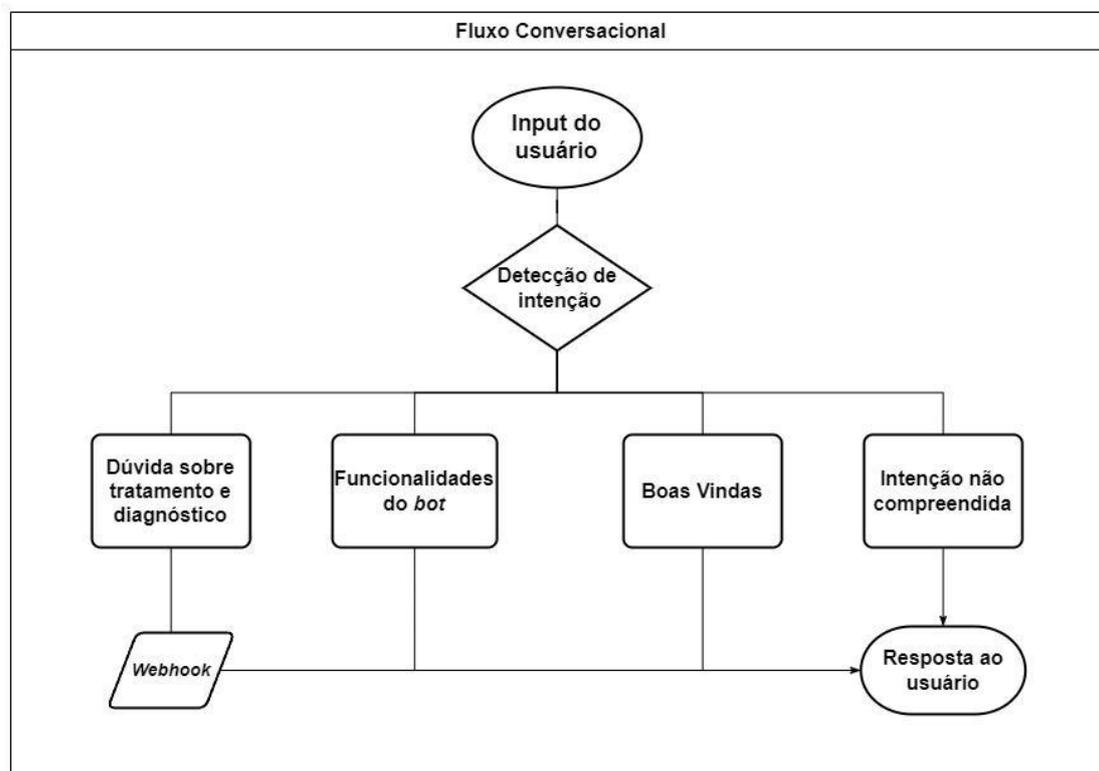
A próxima subseção aborda de forma mais detalhada o funcionamento de cada um dos módulos, assim como a especificação da base de dados e do modelo de linguagem utilizado no desenvolvimento da ferramenta.

### **Chatbot**

Para compor a interface da aplicação, que realiza a comunicação direta com o usuário, foi escolhido o *chatbot*. Os *chatbots* são sistemas computacionais de diálogo, no qual o usuário escreve uma mensagem em linguagem natural e recebe uma resposta correspondente. Também chamados de agentes conversacionais textuais, seu principal

objetivo é compreender a intenção do usuário e retornar uma resposta coerente, em linguagem natural, simulando uma conversa entre dois humanos (BARROS; TEDESCO, 2016). A Figura 2 mostra o fluxo de conversa do agente desenvolvido.

Figura 2 - Fluxo de conversa



Fonte: (AUTORES, 2023)

O processo de interação sempre inicia com o usuário, por meio de uma mensagem de texto. A intenção da mensagem é avaliada, retornando ao usuário uma resposta correspondente. A intenção de “dúvida sobre tratamento e diagnóstico” é a única que não retorna a resposta diretamente ao usuário, mas sim envia a mensagem ao módulo de QA por meio de uma conexão *webhook*, que possibilita a extração da resposta dentro da base de dados.

O *framework* utilizado no desenvolvimento do agente foi o DialogFlow (Disponível em: <https://cloud.google.com/dialogflow/docs?hl=pt-br>), que oferece suporte na etapa de criação, configuração e treinamento das intenções.

O idioma escolhido para a atuação do *chatbot* foi o inglês. O principal motivo é que a grande maioria dos modelos de linguagem foram treinados com textos em inglês, assim como é o idioma do livro que serviu como base de dados para esta pesquisa.

## **QA extrativo**

O módulo de perguntas e respostas executa a IA extrativa. A IA utiliza a pergunta enviada pelo usuário e realiza uma busca pela resposta dentro da base de dados textuais, retornando a resposta obtida, assim como a confiança na resposta. Caso a confiança (representada por uma variável numérica chamada *score*) na resposta seja baixa, o resultado não é enviado ao usuário, mas sim um texto solicitando que ele reformule a pergunta realizada.

## **Modelo de linguagem**

O modelo de linguagem (LM - Language Model) adotado pela pesquisa foi o BioM-ELECTRA-Large, conforme Alrowili e Vijay-Shankler (2021). O LM, assim como suas configurações e parâmetros, estão disponíveis na plataforma Hugging Face, que disponibiliza ferramentas para construção de aplicações com foco em aprendizado de máquina. Testes comparativos de desempenho, realizados com outros modelos de linguagem treinados no domínio biomédico, indicaram a superioridade do BioM-ELECTRA para a tarefa selecionada.

Um dos fatores que pode explicar o bom desempenho do LM escolhido ao extrair respostas de perguntas da área da saúde é o seu treinamento. O BioM-ELECTRA-Large foi pré-treinado com milhões de resumos de trabalhos da base de dados PubMed, disponível no Portal de Periódicos da CAPES. Após esse treinamento, foi realizado um ajuste fino para a tarefa de QA com o conjunto de dados Squad (Stanford Question Answering Dataset) (RAJPURKAR et al., 2016).

## **Base de dados textuais**

A base de dados selecionada para compor a ferramenta foi desenvolvida a partir do livro Clinical guidelines - Diagnosis and treatment, que aborda tratamento e diagnósticos em hospitais primários (MÉDECINS SANS FRONTIÈRES, 2016). O livro aborda de forma prática e científica sintomas e tratamentos no universo do atendimento clínico. O conteúdo do livro, disposto em formato PDF, foi tratado e processado para obter os melhores resultados em conjunto com o modelo de linguagem BioM-ELECTRA-Large.

## **Avaliação da ferramenta e resultados obtidos**

Testes com diversos modelos de linguagem foram realizados, com objetivos de coletar métricas que auxiliem na definição de um LM que se adapte melhor com a temática da pesquisa.

Durante a execução dos testes, três métricas foram coletadas: o tempo de execução da tarefa de extração de resposta, já que impacta diretamente no tempo total de execução do sistema; a qualidade da resposta extraída, avaliando coerência e correlação entre a pergunta; o *score* obtido, que é uma variável numérica retornada pelo modelo junto com a resposta extraída, e que estima a chance da resposta estar relacionada com a pergunta.

Dessas três métricas, o tempo de execução e a qualidade da resposta extraída foram consideradas para avaliação. O *score*, métrica proporcionada pela própria biblioteca utilizada para execução dos testes, não se mostrou condizente, apresentando valores altos quando a qualidade da resposta estava baixa, e valores baixos com respostas consideradas de boa qualidade.

Todos os testes foram executados em uma máquina cujos recursos computacionais disponíveis são uma unidade de processamento Intel(R) Core (TM) i7-7700HQ, e 16 GB de RAM como memória. Todos os modelos de linguagem testados foram submetidos às mesmas condições e configurados com os mesmos parâmetros.

As principais diferenças entre os modelos testados são os conjuntos de dados aos quais eles foram submetidos durante a etapa de pré-treinamento e de ajuste fino. A Tabela 1 apresenta uma comparação entre esses modelos.

Tabela 1 – Comparativo dos modelos de linguagem

Modelo de linguagem	Pré-treinamento	Ajuste fino	Método de treinamento (tarefa)
<b>Bioformer-8L</b>	Resumos do PubMed; Textos completos do PMC	SQuAD1	<i>masked language model</i>
<b>BioBERT Large</b>	Resumos do PubMed; Textos completos do PMC	SQuAD1	<i>masked language model</i>

<b>BioM ELECTRA Large</b>	Resumos do PubMed	SQuAD2	<i>replaced token detection</i>
<b>BiomedNLP_PubMedBERT</b>	Resumos do PubMed; Textos completos do PMC	Conjunto de dados desconhecido	<i>masked language model</i>

Fonte: (AUTORES, 2023)

Dentre os modelos escolhidos, todos foram treinados em conjunto de dados textuais da área da saúde. Referente às suas arquiteturas, três modelos são baseados no BERT (DEVLIN et al., 2019) e um é baseado no ELECTRA (CLARK et al., 2020).

## Resultados

Cada um dos modelos selecionados foi submetido a três perguntas, resultando em 12 execuções no total. Todos os modelos trabalharam com as mesmas perguntas. A Tabela 2 apresenta os resultados de uma das perguntas utilizadas nos testes.

Tabela 2 - Resultados do teste da pergunta “*What is AIDS?*”

Pergunta		<i>What is AIDS?</i>	
Modelo de linguagem	Tempo de execução (s)	Score	Resposta
<b>Bioformer-8L</b>	0.297	0.925	<i>Acquired immune deficiency syndrome</i>
<b>Biobert Large</b>	1.80	0.337	<i>the most advanced stage of infection with human immunodeficiency virus HIV</i>
<b>BioM ELECTRA Large</b>	1.454	0.207	<i>Acquired immune deficiency syndrome AIDS is the most advanced stage of infection with human immunodeficiency virus HIV</i>

BiomedNLP_PubMedBERT	0.425	0.165	AIDS
----------------------	-------	-------	------

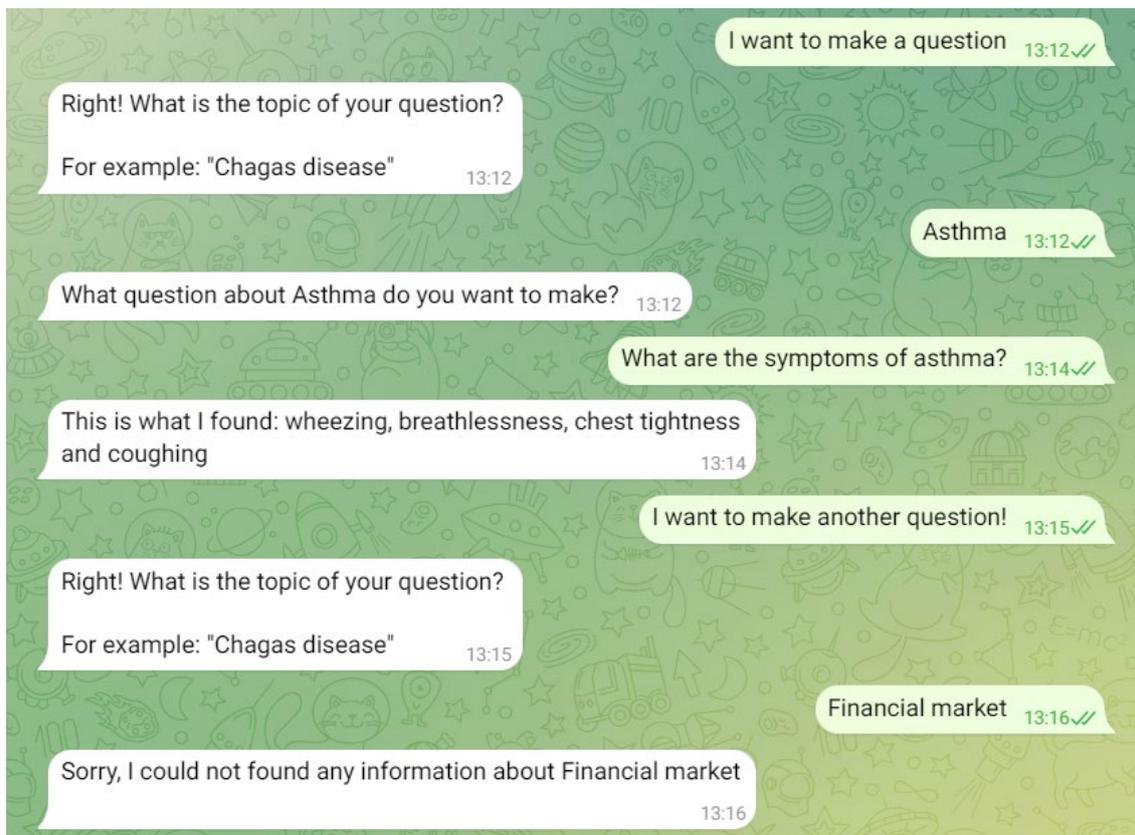
Os tempos de execução seguiram um padrão semelhante aos outros testes realizados, com os modelos definidos como “*Large*” obtendo maiores tempos de execução. Os modelos BioBERT Large e BioM ELECTRA Large foram os únicos que atingiram resultados satisfatórios no campo resposta.

Após realização dos testes, o modelo escolhido foi o BioM ELECTRA Large, pois apresentou respostas satisfatórias em todos os casos. BioBERT Large, apesar de também ter obtido bons resultados, não apresentou um contexto tão completo nas respostas quanto o modelo escolhido. O tempo de execução não foi levado em consideração, pois não influencia o funcionamento do sistema a ponto de se tornar um ruído para o usuário.

### **Exemplo de interação com a ferramenta**

Exemplo de interação com a ferramenta, executado no aplicativo mensageiro Telegram (Disponível em: <https://web.telegram.org/>), com ênfase na intenção “dúvida sobre tratamento e diagnóstico”. A Figura 3 mostra essa interação, com as mensagens do usuário em verde claro e as do agente em branco.

Figura 3 - Exemplo de interação com o agente



Fonte: (AUTORES, 2023)

A primeira ação do *chatbot* questiona o usuário sobre o assunto da pergunta, com objetivo de verificar se esse assunto está presente na base de dados. No exemplo da Figura 3, o usuário menciona o tema sobre qual deseja realizar a pergunta duas vezes. Quando o assunto é “*Asthma*”, o *chatbot* dá continuidade na conversa e solicita que a pergunta seja realizada. Já no segundo exemplo, com “*Financical market*” como assunto, o sistema não encontra nada no banco de dados, então o agente se desculpa com o usuário por não ter encontrado nada.

A resposta do *chatbot* para o usuário, quando questionado sobre “quais são os sintomas da asma”, demorou cerca de 11 segundos para ser retornada. O exemplo de interação com o agente demonstra que o sistema obteve sucesso na tarefa de responder uma dúvida relacionada à área da saúde, com informações científicas validadas, de forma rápida.

## Conclusão

Após análise realizada a partir do referencial teórico, foi possível observar uma crescente aproximação entre as áreas da inteligência artificial, mais especificamente o processamento de linguagem natural, com a Saúde e Educação. Além disso, foi possível compreender os riscos e precauções que devem ser tomadas ao unir a IA generativa com educação.

Assim, este artigo buscou compreender como o PLN pode se encaixar no dia a dia do profissional da saúde, que possui tempo limitado e sofre de sobrecarga no trabalho. As pesquisas realizadas indicaram uma oportunidade de desenvolvimento de uma aplicação, com objetivo de facilitar o acesso desses profissionais a materiais com validação científica.

Ao analisar os resultados obtidos do questionário aplicado em forma de entrevista com estudantes e profissionais da saúde, foi possível compreender melhor o problema abordado neste trabalho. Observou-se que estudantes e profissionais da área da saúde estão utilizando tecnologias baseadas em IA generativa, que podem gerar informações falsas ou imprecisas. Além disso, as entrevistas possibilitaram visualizar a proximidade que esse público tem com aplicativos mensageiros e agentes conversacionais.

Durante a etapa de testagem, foram obtidos resultados satisfatórios em relação às respostas retornadas. Esse processo permitiu a definição de um modelo de linguagem eficiente, que tornou possível a ferramenta retornar respostas coerentes e de forma rápida para o usuário.

Por meio do processo de avaliação da aplicação, foi possível evidenciar que, ao unir técnicas de QA extrativo com agentes conversacionais, é possível auxiliar os profissionais da saúde por meio de uma ferramenta de fácil usabilidade, que retorna respostas com validação científica. Dessa forma, o sistema desenvolvido cumpre com os objetivos estabelecidos no trabalho, otimizando o tempo desses profissionais no processo de busca por informações.

Unindo uma solução computacional ao contexto da educação e da saúde, a pesquisa desenvolvida evidencia a importância da aproximação entre diferentes áreas no ramo da pesquisa acadêmica, auxiliando na construção de conhecimento e no aprendizado.

## Referências

- JÄHNE, J. The future of scientific publication is Open Access, but needs diversity, equability and equality! *Innovative Surgical Sciences*, v. 6, n. 2, p. 49-51, 2021.
- KOVALCHUK, O.; BANAKH, S.; MASONKOVA, M.; BEREZKA, K.; MOKHUN, S.; FEDCHYSHYN, O. Text Mining for the Analysis of Legal Texts. 2022 12th International Conference on Advanced Computer Information Technologies (ACIT), Ruzomberok, Slovakia, pp. 502-505, 2022.
- NOVARETTI, M. C. Z.; SANTOS, E. D. V.; QUITÉRIO, L. M.; DAUD-GALLOTTI, R. M. Sobrecarga de trabalho da Enfermagem e incidentes e eventos adversos em pacientes internados em UTI. 2014 *Revista Brasileira de Enfermagem*, 67, 692-699.
- LUGER, G. F. *Inteligência artificial*. 6. ed. São Paulo: Pearson, 2013. E-book. Disponível em: <https://plataforma.bvirtual.com.br>. Acesso em: 22 abr. 2024.
- RANI, P. J.; BAKTHAKUMAR, J.; KUMAAR, B. P.; KUMAAR, U. P.; KUMAR, S. Voice controlled home automation system using Natural Language Processing (NLP) and Internet of Things (IoT). 2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM), Chennai, India, pp. 368-373. 2017.
- LEE, J., YOON, W., KIM, S., KIM, D., KIM, S., SO, C. H., & KANG, J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*, 36(4), 1234-1240. 2019.
- DING, Y.; HUANG, Z.; WANG, R.; ZHANG, Y.; CHEN, X.; MA, Y.; CHUNG, H.; HAN, S. C. V-Doc: Visual questions answers with Documents. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 21492-21498), 2022.
- WANG, C.; LIU, S.; YANG, H.; GUO, J.; WU, Y.; LIU, J. Ethical considerations of using ChatGPT in health care. *Journal of Medical Internet Research*, 25, e48009, 2023.
- QADIR, J. Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. 2023 *IEEE Global Engineering Education Conference (EDUCON)* (pp. 1-9). IEEE.
- BARROS, F. A.; TEDESCO, P. A. *Agentes inteligentes conversacionais: conceitos básicos e desenvolvimento*. Sociedade Brasileira de Computação, 2016.
- ALROWILI, S.; VIJAY-SHANKER, K. BioM-transformers: building large biomedical language models with BERT, ALBERT and ELECTRA. In *Proceedings of the 20th workshop on biomedical language processing* (pp. 221-227), 2021.
- RAJPURKAR, P.; ZHANG, J.; LOPYREV, K.; LIANG, P. Squad: 100,000+ questions for machine comprehension of text. *arXiv preprint arXiv:1606.05250*, 2016.
- FRONTIÈRES, MÉDECINS SANS (Org.) *Clinical guidelines—Diagnosis and treatment manual*. Geneva: Médecins Sans Frontières, 2016

DEVLIN, J., CHANG, M. W., LEE, K., & TOUTANOVA, K. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018

CLARK, K.; LUONG, M. T.; LE, Q. V.; MANNING, C. D. Electra: Pre-training text encoders as discriminators rather than generators. arXiv preprint arXiv:2003.10555, 2020.