

# Elucidativa: usando modelos de linguagem para explicar resultados de exames complementares

Luana Cruz Machado<sup>1</sup>, Rafael Petri Pinheiro<sup>1</sup>, Felipe Sahb Furtado<sup>2</sup>

<sup>1</sup>Faculdade de Medicina – Universidade Federal do Rio de Janeiro (UFRJ) – Rio de Janeiro, RJ – Brasil

<sup>2</sup>Amparo – São Paulo, SP – Brasil

{luana,rafael, felipe}, luanacruzmachado@hotmail.com,  
rafaelpetri@ufrj.br, felipe@amparo.com.br

**Abstract.** *In Brazil, approximately two billion medical tests are conducted annually, many of which patients struggle to interpret, potentially leading to loss of follow-up and lack of awareness regarding their health status. Language models (LLMs), such as GPT-4, have been leveraged to interpret and elucidate medical tests, thereby fostering patient autonomy. This project utilizes optical character recognition (OCR) in conjunction with GPT-4 to extract and simplify medical reports, making information more accessible to patients. Despite being a proof of concept, qualitative studies are necessary to validate the accuracy of the explanations and assess whether patients retain information better after receiving simplified reports.*

**Resumo.** *No Brasil, cerca de dois bilhões de exames são realizados anualmente, muitos dos quais pacientes têm dificuldade em interpretar, podendo levar a perdas de seguimento e desconhecimento de sua saúde. Modelos de linguagem (LLM), como o GPT-4, têm sido explorados para interpretar e explicar exames, promovendo a autonomia do paciente. Este projeto utiliza reconhecimento óptico de caracteres (OCR) com o GPT-4 para extrair e simplificar laudos médicos, tornando as informações mais acessíveis aos pacientes. Embora seja uma prova de conceito, são necessários estudos qualitativos para validar a precisão das explicações e avaliar se os pacientes retêm melhor as informações após receberem os laudos simplificados.*

## 1. Introdução

No Brasil, exames complementares são frequentemente solicitados para direcionar o tratamento clínico. Somente no sistema de saúde suplementar, 1,1 bilhão de exames foram realizados em 2022, de acordo com o Mapa Assistencial da Saúde Suplementar (AGÊNCIA NACIONAL DE SAÚDE SUPLEMENTAR, 2023). Normalmente, os resultados de exames laboratoriais, de imagem, ou patológicos, como biópsias, são entregues diretamente ao paciente – seja fisicamente ou através de meios digitais – para que sejam apresentados ao médico assistente que os solicitou posteriormente. Esse fluxo apresenta pontos de fragilidade que podem comprometer o seguimento do cuidado. A interpretação equivocada dos resultados de exames pelos pacientes, por exemplo, pode ocasionar perda do acompanhamento (CAVALCANTE; REIS, 2021), o que é especialmente danoso nos casos de exames de rastreamento como Papanicolau, Mamografia e Colonoscopia. Isso pode ocorrer porque a linguagem utilizada nos resultados de exames é técnica, repleta de jargões e complexa, o que dificulta sua compreensão por pessoas sem instrução específica na área da saúde.

Entendendo que esse problema é importante e potencialmente solucionável de maneira escalável com o uso de tecnologias desenvolvidas recentemente, como os grandes modelos de linguagem (LLM) fundacionais, propôs-se o desenvolvimento da ferramenta Elucidativa. Acredita-se que, ao simplificar os resultados de laudos médicos para linguagem comum, abre-se caminho para maior autonomia e assertividade dos pacientes em relação às próprias condições de saúde. Isso reforça os ideais de humanização e centralidade no paciente, permitindo decisões informadas e compartilhadas.

## 2. Materiais e Métodos

Na fase inicial de especificação dos requisitos, adotou-se técnicas de Design Thinking. Para compreender melhor as necessidades dos usuários, foram realizadas entrevistas com 5 pacientes, 2 estudantes de medicina, e 2 médicos. Após estudo da jornada do usuário (i.e., o paciente) através dessas entrevistas, foi evidenciado que o melhor formato de entrada para a aplicação seria o de imagem, visto que, frequentemente, os resultados de exames laboratoriais são compartilhados fisicamente impressos em folhas de papel ou digitalmente como arquivos pdf, que podem estar protegidos ou não. No caso de resultados impressos, é simples utilizar smartphones para fotografar os documentos. Já no caso de exames cujos resultados são enviados eletronicamente, a maior parte dos usuários que entrevistamos se sentia confortável o suficiente para transformar pdfs em imagens através de técnicas como print screen em seus dispositivos móveis.

Com base nos insights gerados na etapa de Design Thinking, chegou-se a funcionalidades centrais como extração óptica do texto de imagens (OCR), envio para processamento em modelos de linguagem (LLM) e posterior geração de explicações simplificadas. Em concordância, adotou-se uma arquitetura modular, composta por módulos de interface com o usuário (UI), processamento de imagens e integração com LLM. O código da aplicação web, desenvolvida em Python, reflete essa separação entre as bibliotecas Streamlit (<https://streamlit.io/>), EasyOCR (<https://github.com/JaidedAI/EasyOCR>) e API do GPT-4 da OpenAI (<https://openai.com/index/gpt-4/>). Depois do upload da imagem contendo o resultado, a imagem é processada pela biblioteca *easyOCR* para extração de texto interpretável pelo GPT-4.

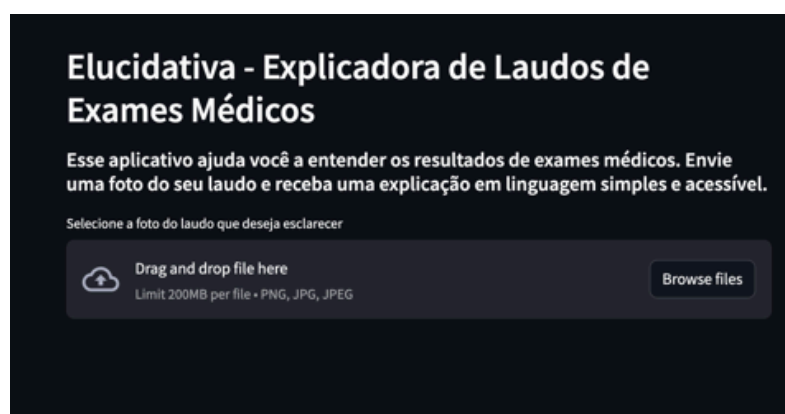
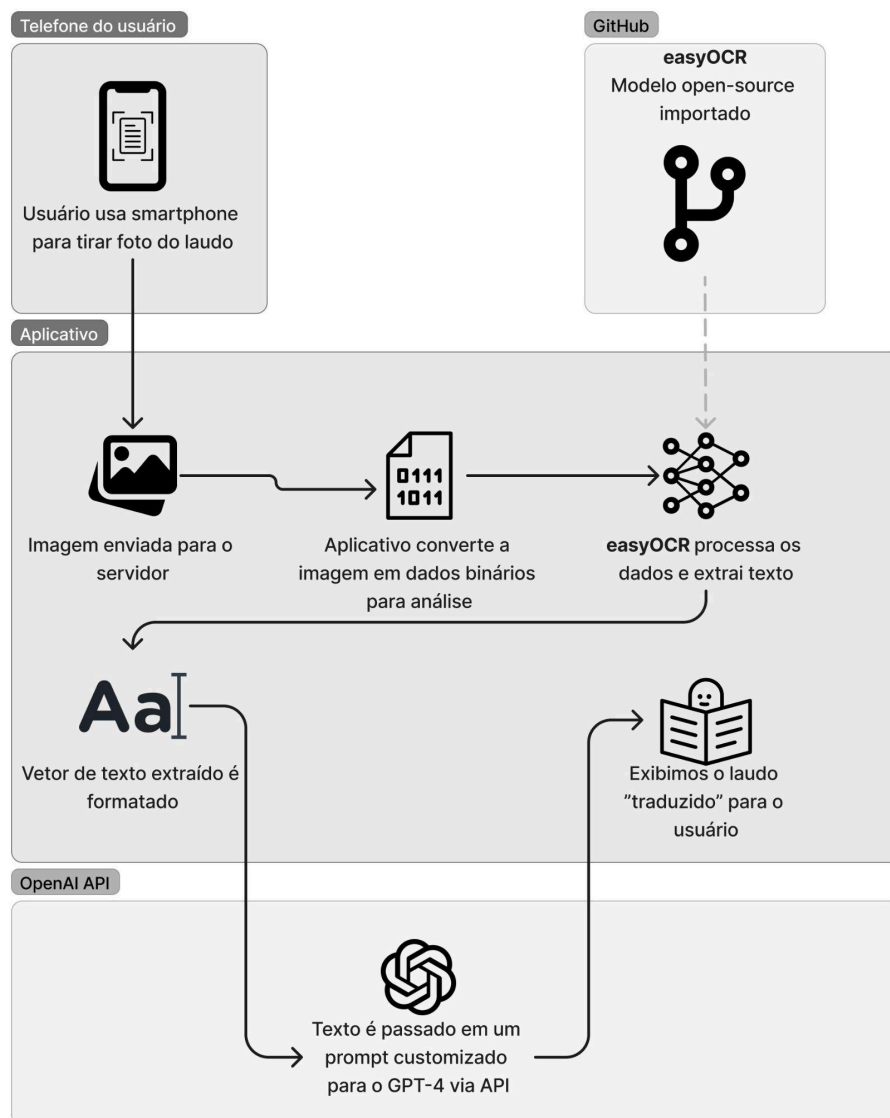


Figura 1. Interface de usuário do aplicativo

O texto extraído é exibido ao usuário para conferência enquanto o GPT-4 transforma o laudo em um texto explicativo para leigos no plano de fundo. Por fim, essa explicação simplificada é exibida na tela.



**Figura 2. Desenho esquemático do funcionamento da aplicação**

O código encontra-se disponível em <https://github.com/Felipe-Furtado/AI4GOOD/tree/main> e o aplicativo é acessível em <https://ai4good.streamlit.app/>. O vídeo de demonstração está disponível no link <https://1drv.ms/v/s!AvzssbcWxqxorWTxx5cphA9rqMf8?e=SWrY4J>.

### 3. Resultados

Testes preliminares de usabilidade e satisfação foram conduzidos com 7 usuários, todos com ensino médio completo. Os participantes demonstraram alta satisfação com a aplicação. No link disponibilizado ao público, 55 visitantes únicos foram registrados até o momento. Usuários são sempre instruídos a procurar seu médico antes de tomar

qualquer outra atitude em resposta ao resultado da interpretação do exame pelo aplicativo. A figura 3 abaixo demonstra exemplos de um laudo real e a explicação fornecida pelo app para ele. Informações que permitissem a identificação do paciente foram ocultadas.

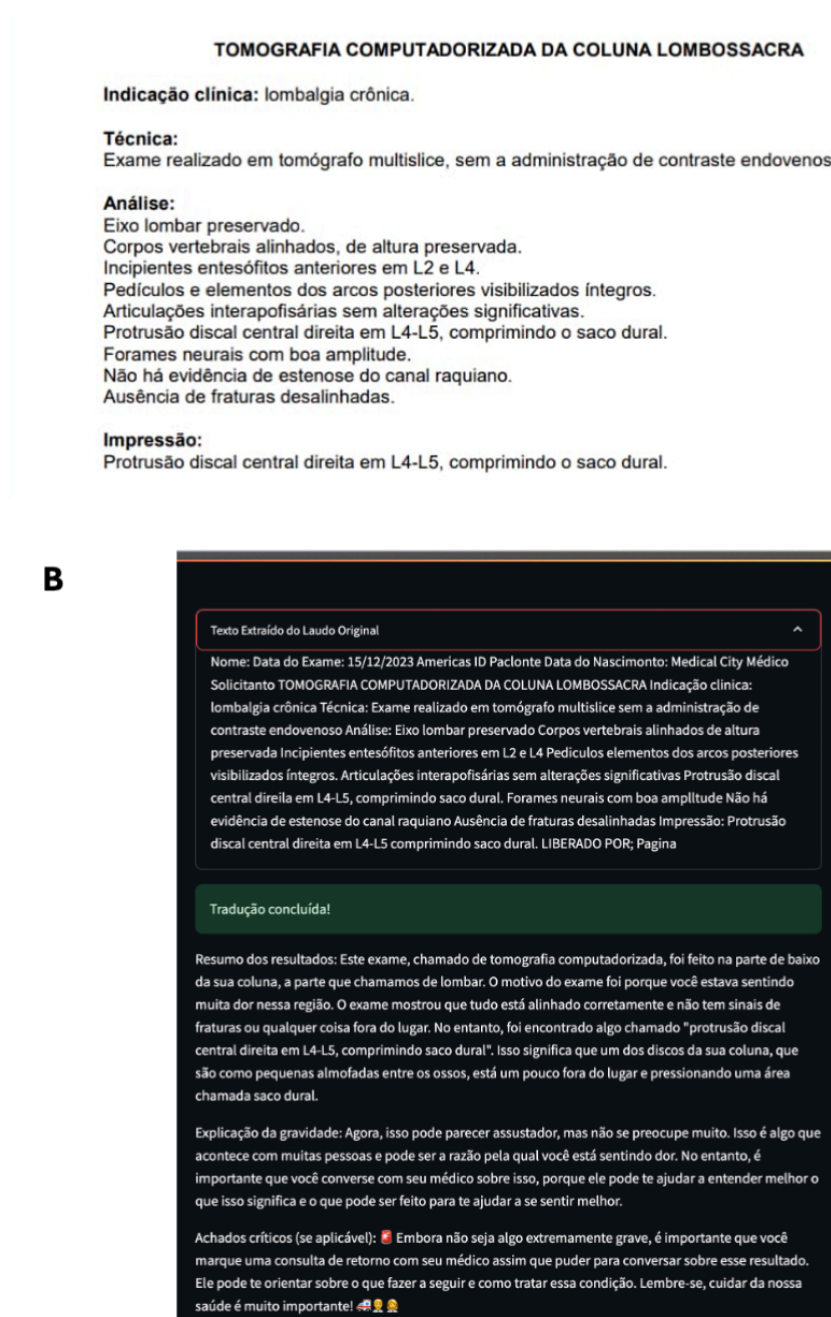


Figura 3. Exemplo de laudo padrão de exame de imagem (A) e de respectiva explicação fornecida ao usuário (B), dado o laudo apresentado.

## 4. Discussão

Algumas iniciativas prévias buscaram aplicar modelos de linguagem para melhorar a comunicação entre médicos e pacientes.

Mokmin & Ibrahim avaliaram a aceitação e desempenho de um chatbot educacional de saúde entre estudantes de graduação na Malásia. Seus resultados mostraram alta satisfação dos usuários e bom desempenho do chatbot em responder às perguntas, com baixas taxas de saída da conversa (MOKMIN; IBRAHIM, 2021). Isso corrobora o potencial da Elucidativa como ferramenta de educação em saúde para aumentar a literacia em saúde entre jovens adultos.

Chow et al. (2023) desenvolveram um chatbot educacional sobre radioterapia usando a plataforma IBM Watson Assistant. O chatbot foi capaz de engajar os usuários em conversas amigáveis, fornecendo informações sobre radioterapia de forma interativa (CHOW et al., 2023). A maioria dos usuários considerou o conteúdo útil e compreensível, apesar de algumas limitações, como respostas fixas dependentes da existência de templates previamente redigidos. Isso reforça a necessidade de usar chatbots mais adaptáveis e flexíveis - como a Elucidativa - para disseminar conhecimentos sobre tópicos de saúde complexos para o público geral.

Balanceando o potencial com a necessidade de cautela, Cadamuro et al. (2023) testaram a capacidade do ChatGPT de interpretar resultados de exames laboratoriais (CADAMURO et al., 2023). Eles verificaram que, apesar de identificar valores anormais, o ChatGPT ainda é superficial, não faz interpretações abrangentes e não considera questões pré-analíticas. Os autores concluem que o ChatGPT ainda não está pronto para interpretar resultados de exames de forma confiável, mas tem potencial futuro se treinado com dados médicos. Isso destaca que chatbots como a Elucidativa ainda requerem supervisão médica e não substituem o julgamento clínico, mas podem complementar a educação do paciente.

Esses estudos reforçam que chatbots educacionais em saúde, como a Elucidativa, têm grande potencial para melhorar a literacia em saúde e engajar os usuários. No entanto, como a tecnologia ainda está em evolução, é fundamental testar sua acurácia e garantir a supervisão de profissionais de saúde antes da ampla implementação clínica. Estudos qualitativos adicionais são necessários para avaliar a retenção de conhecimento pelos usuários e identificar áreas de aprimoramento.

## 5. Conclusão

O aplicativo Elucidativa demonstra a capacidade dos LLMs atuais de interpretar e explicar resultados de exames de maneira satisfatória. Acoplar esse tipo de explicação para leigos a todos os resultados de exames é possível devido à natureza escalável da tecnologia e, potencialmente, pode beneficiar os usuários que terão a oportunidade de ter uma compreensão mais profunda e fundamentada do seu estado de saúde. Esse tipo de abordagem pode empoderar os usuários para a tomada de decisão informada e é mais próximo do modelo de saúde atual, centrado no paciente e em decisões compartilhadas de cuidado. No entanto, é crucial enfatizar que a aplicação proposta ainda é uma prova de conceito e requer validação adicional. Está planejado um estudo qualitativo para

avaliar a acurácia das explicações fornecidas pelo modelo de linguagem, conforme a avaliação de médicos, e se há de fato aumento da retenção de informação pelos pacientes, medindo o entendimento dos resultados antes e depois da apresentação do laudo simplificado. Entre outros aprimoramentos futuros, estão previstos deploy em ambiente com suporte a GPU para melhoria de desempenho, filtros para garantir o processamento apenas de imagens válidas, uso de datasets no treinamento dos modelos de linguagem e emprego de técnicas de Retrieval Augmented Generation (RAG).

## **6. Referências**

AGÊNCIA NACIONAL DE SAÚDE SUPLEMENTAR. Planos de saúde realizaram 1,8 bilhão de procedimentos em 2022.

CADAMURO, J. et al. Potentials and pitfalls of ChatGPT and natural-language artificial intelligence models for the understanding of laboratory medicine test results. An assessment by the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group on Artificial Intelligence (WG-AI). *Clinical chemistry and laboratory medicine*, v. 61, n. 7, p. 1158–1166, 27 jun. 2023.

CAVALCANTE, G. H. O.; REIS, G. J. DOS. Avaliação do seguimento de lesões precursoras de câncer do colo do útero – uma revisão bibliográfica. *Pesquisa e Ensino em Ciências Exatas e da Natureza*, v. 5, 25 jul. 2021.

CHOW, J. C. L. et al. Developing an AI-assisted educational chatbot for radiotherapy using the IBM Watson Assistant platform. *Healthcare (Basel, Switzerland)*, v. 11, n. 17, p. 2417, 29 ago. 2023.

MOKMIN, N. A. M.; IBRAHIM, N. A. The evaluation of chatbot as a tool for health literacy education among undergraduate students. *Education and information technologies*, v. 26, n. 5, p. 6033–6049, set. 2021.