

Título:	COMPARAÇÃO ENTRE PROMPT ENGINEERING E RAG NA EXTRAÇÃO DE CARACTERÍSTICAS DE EXAMES MAMOGRÁFICOS		
Autores:	Hanna Reckziegel Santana de Sousa Fabiane Paulus Geórgia Gonçalves da Rosa Leipelt Giovana da Costa Borges Daniela Duarte da Silva Bagatini Lorenzo Farias Priscila Schmidt Lora Gabriela de Andrade Monteiro Ana Paula Müller Débora Oliveira da Silva Cristiane Drebes Pedron		
Área	<input type="checkbox"/> Humanas <input type="checkbox"/> Sociais Aplicadas <input type="checkbox"/> Biológicas e da Saúde <input checked="" type="checkbox"/> Exatas, da Terra e Engenharias	Dimensão:	<input checked="" type="checkbox"/> Ensino <input type="checkbox"/> Pesquisa <input type="checkbox"/> Extensão <input type="checkbox"/> Inovação
<p>Resumo:</p> <p>O câncer de mama é o tipo mais comum entre as mulheres no Brasil e representa a principal causa de mortalidade por câncer nessa população. A taxa de mortalidade por câncer de mama, ajustada por idade pela população mundial, foi de 12,3 óbitos por 100 mil mulheres, em 2022. As Regiões Sul e Sudeste têm as maiores taxas (13,60 e 13,16 óbitos por 100 mil mulheres, respectivamente) (INCA, 2024). Um dos desafios no combate à doença é a demora na entrega dos resultados dos exames: em 2023, apenas 48,8% dos laudos de mamografias foram liberados em até 30 dias e 36% demoraram mais de 60 dias. Essa demora reduz as chances de cura e piora o prognóstico, evidenciando a necessidade urgente de aprimorar os processos de atendimento e diagnóstico. Diante desse cenário, o presente trabalho teve como objetivo desenvolver uma solução baseada em modelo de Inteligência Artificial (IA) pré-treinada aprimorada com RAG (Retrieval Augmented Generation) e comparar os resultados com o Prompt Engineering na extração de características de laudos de exames mamográficos. O propósito foi obter informações textuais sobre a presença ou ausência de cistos, nódulos, calcificações e microcalcificações, bem como a localização e o tamanho dessas estruturas e classificação do achado (BI-Rads - Breast Imaging Reporting and Data System). A metodologia empregada envolveu: (1) levantamento sobre o impacto do atraso na entrega de exames mamográficos; (2) diálogo com especialistas para mapear as características relevantes utilizadas no <i>prompt</i>; (3) definição do <i>prompt</i>, junto com especialistas, para direcionar o modelo de IA às saídas desejadas; (4) obtenção de exames autorizados por pacientes mediante Termo de Consentimento Livre e Esclarecido (TCLE), dos quais 13 exames em formato PDF foram utilizados para a base de dados do RAG e 3 exames em formato texto para testes da aplicação; (5) levantamento e estudo de <i>frameworks</i></p>			



e bibliotecas para o desenvolvimento do protótipo (LangChain); (6) análise de modelos de IA pré-treinados para o RAG (Mistral small 25.03); e, (7) comparação entre os resultados da extração obtidas pelo Prompt Engineering e com o uso de RAG. A aplicação com RAG, desenvolvida com LangChain, gerou respostas mais organizadas, concisas e com menor ocorrência de alucinações em relação à abordagem por Prompt Engineering, que resultou em alguns erros e informações inventadas. Essas inconsistências não foram observadas no RAG, que ainda forneceu dados complementares ausentes nas respostas geradas por *prompts* simples. Assim, o RAG demonstrou ser alternativa superior, oferecendo respostas mais precisas e contextualizadas. Em trabalhos futuros, a aplicação será aprimorada para extrair automaticamente conteúdo de laudos em PDF e apresentar resultados em interface amigável. Do ponto de vista científico, o projeto permite comparar abordagens com e sem o uso de base externa confiável (RAG). No âmbito social, a extração automatizada das principais características dos laudos e a disponibilização ágil dos resultados pode apoiar a tomada de decisão clínica e ampliar o controle dos pacientes sobre suas informações de saúde. Este trabalho contribui para os ODS 3 e 9 ao ampliar o acesso a serviços de saúde digital de qualidade, com foco na redução da mortalidade prematura e na melhoria da qualidade de vida. Além disso, integra as ações de um projeto de pesquisa apoiado pelo CNPq e pelo Decit/SECTICS/MS, no âmbito da Chamada nº 21/2023 – Estudos Transdisciplinares em Saúde Coletiva.

Instituto Nacional de Câncer (Brasil). **Controle do câncer de mama no Brasil: dados e números: 2024** / Instituto Nacional de Câncer. – Rio de Janeiro: INCA, 2024.

Link do Vídeo:

<https://drive.google.com/file/d/1urzTEt6xzEKsD-fjPz5pERgcVoUSNYXc/view?usp=sharing>