Traduz Saúde: Tradução de Documentos Médicos para Linguagem Acessível com Modelos de Linguagem de Grande Escala

Lucas Matheus Silva[1], **Danyllo Albuquerque**[1], **Damires Yluska Souza**[1], **Emanuel Dantas Filho**[1,2], **Felipe Ramos**[3]

[1] Instituto Federal da Paraíba (IFPB) – Laboratório de Inovação em Software (LIS)

[2]Instituto Federal de Pernambuco (IFPE)

[3]Universidade Federal de Campina Grande (UFCG)

lucas.silva.1@academico.ifpb.edu.br
{danyllo.albuquerque, damires.yluska}@ifpb.edu.br
emanuel.filho@jaboatao.ifpe.edu.br, felipe.ramos@virtus.ufcg.edu.br

Abstract. *This paper presents Traduz Saúde, a web-based system that uses Large Language Models (LLMs) to support the translation of medical documents into accessible language in Brazilian Portuguese. The system combines clinical-domain validation, local sensitive-data handling and structured prompt-based generation to produce explanations in four complementary modes. An exploratory evaluation was conducted with anonymized clinical documents, one medical specialist and nine lay users using an adapted Technology Acceptance Model (TAM) questionnaire. Results provide initial evidence of perceived usefulness, clarity and clinical adherence, while also highlighting methodological and ethical limitations to be addressed in future work.*

Resumo. *Este trabalho apresenta o Traduz Saúde, um sistema web que utiliza Modelos de Linguagem de Grande Escala (LLMs) para apoiar a tradução de documentos médicos em linguagem acessível, em português brasileiro. A solução combina validação de domínio clínico, tratamento local de dados sensíveis e geração baseada em prompts estruturados, produzindo explicações em quatro modos complementares. A avaliação exploratória foi conduzida com documentos clínicos anonimizados, um especialista médico e nove usuários leigos, com base em questionário adaptado do TAM. Os resultados indicam evidências iniciais de utilidade percebida, clareza e aderência clínica, além de limitações metodológicas e éticas a serem tratadas em estudos futuros.*

1. Introdução

Documentos médicos, como laudos, prescrições e resultados de exames, são frequentemente redigidos em linguagem técnica voltada a profissionais de saúde. Para pacientes e familiares, essa linguagem pode dificultar a compreensão dos achados clínicos e das orientações terapêuticas [Yang et al. 2025], comprometendo o entendimento e ampliando a ansiedade associada à incerteza sobre o próprio estado de saúde [Mathes et al. 2021]. No Brasil, esse problema é agravado por desigualdades educacionais: dados do Censo 2022 indicam que parcela expressiva da população possui baixa escolaridade, fator diretamente associado a menor letramento em saúde [Pereira et al. 2024]. Quando documentos clínicos não são compreendidos, pacientes tendem a buscar explicações em fontes genéricas na internet [Maia and Biolchini 2019], favorecendo interpretações equivocadas relacionadas à *cyberchondria* [Starcevic and Berle 2013].

Avanços em LLMs têm criado possibilidades para apoiar comunicação em saúde e educação do paciente [Singhal et al. 2023, Jeblick et al. 2022, Amin et al. 2023]. Contudo, ainda

são menos exploradas soluções em português brasileiro voltadas a documentos clínicos variados, com arquitetura leve, tratamento de dados sensíveis e múltiplos modos de explicação. Neste contexto, apresenta-se o *Traduz Saúde* [Traduz Saúde 2025], sistema web que atua como mediador entre linguagem técnica e compreensão do paciente, sem substituir a avaliação médica profissional.

As contribuições são: (i) um sistema funcional para tradução de documentos médicos em linguagem acessível em português brasileiro; (ii) uma abordagem baseada em engenharia de *prompts* estruturada em quatro técnicas complementares; (iii) uma arquitetura com validação de domínio clínico, tratamento local de dados sensíveis e alertas contextualizados; e (iv) uma avaliação exploratória com especialista médico e usuários leigos baseada no TAM.

2. Contextualização e Problema

Letramento em saúde. O letramento em saúde é entendido como a capacidade de obter, processar e compreender informações necessárias à tomada de decisões em saúde [Nutbeam and Lloyd 2021]. Revisões sistemáticas associam baixos níveis de letramento a piores desfechos clínicos, menor adesão terapêutica e maior dificuldade de interpretação de documentos [Berkman et al. 2011]. Estudos sobre legibilidade de laudos indicam que documentos clínicos podem exigir escolaridade superior à do leitor típico [Martin-Carreras et al. 2019], e a OMS reconhece o letramento em saúde como determinante social relevante [World Health Organization 2025]. Assim, sistemas computacionais que traduzem linguagem técnica para linguagem acessível podem apoiar a compreensão inicial do paciente, desde que apresentem limites claros, não produzam diagnóstico e incentivem o acompanhamento profissional.

Questão de Pesquisa. Este trabalho investiga: *Como LLMs podem apoiar a tradução de documentos médicos em linguagem acessível, favorecendo a compreensão do paciente sem substituir a avaliação médica profissional?*

Exemplo Motivador. Considere o trecho: *"Leve espessamento da fáscia plantar na sua inserção no calcâneo, sugerindo fascite plantar."* Para um paciente leigo, o enunciado é pouco informativo. O *Traduz Saúde* reformula: *"O exame mostrou uma leve inflamação na fáscia plantar, estrutura que liga o calcanhar aos dedos. Essa condição, chamada fasciíte plantar, é uma causa comum de dor na sola do pé. Não é uma emergência, mas é importante levar o resultado ao médico."*

3. Trabalhos Relacionados

LLMs vêm sendo investigados como ferramentas de apoio à comunicação em saúde. Singhal et al. [Singhal et al. 2023] demonstram que esses modelos codificam conhecimento clínico relevante; Jeblick et al. [Jeblick et al. 2022] e Lyu et al. [Lyu et al. 2024] investigam simplificação de relatórios médicos; Amin et al. [Amin et al. 2023] e Yang et al. [Yang et al. 2025] discutem IA generativa para comunicação médico-paciente. Revisões recentes indicam que aplicações de LLMs em educação do paciente ainda enfrentam desafios de legibilidade, acurácia, viés e avaliação sistemática [Aydin et al. 2024, Busch et al. 2025, Bedi et al. 2024], e que técnicas de *prompt engineering* podem melhorar clareza, embora seus efeitos variem conforme modelo e tarefa [Mudrik et al. 2025].

Permanecem lacunas relevantes: (i) poucas soluções são avaliadas em português brasileiro; (ii) muitos estudos focam apenas em radiologia ou perguntas médicas gerais, não em documentos clínicos variados; (iii) há carência de estratégias leves que combinem validação de domínio, anonimização e saída estruturada; e (iv) a avaliação de segurança, clareza e percepção de uso ainda é limitada. O *Traduz Saúde* explora essas lacunas, posicionando o modelo como mediador linguístico e não como ferramenta diagnóstica.

4. Abordagem Proposta

O *Traduz Saúde* é um sistema web para apoiar pacientes e cuidadores na compreensão inicial de documentos médicos. A ferramenta não realiza diagnóstico, não recomenda tratamento e reforça que suas respostas têm finalidade informativa.

A Figura 1 apresenta o fluxo: o usuário informa o tipo de documento e submete o conteúdo; o sistema valida o domínio médico, identifica indícios de autoria profissional, trata dados sensíveis localmente, processa com o LLM e apresenta a explicação.

Figura 1. Fluxo de processamento do Traduz Saúde.

A geração emprega quatro técnicas de engenharia de *prompts* [Girardi et al. 2025], construídas incrementalmente a partir de três objetivos: preservar o significado clínico essencial, reduzir complexidade linguística e impedir recomendações diagnósticas não presentes no documento. A cada iteração, foram inspecionadas saídas geradas para os três cenários avaliados, observando ambiguidades, excesso de simplificação e risco de extrapolação clínica: (a) *role prompting*, definindo o papel do modelo como mediador de linguagem em saúde; (b) *constraint-based prompting*, com regras negativas explícitas como "não invente informações" e "não dê diagnósticos"; (c) *few-shot prompting*, com exemplos que ancoram o estilo acessível da saída; e (d) *structured output*, com esquema JSON para separar os modos de resposta. A implementação utiliza Claude [Anthropic 2024], com *temperature* zero para validação e valor moderado para tradução.

A saída organiza-se em quatro modos: (i) *Resumo*; (ii) *Detalhado*; (iii) *Entenda Fácil*; e (iv) *Termos*. O sistema também gera alertas clínicos contextualizados sobre quando buscar atendimento. A Figura 2 ilustra a interface.

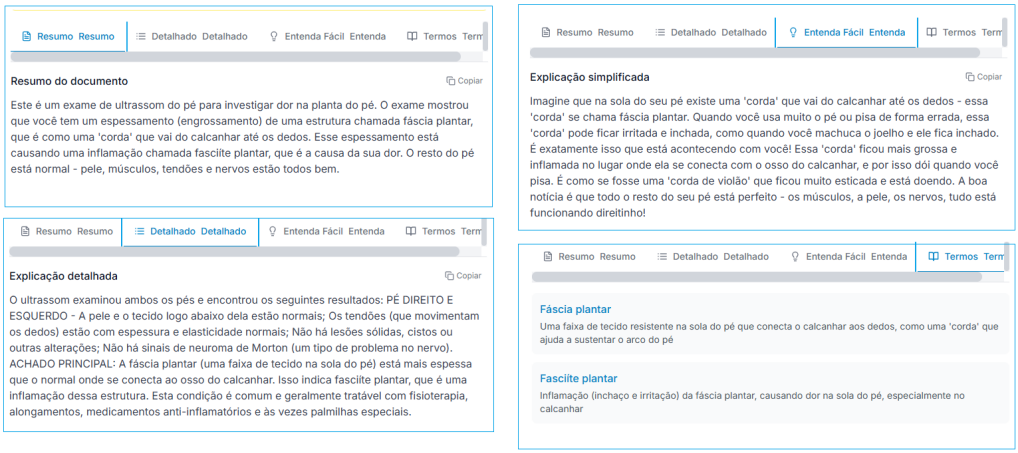

Figura 2. Traduz Saúde: saída nos quatro modos com alertas clínicos.

5. Avaliação Exploratória

A avaliação foi qualitativa e exploratória, combinando: (i) análise por um médico especialista em saúde da família; e (ii) avaliação com nove usuários leigos baseada em questionário adaptado do TAM [Venkatesh et al. 2003]. Foram utilizados três documentos clínicos reais previamente anonimizados: laudo de ultrassonografia, receituário médico e documento relacionado ao pulmão.

5.1. Avaliação por Especialista

O especialista analisou as saídas considerando fidelidade clínica, completude, clareza e segurança da informação, acompanhando o fluxo completo nos quatro modos de saída. No **Cenário 1**

(fasciíte plantar), o sistema preservou o conteúdo central e explicou o termo técnico sem sugerir diagnóstico adicional. No **Cenário 2** (receituário com sumatriptana), explicou finalidade e cuidados do medicamento sem incentivar automedicação. No **Cenário 3** (documento pulmonar), reformulou termos técnicos de forma menos alarmista, sem adicionar conclusões não presentes no documento. As respostas foram avaliadas como úteis como apoio à compreensão inicial, desde que acompanhadas de aviso explícito de que não substituem consulta médica.

5.2. Avaliação com Usuários (TAM)

Nove participantes (cinco mulheres, quatro homens, 27–59 anos, graduação mínima, amostra de conveniência) avaliaram o sistema sobre o cenário do receituário médico, após consentimento. O questionário adaptado do TAM considerou utilidade percebida (UP), facilidade de uso (FU), clareza (CC), confiança (CS) e intenção de uso (IU).

Os resultados (Tabela 1) indicam percepção favorável: nenhum D ou DT em qualquer item. UP e FU apontam concordância quanto à utilidade e intuitividade. CC indica linguagem acessível. CS apresenta cautela esperada em domínio sensível. IU mostra disposição para recomendar, com menor consenso para uso contínuo. Tabela completa disponível como material suplementar [Silva et al. 2026].

Tabela 1. Síntese TAM (n=9). Nenhum D ou DT registrado.

Dim.	Pergunta (abreviada)	CT	C/N
UP	Ajudou a entender o documento.	6	3/0
UP	Facilitou termos desconhecidos.	7	2/0
UP	Reduziria buscas na internet.	8	1/0
FU	É fácil de usar.	8	1/0
FU	Simples inserir o documento.	7	2/0
FU	Navegação intuitiva.	7	2/0
CC	Explicação é clara.	7	2/0
CC	Linguagem fácil de entender.	8	1/0
CC	Exemplos/analogias ajudam.	8	1/0
CS	Transmite confiança.	6	2/1
CS	Sentiria-se seguro usando.	7	1/1
CS	Recomendações cuidadosas.	7	2/0
IU	Usaria novamente.	7	1/1
IU	Recomendaria a outros.	8	1/0
IU	Usaria a cada exame.	5	3/1

6. Discussão, Limitações e Aspectos Éticos

Os achados sugerem que LLMs podem atuar como mediadores entre linguagem clínica e compreensão do paciente quando inseridos em fluxo com restrições explícitas, tratamento de dados sensíveis e alertas de uso responsável. A avaliação por especialista indicou que as saídas preservaram os achados centrais sem introduzir recomendações diagnósticas indevidas. Na avaliação com usuários, a dimensão de confiança apresentou maior dispersão, resultado esperado em domínio sensível. A ausência de respostas discordantes sugere percepção inicial positiva, mas não comprova melhoria objetiva de letramento nem impacto clínico.

A contribuição computacional está em demonstrar que uma arquitetura leve, baseada em *prompt engineering* estruturado, pode produzir resultados úteis em domínio sensível sem RAG, ajuste fino ou bases médicas curadas. O *prompt* é tratado como mecanismo de especificação do comportamento esperado do sistema: define papel, impõe restrições, organiza a saída e reduz riscos de extrapolação clínica [White et al. 2023]. A organização em quatro modos evita resposta única e excessivamente genérica, decisão bem recebida na avaliação qualitativa.

Há limitações importantes: avaliação com um único especialista, nove usuários e documentos previamente selecionados; amostra com escolaridade elevada, não representativa do público mais vulnerável; ausência de comparação quantitativa com RAG ou modelos alternativos; e medição de percepção de uso, não de desfechos clínicos. Em trabalhos futuros, a

comparação com RAG, modelos especializados e métricas objetivas de legibilidade, completude e fidelidade deverá ser priorizada [Bedi et al. 2024].

Do ponto de vista ético, os documentos foram anonimizados antes do processamento e o sistema declara que não substitui consulta, diagnóstico ou conduta profissional. Reconhece-se como limitação a ausência de aprovação prévia por Comitê de Ética para a avaliação com participantes; estudos futuros com pacientes, cuidadores ou dados clínicos reais deverão ser submetidos previamente a Comitê de Ética em Pesquisa.

7. Considerações Finais

Apresentou-se o *Traduz Saúde*, sistema web baseado em LLM para apoiar a tradução de documentos médicos em linguagem acessível em português brasileiro, combinando validação de domínio clínico, tratamento local de dados sensíveis, engenharia de *prompts* estruturada e quatro modos complementares de explicação. A avaliação exploratória indicou evidências iniciais de clareza, utilidade percebida e aderência clínica, com resultados a serem interpretados com cautela pelas limitações metodológicas e éticas.

Como trabalhos futuros: (i) ampliar a avaliação com maior número de especialistas e usuários, incluindo perfis com menor escolaridade; (ii) comparar sistematicamente *prompt engineering*, RAG, ajuste fino e modelos alternativos; (iii) introduzir métricas quantitativas de legibilidade, completude e segurança; e (iv) conduzir estudos com pacientes em contexto real de cuidado, após submissão a Comitê de Ética.

Disponibilidade dos Artefatos e Uso de IA

Os artefatos estão disponíveis como material suplementar [Silva et al. 2026], incluindo protótipo, repositório anonimizado, questionários, roteiros e vídeo demonstrativo. Em conformidade com o Código de Conduta da SBC, declaramos uso de ferramentas de IA como apoio à revisão textual; o conteúdo científico é de responsabilidade exclusiva dos autores.

Referências

Amin, K. S., Khosla, P., Doshi, R., Chheang, S., and Forman, H. P. (2023). Artificial intelligence to improve patient understanding of radiology reports. *The Yale Journal of Biology and Medicine*, 96:407 – 417.

Anthropic (2024). The Claude 3 model family: Opus, Sonnet, Haiku. Model card, Anthropic.

Aydin, S., Karabacak, M., Vlachos, V., and Margetis, K. (2024). Large language models in patient education: a scoping review of applications in medicine. *Frontiers in Medicine*, 11.

Bedi, S., Liu, Y., Orr-Ewing, L., Dash, D., Koyejo, S., Callahan, A., Fries, J., Wornow, M., Swaminathan, A., Lehmann, L. S., Hong, H. J., Kashyap, M., Chaurasia, A., Shah, N. R., Singh, K., Tazbaz, T., Milstein, A., Pfeffer, M. A., and Shah, N. H. (2024). Testing and evaluation of health care applications of large language models: A systematic review. *JAMA*.

Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., Viera, A., Crotty, K., Holland, A., Brasure, M., Lohr, K., Harden, E., Tant, E., Wallace, I., and Viswanathan, M. (2011). Health literacy interventions and outcomes: An updated systematic review. *Evidence Report/Technology Assessment*, 199:1–941.

Busch, F., Hoffmann, L., Rueger, C., van Dijk, E. V., Kader, R., Ortiz-Prado, E., Makowski, M. R., Saba, L., Hadamitzky, M., Kather, J., Truhn, D., Cuocolo, R., Adams, L., and Bressem, K. (2025). Current applications and challenges in large language models for patient care: a systematic review. *Communications Medicine*, 5.

Girardi, C., Fernandes, D. Y. d. S., and Rêgo, A. S. d. C. (2025). Unveiling power on combining prompt engineering techniques: An experimental evaluation on code generation. In *Proceedings of the 40th Brazilian Symposium on Data Bases (SBBD)*, pages 357–370, Fortaleza, CE, Brasil. SBC.

Jeblick, K., Schachtner, B., Dexl, J., Mittermeier, A., Stüber, A. T., Topalis, J., et al. (2022). ChatGPT makes medicine easy to understand: An example for radiology reports. *arXiv preprint arXiv:2212.14882*.

Lyu, Q., Tan, J., Zapadka, M. E., Ponnatapura, J., Niu, C., Myers, K. J., Wang, G., and Whitlow, C. T. (2024). Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning. *Visual Computing for Industry, Biomedicine, and Art*, 6(1):1–11.

Maia, M. R. and Biolchini, J. d. A. (2019). Hiperinformação na era digital: validação das informações sobre saúde. *P2P & Inovação*, 6:285–300.

Martin-Carreras, T., Cook, T. S., and Kahn Jr, C. E. (2019). Readability of radiology reports: implications for patient-centered care. *Clinical Imaging*, 57:116–121.

Mathes, B. M., Norr, A. M., Allan, N. P., Albanese, B. J., and Schmidt, N. B. (2021). Conceptualizations of cyberchondria and relations to the anxiety spectrum: Systematic review and meta-analysis. *Journal of Affective Disorders*, 290:208–216.

Mudrik, A., Nadkarni, G., Efros, O., Soffer, S., and Klang, E. (2025). Prompt engineering in large language models for patient education: A systematic review. *Preprint*.

Nutbeam, D. and Lloyd, J. E. (2021). Understanding and responding to health literacy as a social determinant of health. *Annual Review of Public Health*, 42:159–173.

Pereira, M. C. L., Morais, B. S., de Oliveira, M., Bezerra, L. M. L., dos Santos, I. P., and de Freitas, R. M. L. (2024). Saúde pública no brasil: desafios estruturais e necessidades de investimento sustentáveis para a melhoria do sistema. *Revista Cedigma*, 2(3).

Silva, L. M., Albuquerque, D., and Dantas Filho, E. (2026). Material suplementar do traduz saúde. `https://github.com/L42Matheus/tradutor-de-laudos`. Acesso em: 14 maio 2026.

Singhal, K., Azizi, S., Tu, T., Mahdavi, S. S., Wei, J., Chung, H. W., et al. (2023). Large language models encode clinical knowledge. *Nature*, 620:172–180.

Starcevic, V. and Berle, D. (2013). Cyberchondria: Towards a better understanding of excessive health-related internet use. *Expert Review of Neurotherapeutics*, 13(2):205–213.

Traduz Saúde (2025). Traduz saúde: Sistema para tradução de documentos médicos em linguagem acessível. `https://traduz-saude-app.up.railway.app/`. Acesso em: 21 mar. 2026.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3):425–478.

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., and Schmidt, D. C. (2023). A prompt pattern catalog to enhance prompt engineering with chatgpt. *arXiv preprint arXiv:2302.11382*.

World Health Organization (2025). Health literacy. `https://www.who.int/news-room/fact-sheets/detail/health-literacy`. Acesso em: 20 mar. 2026.

Yang, X., Xiao, Y., Liu, D., Zhang, Y., Deng, H., Huang, J., Shi, H., Liu, D., Liang, M., Jin, X., Sun, Y., Yao, J., Zhou, X., Guo, W., He, Y., Tang, W., and Xu, C. (2025). Enhancing doctor-patient communication using large language models for pathology report interpretation. *BMC Medical Informatics and Decision Making*, 25.