

IARA SUS: INTERFACE MÓVEL COM MODELO DE LINGUAGEM E RECUPERAÇÃO DE INFORMAÇÕES DO SUS E ALERTAS LOCAIS

EDUARDO TROTA CHAVES¹; MURILO COSTA SALEM²; RAFAEL PEREIRA REIS FILHO³; GIANA DA SILVEIRA LIMA⁴; CRISTIANE WIENKE RAUBACH⁵

¹Universidade Federal de Pelotas – eduardo.trota@yahoo.com

²Universidade Federal de Pelotas – mcsalem@inf.ufpel.edu.br

³Universidade Federal de Pelotas – rprfilho@inf.ufpel.edu.br

⁴Universidade Federal de Pelotas – giana.lima@ufpel.edu.br

⁵Universidade Federal de Pelotas – craubach.iqg@ufpel.edu.br

1. DESCRIÇÃO DA INOVAÇÃO

A IAra SUS é uma inovação social de base tecnológica destinada a facilitar o acesso da população a informações confiáveis sobre o Sistema Único de Saúde (SUS). A proposta responde à dispersão de orientações sobre documentos, campanhas, locais de atendimento e disponibilidade de serviços entre diferentes canais institucionais.

O aplicativo integra um assistente conversacional e um feed de informações locais. O usuário formula perguntas em linguagem natural, e o sistema consulta documentos institucionais, notícias e comunicados para apresentar respostas acompanhadas das respectivas fontes.

Seu funcionamento combina um modelo de linguagem de grande porte, ou *Large Language Model* (LLM), com a abordagem *Retrieval-Augmented Generation* (RAG), que recupera informações externas antes da elaboração da resposta (Lewis et al., 2020). As fontes são organizadas em conteúdos gerais do SUS e informações locais, considerando origem, data, abrangência territorial e validade.

O diferencial da IAra SUS está na integração dessas informações em um ambiente conversacional. Ao iniciar uma conversa a partir de uma publicação do feed, seu contexto é preservado. Quando não existem evidências suficientes, o aplicativo sinaliza a indisponibilidade da informação e direciona o usuário aos canais oficiais. A solução busca complementar a comunicação institucional, sem substituir profissionais ou serviços de saúde. Sua interface também considera critérios de acessibilidade previstos nas WCAG 2.2 (W3C, 2024).

2. ANÁLISE DE MERCADO

O público-alvo principal é composto por usuários do SUS que necessitam localizar e compreender informações sobre campanhas, documentos, unidades, horários e formas de acesso aos serviços. A solução pode beneficiar especialmente cidadãos com menor familiaridade com a organização do sistema de saúde ou dificuldade para identificar os canais adequados.

Secretarias de saúde, unidades assistenciais, hospitais públicos, universidades e projetos de extensão constituem o público institucional. Para essas organizações, a IAra SUS pode complementar a divulgação de campanhas, organizar informações e esclarecer dúvidas recorrentes.

As alternativas existentes incluem portais governamentais, aplicativos institucionais, redes sociais e serviços telefônicos. Entretanto, esses canais frequentemente funcionam de maneira independente ou não consideram simultaneamente a procedência, a atualidade e a abrangência territorial da informação. A IAra SUS diferencia-se por reunir essas fontes em uma interface conversacional e apresentar as referências que fundamentam as respostas.

O mercado potencial compreende a população usuária do SUS e as instituições responsáveis pela comunicação pública em saúde. Inicialmente, a solução poderá ser implantada em um município ou conjunto limitado de unidades e, após sua validação, adaptada a outros territórios. Por estar em fase de protótipo, ainda não existem dados para estimar financeiramente esse mercado.

3. ESTRATÉGIA DE DESENVOLVIMENTO E IMPLEMENTAÇÃO

Por se tratar de uma inovação social, a IAra SUS não prevê cobrança direta do cidadão. Sua implantação poderá ocorrer por meio de parcerias com secretarias de saúde, universidades e unidades assistenciais, além de editais e convênios. Futuramente, serviços de implantação, personalização e manutenção poderão contribuir para sua sustentabilidade, preservando o acesso gratuito da população.

A arquitetura adota uma estrutura cliente-servidor: o aplicativo concentra a interação com o usuário, enquanto o backend realiza o roteamento das consultas, a recuperação documental e a geração das respostas. Os documentos institucionais e locais são convertidos em representações vetoriais pelo modelo *multilingual-e5-base* (Wang et al., 2024) e armazenados no PostgreSQL com a extensão pgvector (Kane, 2026), permitindo a busca por similaridade semântica. O Qwen2.5-7B-Instruct (Yang et al., 2024) classifica as consultas conforme a necessidade de informações gerais do SUS, locais ou de ambas as fontes e utiliza os trechos recuperados para elaborar a resposta (Figura 1).

A solução encontra-se em fase de protótipo funcional em ambiente controlado, correspondente ao TRL 4. As próximas etapas incluem validar as fontes, avaliar a qualidade das respostas, realizar testes de usabilidade e acessibilidade e conduzir uma implantação-piloto. Após a consolidação do sistema, poderá ser avaliado o registro do programa de computador.

A avaliação comparará respostas geradas sem recuperação documental, com informações gerais do SUS e com a combinação entre informações gerais e locais. Serão analisadas correção factual, sustentação documental, adequação territorial e temporal, latência e facilidade de uso. A avaliação automatizada utilizará uma LLM de maior capacidade como julgadora, estratégia aplicada à análise de textos gerados por inteligência artificial (Zheng et al., 2023; Liu et al., 2023), acompanhada de revisão humana para verificar sua consistência.

Os principais riscos envolvem informações desatualizadas, falhas na recuperação e respostas sem sustentação. Para mitigá-los, serão adotados filtros territoriais e temporais, identificação das fontes, atualização periódica dos conteúdos e sinalização da ausência de evidências. A aplicação não realizará diagnósticos, prescrições ou orientações clínicas individualizadas e direcionará o usuário aos canais oficiais quando necessário.

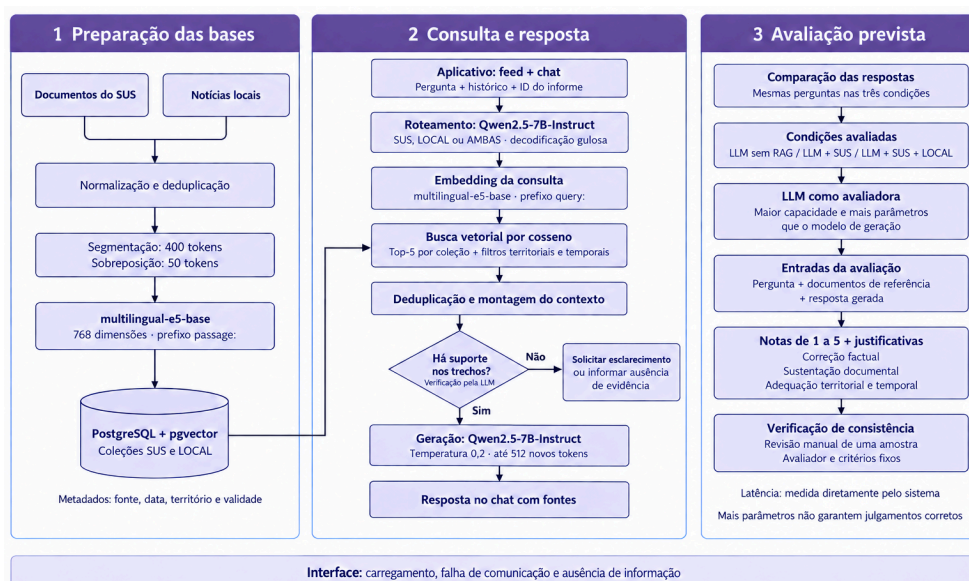


Figura 1. Fluxo de recuperação das informações e geração das respostas.

4. RESULTADOS ESPERADOS E IMPACTO

Espera-se que a Iara SUS facilite a localização e a compreensão de informações públicas de saúde, reduzindo a necessidade de consultar diferentes canais. A apresentação conversacional poderá favorecer o letramento em saúde, enquanto a integração entre orientações gerais e conteúdos locais poderá tornar as respostas mais adequadas às necessidades dos usuários. A interface proposta integra o feed de informações ao chat contextualizado (Figura 2).

Para os serviços, a ferramenta poderá ampliar o alcance das campanhas, organizar a comunicação e reduzir dúvidas recorrentes. O impacto social esperado envolve maior autonomia informacional, aproximação entre cidadãos e serviços públicos e redução das barreiras decorrentes da fragmentação das informações. Esses benefícios, entretanto, ainda precisam ser avaliados com usuários e instituições parceiras.

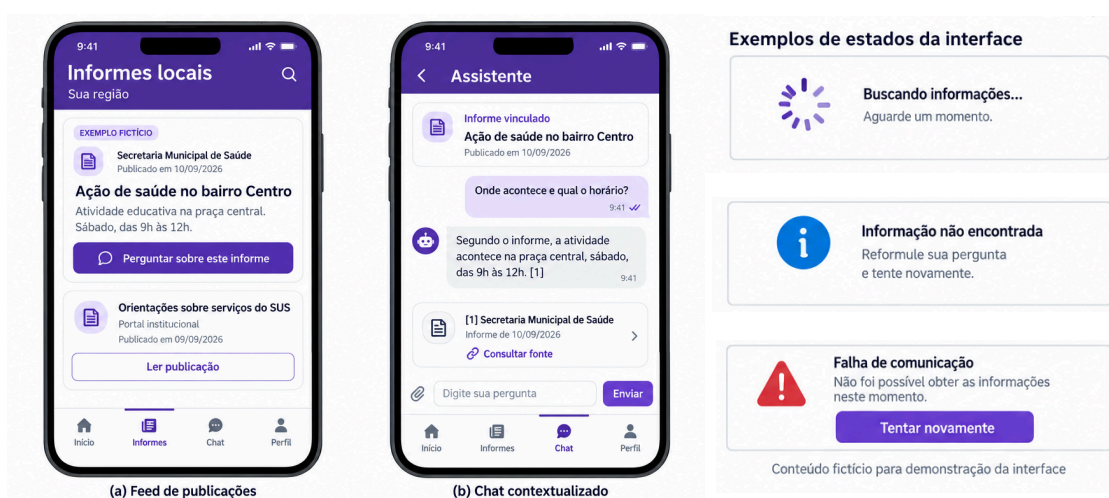


Figura 2. Exemplo da interface móvel com feed e chat contextualizado.

A implantação-piloto permitirá estimar os recursos necessários para manutenção e expansão. Como o projeto ainda está em fase de protótipo, não são apresentadas projeções financeiras nesta etapa. Futuramente, a IAra SUS poderá ser adaptada a diferentes municípios e incorporar notificações, novos recursos de acessibilidade e integração com outros canais públicos.

5. CONCLUSÕES

A IAra SUS integra inteligência artificial, informações institucionais e comunicados locais para facilitar o acesso da população a conteúdos confiáveis e contextualizados sobre o SUS. A solução tem potencial para complementar os canais existentes, favorecer o letramento em saúde e aproximar os cidadãos dos serviços públicos. As próximas etapas incluem avaliar a qualidade das respostas, a usabilidade e a acessibilidade da ferramenta, além de estabelecer parcerias para sua implantação-piloto.

6. REFERÊNCIAS BIBLIOGRÁFICAS

KANE, A. **pgvector: open-source vector similarity search for Postgres**.

GitHub, 2026. Disponível em: <https://github.com/pgvector/pgvector>. Acesso em: 13 set. 2026.

LEWIS, P. et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. In: **CONFERENCE ON NEURAL INFORMATION PROCESSING**

SYSTEMS, 34., 2020. *Advances in Neural Information Processing Systems*, v. 33, p. 9459–9474, 2020. Disponível em: <https://arxiv.org/abs/2005.11401>. Acesso em: 13 set. 2026.

LIU, Y.; ITER, D.; XU, Y.; WANG, S.; XU, R.; ZHU, C. G-Eval: NLG evaluation using GPT-4 with better human alignment. In: **CONFERENCE ON EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING**, 2023, Singapore. *Proceedings...* Association for Computational Linguistics, 2023. p. 2511–2522. DOI: 10.18653/v1/2023.emnlp-main.153.

WANG, L.; YANG, N.; HUANG, X.; YANG, L.; MAJUMDER, R.; WEI, F. Multilingual E5 text embeddings: a technical report. *arXiv preprint*, arXiv:2402.05672, 2024. Disponível em: <https://arxiv.org/abs/2402.05672>. Acesso em: 13 set. 2026.

WORLD WIDE WEB CONSORTIUM. **Web Content Accessibility Guidelines (WCAG) 2.2**. W3C Recommendation, 12 dez. 2024. Disponível em: <https://www.w3.org/TR/WCAG22/>. Acesso em: 13 set. 2026.

YANG, A. et al. Qwen2.5 technical report. *arXiv preprint*, arXiv:2412.15115, 2024. DOI: 10.48550/arXiv.2412.15115. Disponível em: <https://arxiv.org/abs/2412.15115>. Acesso em: 13 set. 2026.

ZHENG, L. et al. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. In: **CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS**, 37., 2023. *Advances in Neural Information Processing Systems*, v. 36, p. 46595–46623, 2023.