Anais do Congresso Brasileiro de Iniciação Científica
(https://periodicos.unisanta.br/COB)

ISSN 2317-1316

TOTEM DE AUTOATENDIMENTO ACESSÍVEL COM INTELIGÊNCIA ARTIFICIAL CONVERSACIONAL E REPRESENTAÇÃO DE ROTA EM PLANTA 2D

ACCESSIBLE SELF-SERVICE KIOSK WITH CONVERSATIONAL ARTIFICIAL INTELLIGENCE AND 2D FLOOR-PLAN ROUTE REPRESENTATION

Davi Xavier de Lima, Kauã Santos Silva, Rafael Luiz Forssell Ferrara Fomin, Sérgio Schina de Andrade

Universidade Santa Cecília, Faculdade de Engenharia, Curso de Engenharia da Computação
E-mails para contato: daavixaavier2511@gmail.com; rafaelforssell@gmail.com; kauas100@gmail.com

RESUMO – Este trabalho apresenta um protótipo de totem de autoatendimento acessível para consulta e localização demonstrativa de produtos em loja simulada, com interação por voz, aplicação web, API em Python/FastAPI, catálogo relacional, inteligência artificial conversacional, síntese de voz e rota em planta 2D/SVG. A validação combinou inspeção funcional, testes automatizados e baterias de intenção e voz. A classificação alcançou 97,3% de acurácia em 75 frases; nos 60 áudios sintéticos, o WER foi 7,59%, com 80% das transcrições sem erro. Nos 41 fluxos que acionaram o modelo, a latência ponta a ponta apresentou p50 de 1,99 s e p95 de 5,53 s. Os resultados sustentam a viabilidade técnica inicial do protótipo, sem caracterizar conformidade plena de acessibilidade.

Palavras-chave: Acessibilidade; Tecnologia Assistiva; Totem de Autoatendimento; Inteligência Artificial Conversacional; Representação de Rota em Planta 2D.

ABSTRACT – This work presents an accessible self-service kiosk prototype for product consultation and demonstrative location in a simulated store, combining voice interaction, a web application, a Python/FastAPI API, a relational catalog, conversational artificial intelligence, speech synthesis, and 2D/SVG floor-plan routing. Validation combined functional inspection, automated tests, and intent and voice test batteries. Intent classification reached 97.3% accuracy over 75 sentences; for 60 synthetic audio samples, WER was 7.59%, with 80% of transcriptions error-free. In the 41 flows that invoked the model, end-to-end latency was 1.99 s at p50 and 5.53 s at p95. Results support the prototype's initial technical feasibility without demonstrating full accessibility compliance.

Keywords: Accessibility; Assistive Technology; Self-Service Kiosk; Conversational Artificial Intelligence; 2D Floor-Plan Route Representation.

ISSN 2317-1316

1 INTRODUÇÃO

A acessibilidade em ambientes físicos e digitais está associada à possibilidade de uso de produtos, serviços, sistemas e informações por pessoas com diferentes capacidades e necessidades. No Brasil, a Lei Brasileira de Inclusão reforça o direito à acessibilidade em tecnologias, comunicações e serviços, enquanto as *Web Content Accessibility Guidelines* (WCAG) estruturam recomendações segundo os princípios de perceptibilidade, operabilidade, compreensibilidade e robustez (BRASIL, 2015; W3C, 2024a).

Em ambientes comerciais, a consulta de produtos e a orientação espacial ainda dependem, em muitos casos, de placas, etiquetas, corredores sinalizados, menus visuais e telas sensíveis ao toque. Para pessoas com cegueira, baixa visão, limitações motoras, idosos ou usuários com baixa familiaridade digital, essa dependência pode reduzir a autonomia e aumentar a necessidade de auxílio de terceiros. Estudos sobre quiosques acessíveis destacam que apenas disponibilizar uma tela de autoatendimento não garante uso autônomo e apontam a importância de voz, contraste, alternativas de entrada e comunicação clara (HAGEN; SANDNES, 2010; PARK et al., 2025).

A orientação em ambientes internos também apresenta desafios próprios. Sistemas de navegação destinados a pessoas com deficiência visual podem exigir infraestrutura de localização, calibração e mecanismos de orientação contextual (GUERRERO et al., 2012). Trabalhos mais recentes investigam assistentes de compras baseados em inteligência artificial para apoiar usuários com deficiência visual, porém com arquiteturas e estratégias de navegação diferentes das adotadas neste estudo (SHIBATA et al., 2025).

Diante desse contexto, este trabalho investiga como um protótipo de autoatendimento baseado em voz pode apoiar a consulta de produtos e a orientação demonstrativa em uma loja simulada, reduzindo a dependência exclusiva de leitura visual e de navegação manual por menus. O objetivo é desenvolver e avaliar parcialmente um totem capaz de receber solicitações por voz, consultar um catálogo relacional, gerar respostas textuais e sonoras condicionadas aos dados disponíveis e representar a localização do produto em planta 2D/SVG.

A proposta é delimitada como protótipo acadêmico. A planta utiliza coordenadas e rotas predefinidas, sem localização em tempo real ou cálculo dinâmico de caminho. O sistema não foi submetido, nesta etapa, a testes formais com usuários com deficiência visual nem a auditoria completa segundo WCAG. As avaliações de voz foram realizadas com áudios sintéticos, o que

ISSN 2317-1316

reduz a variabilidade de pronúncia e ruído em relação ao uso real. Assim, os resultados devem ser interpretados como evidências de integração técnica, desempenho inicial e implementação de recursos de acessibilidade, e não como certificação de conformidade ou validação de usabilidade com o público-alvo.

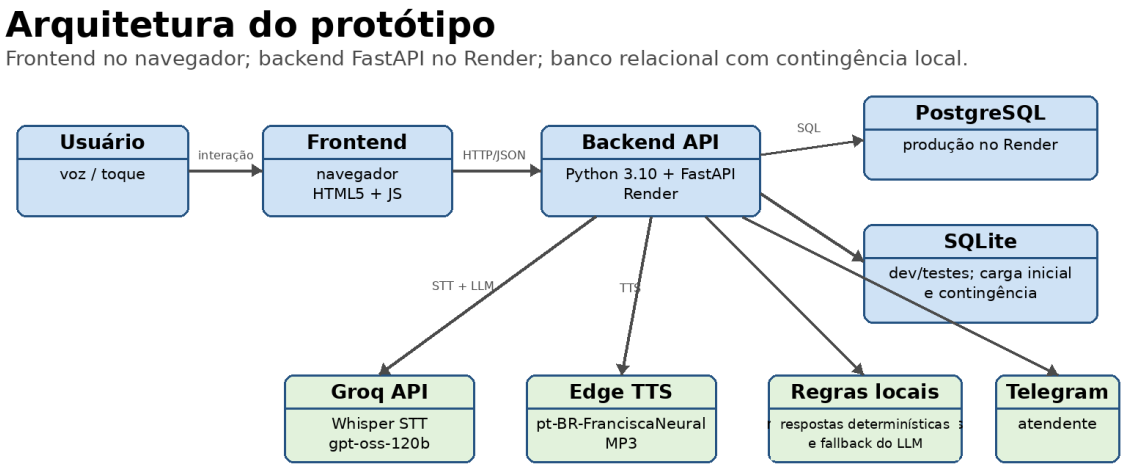
2 MATERIAIS E MÉTODOS

A pesquisa é aplicada quanto à finalidade, com caráter exploratório e descritivo, pois busca construir uma solução tecnológica para um problema prático e descrever o funcionamento do protótipo. Quanto aos procedimentos, caracteriza-se como desenvolvimento experimental e prototipagem, com validação por inspeção funcional, testes automatizados do servidor e baterias quantitativas de classificação de intenção e processamento de voz, classificação compatível com a organização metodológica apresentada por Gil (2017). O ambiente de estudo corresponde a uma loja de roupas simulada, com catálogo reduzido e setores definidos.

O desenvolvimento foi incremental e contemplou modelagem do catálogo, construção da aplicação para navegador, captura de áudio, desenvolvimento da API, integração com serviços de inteligência artificial, recursos de acessibilidade, segurança e estratégia de testes. A arquitetura é formada por interface em navegador, servidor Python 3.10 com FastAPI 0.136.1 implantado no Render por meio de Docker (imagem python:3.10-slim) e serviços externos consumidos por API. O projeto define PostgreSQL no Render como banco principal; SQLite é utilizado no desenvolvimento, nos testes, na carga inicial quando necessário e como contingência automática diante da indisponibilidade da conexão principal. O catálogo relacional fornece preço, estoque, atributos e localização dos produtos. A arquitetura geral está representada na Figura 1.

ISSN 2317-1316

Figura 1: Arquitetura geral do sistema. Fonte: Autoria própria com auxílio de Claude (Anthropic).



Na interface, o protótipo foi implementado como aplicação de página única em HTML5 e JavaScript, com telas de início, conversa, detalhe do produto, mapa e seleção de idioma. O navegador utiliza *MediaRecorder*, *getUserMedia*, *AudioContext* e *AnalyserNode* para capturar áudio e detectar silêncio. O intervalo de encerramento da fala é de 3,0 s (SILENCE_TIMEOUT_MS = 3000). O histórico do Git registra ajustes empíricos de 5,0 s para 1,5 s, retorno a 5,0 s e adoção de 3,0 s em 10/06/2026, sem estudo comparativo sistemático; o limiar de volume permaneceu em 0,02. Após esse intervalo, o áudio é enviado à rota /query-audio. O servidor transcreve o arquivo, classifica a intenção, consulta o catálogo quando necessário e constrói respostas textuais e sonoras condicionadas aos dados do banco.

A configuração final utiliza o modelo *whisper-large-v3-turbo*, da Groq, para transcrição em português, com tempo limite de 30 s. A classificação de intenção utiliza *openai/gpt-oss-120b*, também via Groq, com temperatura 0,0, nível de raciocínio *low*, parâmetro *max_tokens* igual a 512, saída em JSON e tempo limite de 20 s. A geração da resposta ao cliente emprega o mesmo modelo, temperatura 0,0, raciocínio *low*, *max_tokens* igual a 1000, contexto limitado às últimas 10 mensagens e tempo limite de 20 s. A síntese de voz utiliza *Edge-TTS* com a voz *pt-BR-FranciscaNeural* e saída MP3. Não existe contingência via OpenRouter: respostas determinísticas podem ser resolvidas por regras locais e, em falha do modelo de linguagem, essas regras constituem o mecanismo de contingência. O *Whisper tiny* local existe apenas no ambiente de desenvolvimento e não participou dos testes quantitativos apresentados.

ISSN 2317-1316

O catálogo analisado contém 41 produtos cadastrados, com campos para nome, categoria, cor, marca, preço, estoque, imagens, texto alternativo, tamanho e localização simulada. Esse conjunto constitui uma carga demonstrativa destinada à validação do protótipo e não representa a variedade de um catálogo comercial real. A associação produto-setor/corredor-coordenada alimenta a planta 2D/SVG. A rota representada é predefinida entre a entrada e a seção correspondente; portanto, não há posicionamento do usuário por sensores nem cálculo de menor caminho.

Os critérios iniciais de acessibilidade foram definidos a partir das WCAG e de boas práticas da especificação *Accessible Rich Internet Applications* (ARIA) (W3C, 2024a; W3C, 2024b). Foram verificados fluxo por voz, resposta sonora, modo de alto contraste, pausa e retomada do microfone, atributos ARIA no chat, interface simplificada, descrição textual associada ao mapa e possibilidade de atendimento humano. Como esses recursos não foram avaliados com o público-alvo, foram tratados como evidências de implementação, não como comprovação de acessibilidade plena.

A validação combinou inspeção funcional da interface publicada no Netlify, integrada ao backend publicado no Render, e baterias automatizadas executadas por scripts Python no computador de teste (Windows 11 Home Single Language, 64 bits; AMD Ryzen 5 5600; 16 GB de RAM). A inspeção manual utilizou Google Chrome, microfone Kaidi KMF4-C e conexão Wi-Fi; não há registro da versão exata do navegador na data desses testes. As baterias automatizadas não utilizaram o Chrome. A classificação de intenções consultou a Groq diretamente a partir do computador, com 75 frases, 15 por intenção: nova busca, pergunta sobre produto, mapa, encerramento e fora do escopo. A bateria de voz enviou ao servidor no Render 60 áudios sintéticos, derivados de 30 frases em duas vozes. Foram registrados acurácia por intenção, WER, latência ponta a ponta e origem de processamento apurada posteriormente. Os CSVs das execuções válidas e os logs disponíveis foram preservados.

3 RESULTADOS E DISCUSSÃO

Na execução registrada, o comando *python -m pytest -q* retornou 31 de 31 testes automatizados implementados aprovados e zero falhas. A suíte incluiu verificações das rotas principais, do serviço de produtos e do fluxo de processamento, com teste específico para garantir que filtros e atributos do catálogo permanecessem consistentes. Não foi calculada cobertura percentual de código para essa suíte. Esses 31 testes se referem ao servidor; o caminho

ISSN 2317-1316

de voz foi avaliado separadamente pela bateria de 60 áudios descrita adiante. O resultado confirma estabilidade básica no conjunto de casos implementado, sem substituir testes de carga, segurança ou usabilidade. A execução dos testes automatizados é apresentada na Figura 2.

Figura 2: Execução dos testes automatizados do servidor. Fonte: Autoria própria.

```
(.venv) $ python -m pytest -q

.............................

31 passed, 0 failed
Execução local dos testes unitários do backend após refinamentos de tamanho e pipeline.
```

A inspeção funcional demonstrou o fluxo principal de uso. A partir da tela inicial, o usuário pode iniciar o atendimento, formular uma solicitação em linguagem natural e receber uma resposta associada ao catálogo. Como exemplo visual registrado no TCC, a frase “Olá, gostaria de uma camiseta preta” resultou na recuperação de um produto compatível, acompanhado por resposta textual e cartão visual com informações comerciais. Esse exemplo é apenas ilustrativo e não constitui a base quantitativa da avaliação, que foi realizada pelas baterias de 75 frases de intenção e 60 áudios. A separação entre linguagem natural e dados relacionais reduz a dependência de geração livre pelo modelo e torna o resultado verificável no catálogo.

Após a seleção do produto, uma pergunta sobre localização acionou a planta 2D/SVG e a indicação da seção correspondente. A Figura 3 mostra a representação de rota registrada. O recurso é útil como demonstração de orientação espacial, mas sua interpretação deve permanecer limitada: o sistema não conhece a posição real do usuário, não reage a obstáculos e não executa navegação dinâmica.

ISSN 2317-1316

Figura 3: Representação de rota em planta 2D/SVG. Fonte: Autoria própria.



A API publicada apresentou estado operacional e acesso ao catálogo com 41 produtos cadastrados. O PostgreSQL no Render é o banco principal previsto na arquitetura, com contingência automática em SQLite. Contudo, os logs da bateria de voz de 06/10/2026 mostraram indisponibilidade do PostgreSQL e consultas atendidas pelo SQLite, com o mesmo catálogo. O resultado evidencia o funcionamento da contingência nesse ensaio, mas não valida a disponibilidade do banco principal; a persistência de alterações administrativas durante essa condição exige verificação. A documentação automática do FastAPI permitiu inspecionar as rotas de consulta textual e por áudio, listagem e busca de produtos, reinício de sessão e verificação de saúde da aplicação. A rota /health foi restringida para expor apenas estado, banco, quantidade de produtos e horário, mantendo detalhes de erro em registro interno.

Em relação à acessibilidade, a inspeção identificou resposta por voz, modo de alto contraste, botões de grandes dimensões, interface simplificada, pausa da captura de áudio, atributos ARIA, descrição textual associada ao mapa e possibilidade de atendimento humano. Tais decisões são coerentes com recomendações discutidas para quiosques acessíveis e interfaces digitais (HAGEN; SANDNES, 2010; PARK et al., 2025; W3C, 2024a; W3C, 2024b), mas sua efetividade ainda precisa ser verificada com participantes do público-alvo.

Na bateria de classificação de intenção, 73 das 75 frases foram classificadas corretamente, com acurácia geral de 97,3%. Nova busca e pergunta sobre produto obtiveram 14 acertos em 15 casos cada (93,3%); mapa, encerramento e fora do escopo atingiram 15 acertos em 15 casos

ISSN 2317-1316

(100%). Após uma execução sofrer limitação de taxa da Groq, a bateria foi repetida com maior intervalo entre chamadas; na execução válida, as 75 frases foram encaminhadas ao *gpt-oss-120b*, sem acionamento de contingência identificado nos resultados. A amostra de 15 frases por intenção é pequena e não permite generalização. A distribuição dos acertos está na Tabela 1.

Tabela 1. Acurácia da classificação por intenção.

Intenção	Acertos	Acurácia
Nova busca	14/15	93,3%
Pergunta sobre produto	14/15	93,3%
Mapa	15/15	100%
Encerramento	15/15	100%
Fora do escopo	15/15	100%
Geral	73/75	97,3%

Nos 60 áudios, 41 respostas acionaram o gpt-oss-120b e 19 foram resolvidas intencionalmente por regras do pipeline, sobretudo despedidas, pagamento e localização com produto citado. Para identificar a origem, as transcrições devolvidas pelo Render foram reprocessadas localmente com o LLM simulado; frases com zero chamadas ao modelo foram classificadas como regras. O resultado foi confrontado com os tempos da etapa de IA devolvidos pelo Render (até 0,007 s para regras e a partir de 0,297 s para chamadas ao modelo) e confirmado nos logs de 06/10/2026, das 13h52 às 14h19 UTC: 40 das 41 solicitações ao modelo registraram classificação de intenção e uma registrou somente geração da resposta. Não houve consultas ao modelo nas outras 19 nem registros de erro da API, portanto não foi identificado fallback por falha nessa bateria. O reprocessamento e a contingência do Render utilizaram SQLite com o mesmo catálogo; a API, porém, não retorna explicitamente a origem da resposta. Nos 41 fluxos com IA, a latência ponta a ponta teve p50 de 1,99 s e p95 de 5,53 s; no conjunto de 60 áudios, 1,66 s e 3,92 s, respectivamente (Tabela 2). O WER global foi de 7,59%, com 80% das transcrições sem erro. Áudios sintéticos podem superestimar o desempenho com falantes e ruídos reais; o plano gratuito da Groq também limita interações frequentes.

Tabela 2. Latência ponta a ponta na avaliação de voz.

Conjunto	p50	p95
Fluxos com IA (n = 41)	1,99 s	5,53 s

ISSN 2317-1316

Conjunto	p50	p95
Todos os áudios (n = 60)	1,66 s	3,92 s

Os ajustes de segurança reduziram riscos identificados pela equipe: o CORS passou a aceitar apenas origens configuradas; o painel administrativo deixou de ter chave padrão e permanece bloqueado sem ADMIN_KEY, utilizando comparação em tempo constante; a rota /health passou a limitar a exposição de detalhes; o áudio enviado é apagado após o processamento e as respostas de áudio expiram após 10 min; a memória conversacional é separada por sessão, expira em 30 min e /reset atua apenas na sessão correspondente. Posteriormente à implementação, esses ajustes foram relacionados às categorias A01 (controle de acesso), A05 (configuração de segurança) e A07 (identificação e autenticação) do OWASP Top 10:2021 (OWASP, 2021); esse referencial não orientou originalmente as correções. Como a avaliação utilizou áudios sintéticos, não houve coleta de voz de participantes. Testes futuros com pessoas deverão prever consentimento e políticas de tratamento e descarte de registros. Permanecem limitações como dependência de serviços externos, suporte multilíngue parcial, ausência de testes com o público-alvo e navegação por rotas predefinidas. Os resultados indicam viabilidade técnica inicial, não comprovação de acessibilidade plena ou desempenho em lojas reais.

4 CONCLUSÃO

O estudo desenvolveu e avaliou um protótipo acadêmico de autoatendimento com consulta por voz, catálogo relacional, respostas sonoras e representação demonstrativa de rota em planta 2D/SVG. A avaliação registrou 31 testes automatizados aprovados, 97,3% de acurácia em 75 frases, WER de 7,59% em 60 áudios sintéticos e latência p50 de 1,99 s e p95 de 5,53 s nos 41 fluxos com IA. Na bateria de voz, a contingência SQLite manteve as consultas durante a indisponibilidade do PostgreSQL; a estabilidade do banco principal e a persistência de alterações ainda precisam ser verificadas. Os recursos de voz, alto contraste, ARIA e separação de sessões evidenciam medidas de acessibilidade e segurança no cenário simulado, sem demonstrar conformidade plena ou usabilidade com o público-alvo. Trabalhos futuros incluem testes com vozes e usuários reais, auditoria WCAG 2.2, avaliação de carga e escalabilidade considerando os limites da Groq, reforço da autenticação e evolução para localização dinâmica. Os autores agradecem à Universidade Santa Cecília e ao Prof. Me. Sergio Schina de Andrade pelo acompanhamento acadêmico do projeto.

ISSN 2317-1316

5 DECLARAÇÃO DE USO DE IA

Declaração de Uso de IA: Na elaboração deste artigo, utilizou-se ChatGPT (GPT-5.6 Sol, OpenAI) para estruturar e condensar o texto do TCC, revisar a linguagem e adequar o material ao 18º COBRIC. Claude (Anthropic) auxiliou na criação da Figura 1 em código Mermaid, incorporada ao repositório em 30/04/2026, sem registro da versão exata então utilizada. Em outubro de 2026, o Claude Sonnet 5.5 também foi utilizado para consultas e esclarecimentos técnicos pontuais durante o desenvolvimento do projeto. Por meio do Claude Code, foram empregados Claude Sonnet 5 (19/09/2026) em ajustes de código, mapa e atendimento humano e Claude Opus 5.5 (06/10/2026) em correções de segurança, scripts de avaliação, troca de modelo e redação das seções 4.3.4, 4.5 e 5.1 a 5.4 do TCC que fundamentaram este artigo. O conteúdo e os resultados foram revisados pelos autores, que assumem responsabilidade pelo texto final.

6 REFERÊNCIAS

BRASIL. Lei nº 13.146, de 6 de julho de 2015. Institui a Lei Brasileira de Inclusão da Pessoa com Deficiência. Brasília, DF: Presidência da República, 2015. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13146.htm. Acesso em: 27 maio 2026.

GIL, Antonio Carlos. Como elaborar projetos de pesquisa. 6. ed. São Paulo: Atlas, 2017.

GUERRERO, Luis A. et al. An indoor navigation system for the visually impaired. Sensors, v. 12, n. 6, p. 8236-8258, 2012. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC3436027/. Acesso em: 4 jun. 2026.

HAGEN, Simen; SANDNES, Frode Eika. Toward accessible self-service kiosks through intelligent user interfaces. Personal and Ubiquitous Computing, v. 14, p. 715-721, 2010. DOI: 10.1007/s00779-010-0286-8.

OWASP. OWASP Top 10: 2021. 2021. Disponível em: https://owasp.org/Top10/. Acesso em: 8 out. 2026.

PARK, Sunyoung et al. Kiosk accessibility challenges faced by people with disabilities: an analysis of domestic and international accessibility laws/guidelines and user focus group interviews. Universal Access in the Information Society, 2025. DOI: 10.1007/s10209-023-01028-4.

SHIBATA, Larissa R. de S. et al. An AI-Based Shopping Assistant System to Support the Visually Impaired. arXiv, 2025. Disponível em: https://arxiv.org/abs/2509.01246. Acesso em: 4 jun. 2026.

ISSN 2317-1316

W3C. Web Content Accessibility Guidelines (WCAG) 2.2. 2024a. Disponível em: https://www.w3.org/TR/WCAG22/. Acesso em: 4 jun. 2026.

W3C. WAI-ARIA Overview. 2024b. Disponível em: https://www.w3.org/WAI/standards-guidelines/aria/. Acesso em: 4 jun. 2026.