ISSN 2596-2701 - Revista de Ciência da Computação - ReCiC v. 8 n. 1 - jan/dez (2026)
Dossiê Temático - Computação Aplicada à Saúde / Saúde Digital

Avaliação da experiência do usuário na utilização de um protótipo de chatbot para o apoio à prevenção do câncer do colo do útero.

Luiz Fernando Picolo
Instituto Federal de Mato Grosso do Sul (IFMS)
Nova Andradina, Mato Grosso do Sul, Brasil
luiz.picolo@ifms.edu.br

Manuella Silva Sampaio
Instituto Federal de Mato Grosso do Sul (IFMS)
Nova Andradina, Mato Grosso do Sul, Brasil
manuella.sampaio@estudantes.ifms.edu.br

Felipe de Mauro Paulo
Instituto Federal de Mato Grosso do Sul (IFMS)
Nova Andradina, Mato Grosso do Sul, Brasil
felipe.paulo@estudantes.ifms.edu.br

Renato de Souza Garcia
Instituto Federal de Mato Grosso do Sul (IFMS)
Nova Andradina, Mato Grosso do Sul, Brasil
renato.garcia@ifms.edu.br

***Resumo* —Este estudo avaliou a usabilidade, a aceitação e a experiência de fluxo durante a interação com o protótipo do chatbot Liz, desenvolvido para apoiar a prevenção do câncer do colo do útero na saúde pública de Nova Andradina/MS. Foi realizado um estudo de abordagem quanti-qualitativa com 59 participantes, utilizando os instrumentos System Usability Scale (SUS), Technology Acceptance Model (TAM) e Flow State Scale (FSS-short), além de uma questão aberta. Os dados foram analisados por meio dos escores dos instrumentos e de análise lexical no IRAMUTEQ. O SUS apresentou média de 75,7 pontos (DP = 17,6; mediana = 77,5), acima do ponto de referência de 68. No TAM, as médias (DP; alfa de Cronbach) foram 4,14 (0,74; 0,72) para UP, 4,34 (0,75; 0,74) para FUP e 4,25 (0,78; 0,71) para IUP. Para os nove itens FSS-short, a média foi 4,12 (DP = 0,81; alfa = 0,92). Na análise textual, o termo agendamento destacou-se, associado à facilidade, à clareza e à funcionalidade. Contudo, foram relatadas dificuldades pontuais de digitação, nas respostas do chatbot e na ocorrência de erros. Os resultados indicam avaliação global favorável, com oportunidades específicas de melhoria.**

***Palavras-chave* —Agentes conversacionais, Usabilidade, Experiência do usuário, Saúde digital, Câncer do colo do útero.**

I. Introdução

O câncer é um importante problema de saúde pública e ocorre quando as células do corpo crescem de forma descontrolada, podendo atingir tecidos próximos ou outros órgãos. No Brasil, estima-se que sejam registrados cerca de 781 mil novos casos de câncer por ano no triênio 2026–2028, considerando todos os tipos dessa doença [10], [19].

Entre os tipos que afetam as mulheres, destaca-se o câncer do colo do útero, que está relacionado à infecção persistente por alguns tipos do papilomavírus humano (HPV) [9]. Essa infecção é comum e, na maioria dos casos, desaparece naturalmente. No entanto, em algumas situações, pode causar alterações nas células que, com o tempo, podem evoluir para lesões que antecedem o câncer e, posteriormente, para a doença. A identificação dessas alterações por meio de exame preventivo permite iniciar medidas de prevenção e tratamento no momento adequado [8], [9], [19]. Nesse sentido, o exame citopatológico, conhecido como Papanicolau, é importante para identificar alterações nas células do colo do útero e evitar que a doença se desenvolva [33].

No município de Nova Andradina, localizado na região leste de Mato Grosso do Sul, o rastreamento do câncer do colo do útero é realizado principalmente no âmbito da Estratégia Saúde da Família (ESF). Essa estratégia busca reorganizar a assistência à saúde, articulando ações de prevenção, promoção e acompanhamento da população [7]. Além disso, a atenção primária constitui um espaço relevante para a organização das ações de rastreamento e para a ampliação do acesso à informação sobre prevenção e cuidados em saúde [37].

Contudo, um dos obstáculos apontados no trabalho desenvolvido por Picolo et al. [29] é a comunicação pré e pós agendamento que visa lembrar a paciente sobre o exame ou outras dúvidas que possam surgir. Juntamente com esse problema, Almeida [2], aponta que o câncer do colo do útero afeta principalmente grupos socioeconômicos mais vulneráveis, e fatores sociais influenciam o adoecimento e o acesso ao tratamento oncológico. Diante dessa realidade, buscou-se um meio de comunicação amplamente utilizado, capaz de alcançar a maioria das pacientes, independentemente do nível socioeconômico ou de escolaridade.

Nesse contexto, o WhatsApp se destaca como um dos aplicativos de comunicação mais utilizados. Está presente em 99% dos smartphones dos brasileiros, sendo que 93% utilizam o aplicativo diariamente. Entre os brasileiros, 90% o utilizam para enviar mensagens de texto, e 81% se comunicam por

áudio [38]. Na área médica, a utilização do WhatsApp como ferramenta para fortalecer a relação entre profissionais de saúde e pacientes tem apresentado resultados positivos, em conformidade com o Parecer CFM nº 14/2017 [28].

Nesse cenário, os chatbots, também denominados agentes conversacionais, apresentam potencial para apoiar a disseminação de informações sobre saúde e facilitar o esclarecimento de dúvidas relacionadas à prevenção. Entretanto, a utilização desse tipo de tecnologia na saúde exige avaliações que considerem não apenas seu funcionamento técnico, mas também a qualidade da interação, a usabilidade e a percepção dos usuários [1], [41], [42], [44]. Na perspectiva da Interação Humano-Computador (IHC), torna-se fundamental investigar como os usuários percebem a utilidade, a facilidade de uso, a clareza, a satisfação e a confiabilidade da interação com esses sistemas [20], [41].

Assim, com base nos conceitos e no problema apresentado, desenvolveu-se um protótipo de chatbot chamado Liz [29], como forma de avaliação inicial da proposta. Essa iniciativa contou com a parceria entre o Instituto Federal de Mato Grosso do Sul, campus de Nova Andradina (IFMS/NA), a Prefeitura de Nova Andradina (PMNA) e o apoio da Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul (FUNDECT - PICTEC 1 - 02/2021). Sua função é automatizar a comunicação entre os profissionais da saúde e as pacientes, atuando principalmente no lembrete, agendamento e reagendamento do exame preventivo, além de fornecer informações básicas sobre os problemas e conceitos relacionados ao câncer do colo do útero.

O protótipo foi submetido a uma avaliação centrada no usuário, considerando a experiência das usuárias durante a interação com a solução. Segundo Preece [32], essa etapa permite analisar aspectos relacionados à usabilidade, à facilidade de interação, à compreensão das funcionalidades, à satisfação e à adequação do sistema às necessidades das usuárias e ao contexto de utilização. A avaliação deve envolver usuários, que executem tarefas previamente definidas com o protótipo, possibilitando a coleta de dados quantitativos e qualitativos sobre a experiência de uso [3]. Conforme os princípios de projeto centrado no ser humano estabelecidos pela ISO 9241-210, o feedback dos usuários é fundamental para identificar limitações e oportunidades de melhoria, permitindo o refinamento progressivo das soluções a partir de avaliações realizadas em contextos representativos de uso [20]. Dessa forma, os resultados obtidos na avaliação podem subsidiar novas iterações do protótipo, contribuindo para o aprimoramento de suas funcionalidades e para a redução de incertezas relacionadas à usabilidade, aos requisitos e à adequação da solução antes de uma eventual implantação.

Portanto, a proposta deste trabalho foi avaliar o protótipo já desenvolvido. A avaliação centrada no usuário permite examinar a usabilidade, a aceitação e as percepções de quem interage com sistemas, considerando suas necessidades, tarefas e contexto de uso [3], [20]. Diante desse contexto, estabelece-se a seguinte pergunta de pesquisa: como as usuárias avaliam a usabilidade, a aceitação e a experiência de fluxo durante a interação com o protótipo de chatbot Liz, desenvolvido para apoiar a comunicação e fornecer informações relacionadas à prevenção do câncer do colo do útero?

A hipótese deste trabalho é que o protótipo Liz apresentará resultados positivos na avaliação da usabilidade e da aceitação pelos usuários, bem como na percepção da experiência de fluxo durante a interação. Espera-se que os resultados indiquem que a solução é compreensível, apresenta facilidade de uso e proporciona uma experiência de interação satisfatória, considerando as necessidades e o contexto de utilização. Essa hipótese será examinada com base nos indicadores coletados durante a avaliação do protótipo, sem pressupor efeitos sobre a adesão ao exame preventivo, uma vez que esse desfecho não foi mensurado no estudo.

II. Fundamentação Teórica

A. Usabilidade e experiência do usuário

A ISO 9241-11 define usabilidade como o grau em que um sistema, produto ou serviço pode ser utilizado por usuários específicos para alcançar objetivos específicos com eficácia, eficiência e satisfação em um contexto de uso determinado [21]. A eficácia diz respeito à exatidão e à completude com que os objetivos são alcançados; a eficiência relaciona os resultados obtidos aos recursos empregados; e a satisfação corresponde às respostas físicas, cognitivas e emocionais decorrentes do uso. A usabilidade, portanto, não constitui uma propriedade absoluta da tecnologia, pois depende dos usuários, das tarefas, dos objetivos e do ambiente em que a interação ocorre.

A experiência do usuário possui escopo mais amplo e compreende as percepções e respostas resultantes do uso ou da expectativa de uso de um sistema. A ISO 9241-210 orienta que o projeto centrado no ser humano considere explicitamente usuários, tarefas e ambientes, envolva os usuários ao longo do desenvolvimento, seja conduzido e refinado por avaliações e trate a experiência de forma integral [20]. Dessa forma, uma solução pode permitir a conclusão de uma tarefa e, ainda assim, produzir uma experiência insatisfatória, confusa ou pouco confiável.

Nas interfaces conversacionais, a interação ocorre por turnos de diálogo, e não apenas pela manipulação de elementos gráficos. Sua avaliação deve considerar, além da conclusão da tarefa e do tempo despendido, a compreensão das mensagens, a coerência e a continuidade do diálogo, a recuperação diante de entradas inesperadas, a adequação das respostas, a orientação fornecida ao usuário e a confiança na interação. Métricas convencionais oferecem informações relevantes, mas podem não captar isoladamente todos os aspectos próprios de um chatbot; por isso, recomenda-se combinar medidas padronizadas com observação das tarefas e dados qualitativos sobre o diálogo [25].

B. System Usability Scale

A *System Usability Scale* (SUS), proposta por Brooke [5], é um instrumento breve e independente de tecnologia para mensurar a percepção global de usabilidade. A escala contém

dez afirmações respondidas em uma escala Likert de cinco pontos, alternando itens positivos e negativos. Para calcular o escore, subtrai-se 1 da resposta de cada item ímpar e subtrai-se de 5 a resposta de cada item par. As dez contribuições, cada uma variando de 0 a 4, são somadas e multiplicadas por 2,5, resultando em um escore entre 0 e 100. Esse valor não deve ser interpretado como porcentagem nem a partir de itens isolados, mas como uma medida global comparável a dados normativos.

Bangor, Kortum e Miller [4] relacionaram os escores da SUS a adjetivos atribuídos pelos usuários: valores em torno de 71,4 foram associados a "bom", enquanto valores próximos de 85,5 foram associados a "excelente". Sauro e Lewis [35] situam o escore 68 aproximadamente no centro da distribuição de referência; assim, escores superiores a esse valor indicam usabilidade percebida acima da média. Neste estudo, foi utilizada a *Escala de Usabilidade de Sistema – versão brasileira*, traduzida e adaptada transculturalmente por Lourenço, Carmona e Lopes [24]. Essa versão preserva os dez itens e a forma de pontuação do instrumento original e apresentou equivalência semântica, idiomática, conceitual e cultural, além de consistência interna e validade de construto adequadas.

C. Technology Acceptance Model

O *Technology Acceptance Model* (TAM), formulado por Davis [17], explica a aceitação de sistemas de informação principalmente por meio da utilidade percebida e da facilidade de uso percebida. A utilidade percebida representa o grau em que uma pessoa acredita que o uso da tecnologia melhorará seu desempenho ou favorecerá a realização de uma tarefa. A facilidade de uso percebida corresponde ao grau em que a pessoa acredita que utilizar a tecnologia exigirá pouco esforço. Esses construtos influenciam a atitude e a intenção comportamental de uso, entendida como a predisposição para utilizar ou continuar utilizando a solução. Neste estudo, aplicou-se em português um questionário de 11 itens, organizado em utilidade percebida (UP), facilidade de uso percebida (FUP) e intenção de uso (IUP), com enunciados contextualizados ao chatbot Liz a partir dos construtos do modelo de Davis. A documentação disponível não identifica uma tradução brasileira validada específica para esse conjunto de itens; por isso, ele é descrito como uma adaptação contextual em português baseada no TAM original, e não como uma versão validada independente.

Extensões posteriores ampliaram a explicação da aceitação tecnológica. O TAM2 incorporou processos de influência social e processos cognitivo-instrumentais como antecedentes da utilidade percebida e da intenção de uso [39]. A *Unified Theory of Acceptance and Use of Technology* (UTAUT), por sua vez, integrou construtos de diferentes modelos em torno da expectativa de desempenho, expectativa de esforço, influência social e condições facilitadoras [40]. Embora essas extensões contextualizem a evolução teórica do campo, a avaliação realizada neste estudo concentrou-se nos construtos do TAM original: utilidade percebida, facilidade de uso percebida e intenção de uso.

D. Experiência de fluxo

A experiência de fluxo está relacionada ao estado de envolvimento intenso no qual a atenção se concentra na atividade, as ações parecem ocorrer de forma fluida e a realização da tarefa se torna intrinsecamente satisfatória [15], [16]. Esse estado é favorecido quando existe equilíbrio entre o desafio apresentado e as habilidades da pessoa, quando os objetivos são claros e quando a atividade fornece retorno compreensível sobre seu andamento.

A *Flow State Scale-2* (FSS-2), desenvolvida por Jackson e Eklund [22], operacionaliza o fluxo em nove dimensões: equilíbrio entre desafio e habilidade; fusão entre ação e consciência; objetivos claros; retorno inequívoco; concentração na tarefa; senso de controle; perda da autoconsciência; transformação da percepção do tempo; e experiência autotélica. A versão curta brasileira para atividades gerais, denominada FSS-BR-S, foi traduzida, adaptada e avaliada psicometricamente por Pinto et al. [6]; ela contém nove itens, um para cada dimensão, seguindo a proposta das versões breves da escala [23]. No presente estudo, o instrumento foi aplicado em português e identificado como FSS-short, com os nove itens utilizados na coleta. Assim, os resultados são tratados como referentes ao conjunto completo de nove itens da FSS-short. Essa medida complementa a avaliação funcional ao abordar concentração, controle, clareza e envolvimento durante a interação.

E. Articulação entre usabilidade, aceitação e fluxo

Usabilidade, aceitação e fluxo são dimensões distintas e complementares da interação. A usabilidade, como apontado na seção anterior, indica se as participantes conseguem realizar as tarefas com eficácia, eficiência e satisfação, além de revelar como percebem a facilidade global do sistema [21]. A aceitação examina se reconhecem utilidade na solução, se consideram seu uso pouco oneroso e se demonstram intenção de utilizá-la [17]. Já o fluxo acrescenta a dimensão experiencial, avaliando concentração, controle, clareza e envolvimento durante a atividade [22].

Esses construtos não são equivalentes: um sistema pode ser fácil de usar, mas percebido como pouco útil; pode ser considerado útil, mas produzir uma interação fragmentada; ou pode ser envolvente sem assegurar a conclusão correta da tarefa. Logo, a análise conjunta permite verificar se o chatbot Liz combina qualidade operacional, disposição para adoção e uma experiência de interação satisfatória. A análise qualitativa amplia essa triangulação ao revelar, nas manifestações dos participantes, aspectos do diálogo e do contexto que os escores quantitativos não descrevem isoladamente.

F. Análise lexical de dados textuais

O IRAMUTEQ é um software gratuito, desenvolvido sob a lógica *open source* e licenciado pela GNU GPL (v2), que funciona ancorado no ambiente estatístico R e permite realizar diferentes tipos de análise de dados textuais. Entre suas funcionalidades estão análises simples, como a lexicografia básica e o cálculo da frequência de palavras, e procedimentos

multivariados, como a classificação hierárquica descendente e a análise de similitude. Nas análises lexicais clássicas, o programa identifica e reformata as unidades de texto, transforma Unidades de Contexto Iniciais (UCI) em Unidades de Contexto Elementares (UCE), calcula a frequência média e o número de *hapax*, realiza lematização e cria dicionários de formas reduzidas, distinguindo formas ativas e suplementares [13].

A nuvem de palavras organiza graficamente os termos de acordo com sua frequência, permitindo a identificação rápida das palavras-chave de um corpus. A análise de similitude, fundamentada na teoria dos grafos, identifica coocorrências entre os termos e evidencia suas conexões, contribuindo para a compreensão da estrutura do texto e de seus padrões de associação semântica [13].

A Análise Fatorial de Correspondência (AFC) permite visualizar a distribuição das palavras em um plano cartesiano e identificar padrões de associação e oposição entre os termos. Sua interpretação considera a posição das palavras em relação aos eixos fatoriais: termos próximos tendem a compartilhar significados ou contextos semelhantes, enquanto termos distantes apresentam menor relação semântica [26].

III. Trabalhos relacionados

Carvalho et al. [14] conduziram um mapeamento sistemático de 27 estudos sobre avaliação de usabilidade de chatbots. Os autores encontraram predomínio de questionários (86,95%), com destaque para SUS e TAM, e observaram que os estudos se concentravam na verificação, validação e teste, depois das decisões iniciais de requisitos e design. Esse resultado é útil para situar a presente pesquisa, que avalia um protótipo em uma etapa de validação inicial e pretende retroalimentar suas próximas iterações.

No domínio da saúde, Holmes et al. [25] avaliaram um chatbot de controle de peso utilizando observação de tarefas, SUS, UEQ e o Chatbot Usability Questionnaire (CUQ). O estudo mostrou que o SUS e outras medidas convencionais são úteis, mas não capturam sozinhas todos os aspectos da usabilidade conversacional, especialmente a continuidade do diálogo, o tratamento de erros e a adequação das respostas. Esse achado reforça a importância de combinar escalas com evidências qualitativas.

Ainda em saúde, Le Bonniec et al. [43] desenvolveram e validaram o AppDate-You, um chatbot de apoio à decisão sobre rastreamento do câncer do colo do útero para mulheres em situação socioeconômica desfavorável. O processo foi centrado no usuário e incluiu entrevistas, grupos focais, teste do protótipo e teste em campo. As participantes relataram satisfação, valor educativo e facilidade de uso, embora tenham sido identificadas respostas repetitivas e necessidade de esclarecer alguns termos médicos. O estudo é próximo ao presente trabalho pelo tema de rastreamento, mas avalia uma solução de decisão e prepara uma etapa posterior para medir a participação efetiva no rastreamento.

Silva et al. [36] investigaram a intenção de uso de chatbots no comércio eletrônico, com ênfase em confiança, risco percebido e experiência de *flow*. O estudo quantitativo contou com 226 participantes europeus e utilizou Modelagem de Equações Estruturais. A confiança aumentou a intenção de uso e a experiência de *flow*, oferecendo evidência para a articulação entre aceitação e experiência subjetiva, embora em um domínio não relacionado à saúde.

Martins et al. [27] desenvolveram, validaram e avaliaram a usabilidade do chatbot ChatLinoHans para apoiar profissionais de saúde na abordagem de lesões da hanseníase. A validação envolveu profissionais de saúde e Tecnologia da Informação, obtendo IVC global de 0,98, concordância de 85,7% a 100% na avaliação da aparência e pontuações de usabilidade entre 77,5 e 100. O IRaMuTeQ foi empregado na construção da base de conhecimento, e não como análise lexical das percepções dos participantes.

Em conjunto, como pode ser visualizado na Tabela I, os trabalhos relacionados evidenciam avanços na avaliação da usabilidade, da aceitação e da experiência de uso de chatbots, especialmente no domínio da saúde. Entretanto, também revelam lacunas quanto à integração de diferentes perspectivas de avaliação e à utilização sistemática dos resultados para orientar as etapas iniciais do desenvolvimento. Enquanto parte dos estudos privilegia instrumentos quantitativos padronizados, outros enfatizam processos de validação centrados no usuário, sem necessariamente articular medidas de usabilidade, aceitação, experiência subjetiva e análise das percepções expressas livremente pelos participantes. Além disso, embora existam iniciativas voltadas ao rastreamento do câncer do colo do útero, permanecem oportunidades para investigar, de forma integrada, a experiência de uso de chatbots destinados à prevenção e ao apoio ao agendamento do exame. Nesse contexto, o presente estudo buscou contribuir para o preenchimento dessa lacuna ao combinar SUS, TAM e FSS-*short* com análise lexical das respostas a uma pergunta aberta, permitindo examinar diferentes dimensões da interação com o protótipo Liz. Essa abordagem não pretende superar isoladamente as limitações dos estudos anteriores, mas ampliar a compreensão sobre a experiência dos participantes por meio da complementaridade entre medidas estruturadas e percepções textuais. Por se tratar de uma validação inicial, os resultados devem ser interpretados como subsídios para o aprimoramento do protótipo, e não como evidência de efetividade na ampliação do rastreamento. Assim, destaca-se a necessidade de estudos posteriores que investiguem a experiência de uso em contextos reais, a diversidade dos públicos atendidos e os efeitos da ferramenta sobre o acesso e a adesão ao exame preventivo.

IV. Descrição do artefato

O chatbot Liz foi desenvolvido como um protótipo de apoio à conscientização e à prevenção do câncer do colo do útero no contexto da rede pública de saúde de Nova Andradina. A proposta inicial, descrita por Picolo et al. [29], foi elaborada a partir de levantamento bibliográfico, requisitos identificados com profissionais da Estratégia Saúde da Família, testes de tecnologias e prototipagem. A revisão bibliográfica dos autores

TABELA I
COMPARAÇÃO ENTRE ESTUDOS RELACIONADOS E O PRESENTE ESTUDO

Estudo	Objetivo e domínio	Instrumentos	Amostra	Principais resultados e posicionamento
Carvalho et al. [14]	Mapear avaliações de usabilidade de chatbots	Revisão sistemática; SUS e TAM entre os instrumentos identificados	27 estudos	Predomínio de questionários e avaliações tardias; fundamenta a necessidade de combinar métodos e avaliar protótipos antes da implantação.
Holmes et al. [25]	Avaliar chatbot de controle de peso em saúde	Observação de tarefas, SUS, UEQ e CUQ	Usuários de chatbot de saúde	Métricas convencionais não cobrem toda a usabilidade conversacional; sustenta a inclusão de percepções qualitativas.
Le Bonniec et al. [43]	Apoiar decisão sobre rastreamento do câncer do colo do útero	Entrevistas, grupos focais, testes alfa e beta	Mulheres de 30–65 anos e profissionais; beta: 15 mulheres	Boa aceitação e valor educativo, com problemas pontuais de repetição e terminologia; aproxima-se do domínio deste estudo.
Silva et al. [36]	Explicar aceitação e fluxo em chatbot de comércio	TAM ampliado, confiança, risco e SEM	226 participantes	Confiança associada à intenção de uso e ao fluxo; oferece base para a articulação dos construtos.
Martins et al. [27]	Validar chatbot para abordagem da hanseníase	Validação de conteúdo, aparência, usabilidade e IRaMuTeQ	Profissionais de saúde e TI	Evidências satisfatórias de validade e usabilidade; IRaMuTeQ aplicado à base de conhecimento.
Presente estudo	Avaliar o protótipo Liz para prevenção e agendamento do exame do colo do útero	SUS, TAM, FSS-*short* e análise lexical de pergunta aberta	59 participantes	Integra escalas padronizadas e percepções lexicais em uma validação inicial do protótipo.

também identificou o uso de chatbots em ações de informação, prevenção e acompanhamento em saúde [30].

A. Arquitetura e tecnologias

O chatbot Liz, em sua versão mais recente, é organizado em uma aplicação principal Node.js/Express, uma camada de persistência com Sequelize e PostgreSQL e um serviço de integração com o WhatsApp. O projeto também utiliza AdminJS para administração, Redis e Node-Resque para enfileiramento assíncrono, RiveScript para a base de diálogos, node-telegram-bot-api para o Telegram e node-schedule para tarefas programadas [31].

Na integração com o WhatsApp, o serviço separado utiliza whatsapp-web.js com LocalAuth para manter a sessão autenticada. As mensagens recebidas são encaminhadas por um stream gRPC para o worker do sistema principal; o worker coloca o evento em uma fila Redis, processa a mensagem e devolve a resposta pelo mesmo serviço gRPC. Para o envio, o serviço normaliza o identificador do destinatário para o formato do WhatsApp e chama a API local do cliente conectado. O Telegram segue fluxo semelhante de recepção por webhook/polling e processamento pelo trabalhador.

Os dados de domínio são persistidos principalmente nas entidades Paciente, Esf, Agendamento, Informacao, Avaliacao e Usuario. O cadastro relaciona a paciente ao posto da ESF e ao identificador do canal de mensagens. Os agendamentos armazenam a data e o horário previstos, a data confirmada e o estado de processamento. O serviço de agendamento executa uma rotina periódica, consulta exames próximos e envia lembretes para os identificadores de WhatsApp e Telegram cadastrados.

B. Abordagem conversacional

O protótipo é predominantemente baseado em regras. A biblioteca RiveScript carrega arquivos **.rive** com tópicos, padrões de entrada, respostas, variáveis de contexto e chamadas para funções JavaScript. O processamento não corresponde ao treinamento de um modelo de aprendizado profundo: as mensagens são normalizadas, comparadas às regras e encaminhadas para funções de negócio que consultam ou alteram o banco de dados. A base conversacional pode ser inspecionada nos tópicos de cadastro, boas-vindas, menus, informações, agendamento, finalização e avaliação, enquanto as operações de persistência ficam encapsuladas no modelo ChatBot [29], [30].

C. Fluxos e funcionalidades

O fluxo inicial verifica se o identificador do canal já está associado a uma paciente. Quando não está, o chatbot oferece o cadastro, que solicita nome, CPF, ESF e telefone, confirma cada dado e associa o registro ao WhatsApp ou Telegram. O CPF é validado antes da gravação e, quando já existe um cadastro, o identificador do novo canal pode ser associado ao registro encontrado.

Depois do cadastro, o menu principal oferece quatro caminhos: (1) agendar o exame preventivo; (2) consultar o exame marcado; (3) consultar informações sobre o câncer do colo do útero; e (4) finalizar o atendimento. O agendamento é conduzido por etapas: escolha do dia da semana, confirmação do dia, escolha do intervalo de horário e confirmação do pedido. A consulta recupera a data registrada; o módulo de informações lista temas mantidos no banco e apresenta o conteúdo selecionado; ao final, a paciente pode atribuir uma nota de 0 a 10 ao atendimento.

Neste artigo, o roteiro de uso apresentado na Figura 1 sintetiza a sequência de cadastro, consulta, agendamento e acesso às informações. A implementação original também disponibiliza diagramas de casos de uso, classes, entidade-relacionamento e fluxo de diálogo no repositório do projeto [31].

D. Tratamento de falhas e entradas não reconhecidas

O tratamento de falhas combina validações de entrada, confirmações intermediárias e respostas coringa. Nos tópicos RiveScript, entradas diferentes das opções esperadas acionam mensagens como "não tenho resposta para essa opção" e reapresentam o menu ou a etapa corrente. O cadastro verifica CPF e entradas numéricas; o fluxo de agendamento confirma dia e horário antes de salvar; e as chamadas de persistência retornam mensagens de erro quando o cadastro, o agendamento ou a avaliação não podem ser gravados.

Na camada de comunicação, o serviço do WhatsApp informa quando o cliente ainda não está pronto e expõe o estado da conexão por gRPC. A fila Redis desacopla a recepção do processamento e reduz o risco de bloquear o canal enquanto o sistema consulta o banco. Essas estratégias explicam por que termos como "erro" e "problema" podem aparecer nas percepções das participantes: eles podem se referir tanto a entradas fora do menu quanto a dificuldades de compreensão, persistência ou comunicação. Como o presente estudo não analisou logs, essa interpretação permanece contextual e não permite atribuir cada termo a uma falha técnica específica.

V. Metodologia

A. Aspectos éticos

A Resolução nº 466/2012 do Conselho Nacional de Saúde (CNS) [11] estabelece que todo projeto de pesquisa envolvendo seres humanos deve seguir as diretrizes nela previstas. No entanto, conforme o Artigo 1º, parágrafo único, da Resolução nº 510, de 7 de abril de 2016 [12], certas atividades não precisam ser registradas nem avaliadas pelo sistema CEP/CONEP, como é o caso de pesquisas de opinião pública com participantes não identificados.

Neste estudo, essa situação é descrita na seção sobre o percurso metodológico, especificamente na etapa 3 — aplicação do teste. Nessa fase, não houve identificação dos respondentes nem coleta de dados que permitissem associar as respostas a pessoas específicas, assegurando a privacidade e o sigilo dos participantes. Ainda assim, em conformidade com o item II.3 da referida normativa, foi assegurada assistência aos participantes durante todo o processo, mesmo com a manutenção do anonimato.

A Resolução nº 466/2012 também destaca, em seu item II.23, que o respeito à dignidade humana exige a obtenção do consentimento livre e informado dos participantes. Essa etapa está detalhada no percurso metodológico: os participantes foram devidamente informados sobre os objetivos do estudo e cientes de que sua participação seria voluntária. Para isso, foi utilizado o Termo de Consentimento Livre e Esclarecido (TCLE) antes do preenchimento do formulário, garantindo que o participante tivesse conhecimento completo sobre a pesquisa antes de iniciar sua participação.

Por fim, cabe ressaltar que, em conformidade com a Resolução nº 466/2012 [11], não houve envolvimento de menores de idade e que todos os dados coletados foram utilizados exclusivamente para os fins previstos neste estudo, sem qualquer uso adicional.

B. Tipo de pesquisa

O estudo adotou uma abordagem quanti-qualitativa. A dimensão quantitativa utilizou procedimentos de quantificação na coleta das informações e na análise dos dados por meio de técnicas estatísticas [34]; a dimensão qualitativa buscou compreender detalhadamente os significados e as características situacionais apresentadas [34].

A pesquisa teve natureza aplicada, buscando produzir conhecimentos voltados à utilização prática, orientados para a solução de problemas específicos. Quanto aos objetivos, foi classificada como exploratória, por procurar ampliar a compreensão sobre o problema e torná-lo mais claro. Como procedimento metodológico, empregou-se o levantamento, com a intenção de reunir dados organizados que possibilitassem uma análise mais aprofundada [18].

C. Percurso metodológico

A avaliação do chatbot seguiu a estrutura metodológica proposta por Barbosa et al. [3], que organiza o processo em cinco atividades fundamentais, garantindo o rigor científico e a validade dos resultados coletados.

1. Contextualização e análise do domínio: Inicialmente, realizou-se um estudo sobre a situação atual para compreender o domínio do problema, os perfis dos usuários e o contexto de uso do sistema. Esta etapa permitiu identificar os objetivos e as tarefas críticas que serviriam de base para os cenários de avaliação.

2. Planejamento e preparação: Nesta fase, foram definidos os objetivos específicos da avaliação e o seu escopo. O planejamento incluiu o recrutamento de participantes, a preparação da infraestrutura tecnológica e dos materiais de apoio, como termos de consentimento livre e esclarecido (TCLE) e roteiros. Para a avaliação quantitativa foram utilizados: (a) um questionário TAM aplicado em português, com 11 itens contextualizados ao chatbot Liz e distribuídos entre UP, FUP e IUP, baseado nos construtos de Davis [17]; (b) a versão brasileira da *System Usability Scale* (SUS) [24]; e (c) o instrumento identificado como FSS-short, tendo como referência brasileira a FSS-BR-S de Pinto et al. [6]. Para a parte qualitativa foi adicionada uma pergunta aberta, na qual o participante poderia expor sua opinião após utilizar o software.

3. Coleta de dados e observação: A coleta ocorreu remotamente, em ambiente online, por meio de formulário eletrônico no período entre 31 de outubro e 9 de dezembro de 2025. A execução seguiu os preceitos éticos, com a aceitação do termo de consentimento. Os participantes foram orientados a utilizar as funcionalidades da aplicação por meio do roteiro da Figura 1: realizar o cadastro, consultar informações, simular o agendamento do exame, consultar o horário agendado e concluir a interação com a avaliação. Após o uso, responderam aos questionários e à questão aberta.

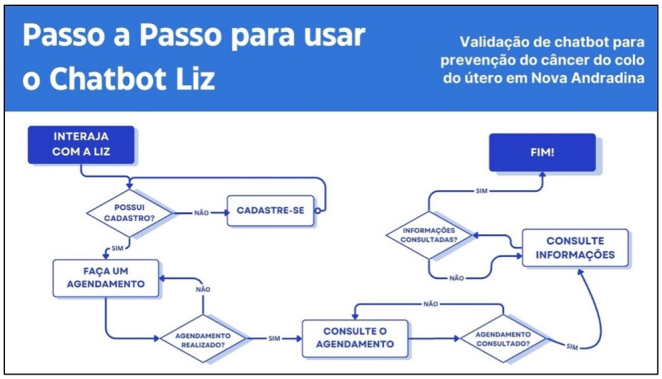


Fig. 1. Roteiro para uso do Chatbot Liz

Os dados foram registrados por meio de formulários eletrônicos e mantidos para análise posterior.

4. Interpretação dos dados: Os dados registrados foram submetidos à análise, seguindo as especificações de cada técnica. Os avaliadores buscaram atribuir significado aos dados coletados. Para os dados qualitativos, foi utilizado o software IRAMUTEQ. Esta etapa teve como foco identificar rupturas na comunicação e responder às perguntas de pesquisa formuladas no planejamento, podendo incluir análises manuais ou automáticas, conforme o tipo de dado registrado.

5. Consolidação e relato dos resultados: Por fim, os achados individuais foram consolidados em uma análise para identificar recorrências e tendências de problemas que afetam grupos de usuários. As descobertas foram organizadas em um relatório final que detalha os objetivos, o método aplicado, o perfil dos participantes, o resumo dos dados, a lista de problemas identificados e sugestões fundamentadas para o reprojeto da interface e da interação.

VI. Análise dos Resultados

O estudo contou com 59 participantes. Quanto ao sexo, 37 participantes (62,7%) foram registrados como masculinos e 22 (37,3%) como femininos. A distribuição etária foi: 37 participantes (62,7%) entre 18 e 24 anos; 6 (10,2%) entre 25 e 34 anos; 8 (13,6%) entre 35 e 44 anos; 7 (11,9%) entre 45 e 54 anos; e 1 (1,7%) entre 55 e 64 anos. Embora o chatbot seja direcionado ao agendamento do exame para o público feminino, os registros incluem participantes de ambos os sexos devido ao objetivo proposto. Assim, os resultados devem ser interpretados como uma validação inicial da experiência de uso, e não como evidência específica de adesão ao rastreamento nessa população.

A. Análise quantitativa

1) ***Usabilidade****:* Os resultados da aplicação do SUS, apresentados na Figura 2, mostram respostas heterogêneas quanto à usabilidade do sistema. Nos itens formulados positivamente, os itens 3 e 7 registraram 83% e 86% de concordância (respostas 4 ou 5), respectivamente, sugerindo facilidade de uso e aprendizagem para a maioria dos participantes.

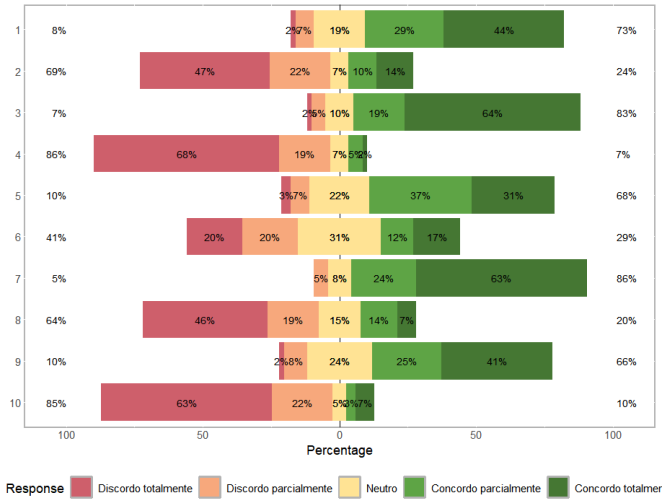


Fig. 2. Resultados do SUS para o Chatbot Liz

Nos itens formulados negativamente, a interpretação requer considerar a direção desfavorável das afirmações. As respostas originais indicam que 87% discordaram (respostas 1 ou 2) que precisariam de ajuda técnica (item 4), 85% discordaram que precisariam aprender várias coisas antes de usar o sistema (item 10), 69% discordaram que o sistema era difícil de usar (item 2) e 65% concordaram que era atrapalhado de usar (item 8). Assim, os itens 4 e 10 indicam baixa necessidade percebida de ajuda ou aprendizagem prévia, enquanto os itens 2 e 8 revelam baixa dificuldades de uso para parte dos participantes. A figura apresenta as respostas originais dos dez itens. Para o cálculo do SUS, os itens negativos foram invertidos individualmente conforme a pontuação padrão (respostas 4 e 5 recebem contribuição menor), e os percentuais desta descrição referem-se às respostas originais.

No item 6, que afirma que o chatbot apresenta muitos erros, 29% discordaram (respostas 1 ou 2), 31% permaneceram neutros e 40% concordaram (respostas 4 ou 5), indicando que a percepção de inconsistências merece investigação adicional. Em conjunto, as respostas aos itens negativos relativizam os resultados favoráveis observados nos itens 3 e 7.

O escore SUS foi calculado individualmente para os 59 participantes, aplicando-se a pontuação padrão aos dez itens de cada respondente. As estatísticas descritivas estão reunidas na Tabela II. A média ficou acima do ponto de referência de 68 discutido na fundamentação teórica, indicando usabilidade global favorável, embora os itens negativos revelem oportunidades específicas de melhoria.

2) ***Aceitação****:* A análise dos resultados do TAM evidencia uma forte tendência de concordância entre os participantes, especialmente nos itens relacionados à utilidade percebida (UP) e à facilidade de uso percebida (FUP). Os itens UP1, UP2 e UP3 apresentaram percentuais elevados de concordância, indicando que os usuários perceberam o chatbot como uma

TABELA II
SÍNTESE DAS ESTATÍSTICAS DESCRITIVAS E DA CONSISTÊNCIA INTERNA

Medida/construto	Itens	n	Média	DP	Mediana	α
SUS	10	59	75,7	17,6	77,5	—
UP	4	59	4,14	0,74	—	0,72
FUP	4	59	4,34	0,75	—	0,74
IUP	3	59	4,25	0,78	—	0,71
FSS-short*	9	59	4,12	0,81	4,22	0,92

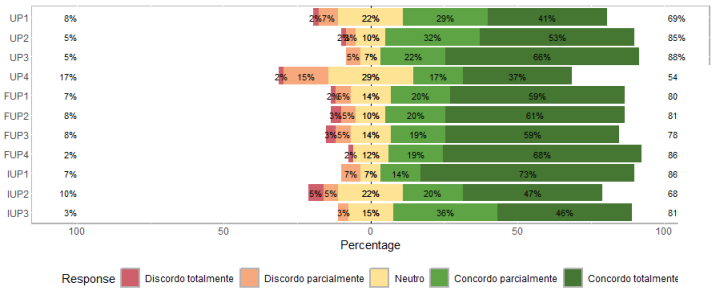


Fig. 3. Resultados do Questionário de Aceitação (TAM)

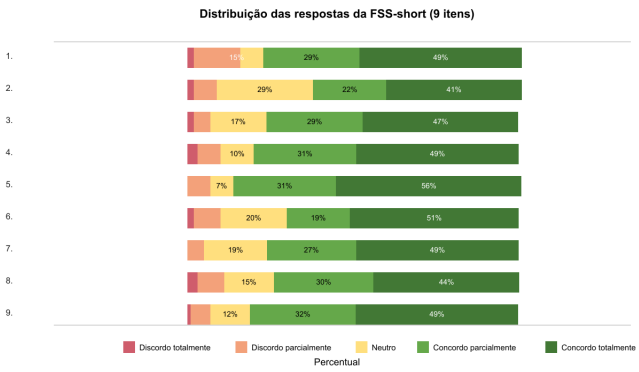


Fig. 4. Distribuição das respostas aos nove itens FSS-short (n = 59)

ferramenta útil para o agendamento do exame. De maneira semelhante, os itens FUP1, FUP2 e FUP3 demonstraram que os participantes consideraram o sistema simples de operar, reforçando a percepção de facilidade de uso.

Os itens associados à intenção de uso percebida (IUP) também apresentaram predominância de respostas favoráveis. As médias, os desvios-padrão amostrais dos escores individuais e os alfas de Cronbach de UP, FUP e IUP estão apresentados conjuntamente na Tabela II.

O item UP4, com 54% de concordância, é menos favorável que os demais itens de utilidade percebida e deve ser analisado em conjunto com a funcionalidade a que se refere, especialmente o agendamento e a consulta de informações.

*3) **Experiência de Fluxo**:* O questionário identificado na coleta como FSS-*short* foi aplicado para avaliar as sensações psicológicas dos participantes durante a utilização da ferramenta. A Figura 4 apresenta a distribuição das respostas desses nove itens, e as estatísticas da Tabela II.

As estatísticas descritivas e a consistência interna dos nove itens apresentados na Figura 4 estão reunidas na Tabela II. As respostas foram predominantemente favoráveis ao envolvimento durante a interação.

B. Análise qualitativa

As respostas à questão aberta foram processadas no IRAMUTEQ por meio de nuvem de palavras, análise de similitude e Análise Fatorial de Correspondência (AFC). O corpus preparado reuniu 47 textos (respostas não vazias), organizados em 47 segmentos de texto, com 1.929 ocorrências; o aproveitamento foi de 79,7% (47 de 59 respostas). Os resultados de cada procedimento são apresentados a seguir.



Fig. 5. Nuvem de Palavras

1) ***Nuvem de Palavras****:* A análise evidencia que as palavras de maior destaque, apresentadas na Figura 5, como *agendamento*, *chatbot*, *mensagem*, *muito*, *achar*, *bem* e *vez*, indicam que os usuários associaram fortemente a ferramenta ao processo de marcar consultas e à comunicação automatizada. Outros termos relevantes, como *cadastro*, *confirmar*, *informação* e *atendimento*, também aparecem no corpus. A ocorrência de palavras como *fácil*, *bom*, *erro* e *pessoa* mostra que as respostas incluem vocabulário relacionado a diferentes aspectos da interação; contudo, a frequência representada na nuvem, por si só, não determina a valência nem o sentido de cada termo. Em particular, *pessoa* não é interpretado isoladamente como indicação de dificuldade. Essa leitura é compatível com a AFC, que posiciona *pessoa* no quadrante superior esquerdo, separado de *intuitivo* e *geral*, no quadrante inferior esquerdo. Assim, a nuvem complementa os resultados quantitativos pela recorrência lexical, enquanto o sentido dos termos depende de sua distribuição e do contexto das respostas [13].



Fig. 6. Árvore de Similitude

2) ***Análise de similitude****:* A análise de similitude, apresentada na Figura 6, evidencia que o termo *agendamento* ocupa posição central na rede semântica, funcionando como núcleo das interações relatadas pelos participantes. As conexões mais fortes com palavras como *mensagem*, *vez*, *bem* e *achar* indicam que o processo de marcar consultas foi o aspecto mais recorrente e estruturador da experiência de uso do chatbot. Além disso, ramificações secundárias que incluem termos como *erro*, *pessoa* e *fácil* revelam tanto percepções positivas de simplicidade quanto relatos de dificuldades pontuais. A presença de palavras como *cadastro* e *confirmação* reforça a importância dos elementos funcionais da interação, enquanto *ideia* e *problema* apontam para reflexões críticas sobre o uso da ferramenta. Em conjunto, essa análise qualitativa complementa os resultados quantitativos, demonstrando que o chatbot Liz foi fortemente associado ao ato de agendar exames, mas também trouxe à tona nuances ligadas à clareza das mensagens, à facilidade de uso e às limitações percebidas pelos usuários.

3) ***Análise Fatorial de Correspondência****:* A AFC apresentada na Figura 7 distribui esses termos em quadrantes distintos: *ótimo* aparece no quadrante superior direito; *pessoa* e *até*, no superior esquerdo; e *intuitivo*, *geral*, *único* e *responder*, no inferior esquerdo, em posições diferentes entre si. Já *digitar* aparece no quadrante inferior direito. Portanto, esses termos não formam um único agrupamento. A figura permite descrever suas posições relativas nos eixos fatoriais, mas, isoladamente, não sustenta a atribuição de uma avaliação comum a palavras situadas em quadrantes diferentes.

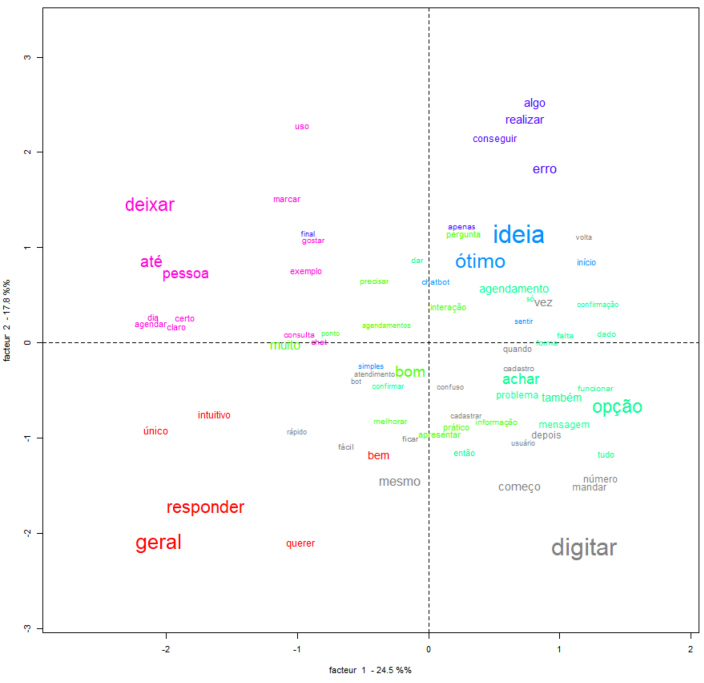


Fig. 7. Análise Fatorial de Correspondência

Os eixos fatoriais, que explicam 24,5% e 17,8% da variância, respectivamente, organizam duas dimensões principais do discurso: uma relacionada à avaliação da experiência e outra à operacionalidade da interação. A leitura deve ser restrita à posição efetivamente exibida na Figura 7: termos próximos apresentam associação ou proximidade no plano, enquanto termos distantes não devem ser tratados como evidência de dependência causal. Assim, a AFC sugere agrupamentos e contrastes lexicais, mas não permite, isoladamente, inferir relações de dependência entre as palavras.

a) Excertos das respostas abertas.: Para contextualizar os termos *erro*, *problema*, *digitar* e *responder*, apresentam-se cinco excertos da pergunta aberta. Os códigos P1–P59 foram atribuídos pela ordem das respostas na planilha; mantiveram-se a redação e a grafia originais, e as supressões estão indicadas por colchetes. Os trechos ilustram experiências relatadas e não representam, por si sós, a frequência de cada tema.

P2 "Tive alguns problemas com ter que mandar a mesma informação mais de uma vez. De resto, achei rápido e prático."

P10 "Apenas com alguns pontos a melhorar em relação ao tratamento de erro, e também acho que por se tratar de um bot de atendimento precisaria ser um pouco mais expressiva nas interações."

P11 "foi uma boa experiência, no começo apresentou erros ao responder minha primeira interação"

P16 "No começo fiquei confuso se era para digitar sim ou não, mas depois que li era para ser digitado apenas o valor do numero [...]"

P24 "Foi fácil, porém apresentou alguns erro: [...] Tive que digitar varias vezes que queria fazer o cadastro [...]"

Os relatos descrevem repetição de informações, dificuldades no tratamento de erros e dúvidas sobre as entradas esperadas. Em conjunto, ajudam a contextualizar os termos observados nas análises lexicais, sem substituir a leitura da distribuição completa das respostas.

VII. Discussão

O escore SUS médio de 75,7 (DP = 17,6; mediana = 77,5) situa o chatbot Liz acima do ponto de referência de 68 utilizado na literatura e distante da faixa adjetivada como boa por Bangor, Kortum e Miller [4]. Essa interpretação, junto às respostas aos itens negativos, aponta oportunidades de melhoria em facilidade de uso, necessidade de apoio e tratamento de erros. Holmes et al. [25], por exemplo, mostraram que o SUS é útil para uma visão global da usabilidade, mas não descreve sozinho aspectos conversacionais como continuidade do diálogo, tratamento de erros e adequação das respostas. No presente estudo, os termos "erro", "problema", "digitar" e "responder" complementam a medida global. Como os trabalhos utilizam amostras, tarefas e versões de instrumentos distintas, não é adequado estabelecer uma hierarquia direta de escores entre os chatbots; a comparação deve ser interpretada como convergência metodológica.

Os resultados do TAM também dialogam com estudos de agentes conversacionais que relacionam utilidade percebida, facilidade de uso, confiança e intenção de uso. Silva et al. [36] mostram que a aceitação pode depender de fatores adicionais ao TAM original, enquanto o estudo de Liz evidencia que a utilidade do agendamento e a simplicidade da interação são centrais para a percepção dos participantes. O resultado menos favorável de UP4 indica que a utilidade não é homogênea entre as funcionalidades: uma operação pode ser considerada simples, mas não necessariamente percebida como suficientemente necessária ou vantajosa. Essa distinção deve orientar a revisão do conteúdo e da apresentação das funções do chatbot.

Na dimensão do fluxo, a tendência favorável da FSS-short sugere que menus objetivos, metas claras, confirmação de etapas e retorno imediato podem favorecer concentração e sensação de controle em uma interação curta. Essa evidência é complementar, e não equivalente, à usabilidade ou à aceitação: uma interface pode ser fácil e útil sem produzir envolvimento, assim como uma interação envolvente pode esconder dificuldades operacionais. A articulação entre os três construtos permite, portanto, distinguir eficiência funcional, valor percebido e qualidade subjetiva da experiência. Ainda são necessários estudos com registros temporais e medidas comportamentais para verificar se a percepção de fluxo se mantém em uso prolongado.

Os achados permitem formular recomendações concretas para o reprojeto da interação. O Quadro III relaciona cada problema identificado à evidência disponível e a uma proposta de melhoria. Para as entradas não reconhecidas, a evidência combina o comportamento descrito na implementação e relatos de dificuldade de resposta; como não foram analisados logs, não se atribui cada relato a uma falha específica de reconhecimento. As recomendações também consideram desafios descritos em chatbots de saúde, como respostas repetitivas e necessidade de esclarecer termos médicos [43].

Há, portanto, convergência parcial entre os resultados quantitativos e qualitativos: o agendamento aparece como funcionalidade central e há percepções favoráveis de utilidade e envolvimento, enquanto o SUS e as respostas aos itens negativos apontam dificuldades de usabilidade. Os termos relacionados a digitação, respostas e erros ajudam a contextualizar essas dificuldades. A interpretação poderá ser aprofundada com análises estratificadas por características dos participantes e com registros de interação.

VIII. Limitações do Projeto

Embora a validação inicial do protótipo Liz tenha produzido resultados favoráveis em algumas dimensões e apontado dificuldades de usabilidade, algumas limitações precisam ser destacadas para orientar futuras pesquisas e aprimoramentos:

- **Tamanho da amostra:** O estudo contou com 59 participantes, o que restringe a generalização dos resultados para populações maiores e mais diversas.
- **Composição da amostra:** O estudo contou com 59 participantes. Também não foram reportadas as distribuições por escolaridade, familiaridade digital e vínculo com a atenção primária. Estudos futuros devem recrutar participantes do público-alvo e apresentar análises estratificadas.
- **Medidas autorrelatadas:** SUS, TAM e FSS-short capturam percepções declaradas e podem sofrer influência de memória, expectativa e desejabilidade social. Avaliações futuras devem combiná-las com taxa de conclusão, tempo por tarefa, erros, pedidos de ajuda e registros de conversa.
- **Questão aberta única:** A análise qualitativa baseou-se em uma única pergunta aberta, limitando a profundidade e a possibilidade de esclarecer termos como "erro" e "problema". Entrevistas, protocolos de pensar em voz alta e trechos dos diálogos podem ampliar essa explicação.

TABELA III
PROBLEMAS, EVIDÊNCIAS E RECOMENDAÇÕES DE MELHORIA PARA O REPROJETO

Problema	Evidência	Recomendação de melhoria
Usabilidade e orientação	No SUS, 86% discordaram que precisariam de ajuda técnica, 85% discordaram que precisariam aprender várias coisas antes de usar o sistema e 71% concordaram que o sistema era difícil de usar.	Organizar a conversa em etapas curtas, usar linguagem direta e menus claros, e incluir exemplos e ajuda contextual junto à opção solicitada.
Percepção de erros	No item 6 do SUS, 41% concordaram que o chatbot apresentava muitos erros; 31% ficaram neutros e 29% discordaram.	Revisar os fluxos com maior incidência de falhas, apresentar mensagens de erro claras e acionáveis e monitorar erros após cada nova versão.
Resolução de dúvidas (UP4)	54% concordaram que o chatbot tirou suas dúvidas sobre o agendamento, percentual inferior aos demais itens de utilidade percebida.	Ampliar e organizar as informações sobre agendamento e preparo para o exame, responder diretamente à dúvida e indicar o próximo passo ou um canal de apoio quando a resposta não estiver disponível.
Digitação e repetição de dados	P2 relatou ter de enviar a mesma informação mais de uma vez; P16 ficou em dúvida se deveria digitar "sim" ou "não"; P24 relatou repetir a entrada numérica durante o cadastro.	Mostrar o formato esperado antes da entrada, validar o dado no próprio passo e permitir corrigir ou repetir somente o campo necessário, preservando informações já confirmadas.
Entradas não reconhecidas	A implementação prevê resposta coringa quando a entrada não corresponde às opções; P10 pediu melhor tratamento de erros e P11 relatou erro ao responder à primeira interação. Não há logs para confirmar a causa dos relatos.	Na resposta coringa, informar quais opções são aceitas, oferecer tentativa novamente ou retorno ao menu e preservar o estado da conversa. Registrar entradas não reconhecidas para orientar ajustes posteriores.

- **Contexto controlado:** A aplicação foi realizada em ambiente específico e supervisionado, não refletindo integralmente as condições reais de uso. A avaliação em campo, com usuárias da atenção primária e diferentes níveis de letramento digital, é necessária para verificar a transferibilidade dos resultados.
- **Escopo funcional:** O chatbot foi projetado para funções básicas de lembrete e agendamento. Recursos mais avançados, como integração com sistemas de prontuário eletrônico ou suporte a múltiplos idiomas, ainda não foram implementados.
- **Dependência tecnológica:** O uso do WhatsApp como principal canal de comunicação pode limitar o alcance em grupos populacionais que não utilizam o aplicativo ou que possuem restrições de acesso à internet.
- **Avaliação longitudinal:** O estudo não contemplou análises de longo prazo sobre adesão contínua, impacto na cobertura do exame preventivo ou redução efetiva de casos de câncer do colo do útero. Essas consequências não podem ser inferidas dos escores de experiência e devem ser testadas com indicadores de saúde e seguimento temporal.

Essas limitações indicam a necessidade de novos estudos com amostras ampliadas, em diferentes contextos de aplicação e com funcionalidades adicionais, de modo a fortalecer a robustez dos resultados e ampliar o potencial de impacto do protótipo Liz na saúde pública.

IX. CONSIDERAÇÕES FINAIS

Este estudo teve como objetivo avaliar como os participantes percebem a usabilidade, a aceitação e a experiência de fluxo durante a interação com o chatbot Liz. Em resposta à pergunta de pesquisa, os dados indicam percepção favorável quanto à aceitação e ao envolvimento, mas o SUS apresentou média de 75,7 (DP = 17,6; mediana = 77,5), acima do ponto de referência de 68, embora persistam dificuldades específicas nos itens negativos. Esses resultados caracterizam uma validação inicial da experiência percebida, não uma demonstração de impacto clínico ou de adesão efetiva ao rastreamento.

A análise realizada por meio dos instrumentos SUS, TAM e FSS-*short* mostrou percepções favoráveis de utilidade e envolvimento, mas também apontou dificuldades de usabilidade, sobretudo nas respostas aos itens formulados negativamente. Assim, não se conclui que o chatbot seja globalmente simples de utilizar; os resultados indicam aspectos concretos a aprimorar antes de avaliar em campo sua contribuição potencial para a comunicação e as práticas preventivas.

Os resultados qualitativos complementam essas evidências. A nuvem de palavras destacou *agendamento*, e a análise de similitude o posicionou como núcleo da rede, conectado a termos recorrentes nas respostas. Na AFC, os termos não se concentram todos no mesmo quadrante: *ótimo* está no superior direito; *pessoa*, no superior esquerdo; *intuitivo* e *geral*, no inferior esquerdo; e *digitar*, no inferior direito. A posição de *pessoa* em quadrante distinto de *intuitivo* e *geral*, somada à ausência de valência inerente na frequência da nuvem, impede classificá-la isoladamente como avaliação positiva ou dificuldade. Essa distribuição indica posições distintas no plano fatorial e não deve ser resumida como um agrupamento único de avaliações positivas. A leitura da AFC, portanto, complementa as outras análises ao mostrar contrastes na organização lexical das respostas, sem estabelecer, por si só, relações causais ou uma avaliação global do chatbot.

Contribuições e trabalhos futuros. Para a Computação, o trabalho contribui para a avaliação de IHC conversacional ao articular SUS, TAM, FSS-short e análise lexical. O produto

dessa análise é o Quadro III, que relaciona problemas de usabilidade, percepção de erros, resolução de dúvidas, digitação e entradas não reconhecidas às respectivas evidências e recomendações de melhoria. Para a Saúde, oferece evidências preliminares sobre a aceitabilidade de um chatbot voltado a informações e agendamento do exame preventivo. A redução de barreiras de comunicação e o favorecimento da adesão são hipóteses que precisam ser verificadas com indicadores de serviço e saúde. Recomenda-se realizar avaliação em campo com usuárias da atenção primária; analisar logs e conversas para medir conclusão, erros, tempo e pedidos de ajuda; comparar subgrupos por idade, escolaridade e letramento digital; e conduzir estudo longitudinal que acompanhe agendamentos, comparecimento e adesão ao exame.

X. Agradecimentos

Agradecemos à Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul (FUNDECT) pelo suporte financeiro e incentivo à pesquisa, fundamentais para a realização deste estudo. Manifestamos também nossa gratidão ao Instituto Federal de Mato Grosso do Sul (IFMS) pelo apoio institucional e pela infraestrutura disponibilizada, que contribuíram decisivamente para o desenvolvimento e a consolidação deste trabalho.

XI. Declaração de disponibilidade dos dados da pesquisa

Caso explicitamente justificado, os dados de pesquisa estão disponíveis sob demanda. Já o código utilizado pode ser acessado em https://github.com/luizpicolo/projeto-chatbot-saude.

XII. Contribuição de autoria

Luiz Fernando Picolo: Conceptualization, Data curation, Formal analysis, Investigation, Software, Visualization, Writing – original draft. Manuella Silva Sampaio: Conceptualization, Methodology, Supervision, Writing – review & editing. Felipe de Mauro Paulo: Conceptualization, Methodology, Validation, Writing – review & editing. Renato de Souza Garcia: Data curation.

XIII. Conflito de interesses

Os autores declaram que não há conflito de interesses a mencionar.

XIV. Declaração sobre o uso de inteligência artificial

Em conformidade com o Art. 2º, inciso III, do Manual de Publicações da SBC, os autores declaram que o uso de Inteligência Artificial (IA) generativa foi realizado apenas para a revisão textual. Todo o uso de IA ocorreu sob supervisão dos autores, que assumem integral responsabilidade pelo conteúdo científico.

Referências

[1] ABU SHAWAR, Bayan; ATWELL, Eric. Chatbots: Are they Really Useful?. *Journal for Language Technology and Computational Linguistics*, v. 22, n. 1, 2007.

[2] DE ALMEIDA, Karina Isabel Vieira. Desigualdade social e câncer do colo do útero: uma revisão sistemática. *Anais do XVI Encontro Nacional de Pesquisadores em Serviço Social*, v. 16, n. 1, 2018. Disponível em: https://periodicos.ufes.br/abepss/article/view/23409. Acesso em: 21 maio 2026.

[3] BARBOSA, Simone Diniz Junqueira et al. *Interação humano-computador e experiência do usuário*. 1. ed. Rio de Janeiro: Simone Diniz Junqueira Barbosa, 2021. E-book.

[4] BANGOR, Aaron; KORTUM, Philip; MILLER, James. Determining what individual SUS scores mean: adding an adjective rating scale. *Journal of Usability Studies*, v. 4, n. 3, p. 114–123, 2009.

[5] BROOKE, John. SUS: A "quick and dirty" usability scale. In: JORDAN, Patrick W.; THOMAS, Bruce; WEERDMEESTER, Bernard A.; MCCLELLAND, Ian L. (ed.). *Usability Evaluation in Industry*. London: Taylor & Francis, 1996. p. 189–194.

[6] PINTO, Ig Ibert Bittencourt Santana et al. Psychometric properties of the Brazilian-Portuguese Flow State Scale Short (FSS-BR-S). *PLOS ONE*, v. 19, n. 2, e0286639, 2024. DOI: https://doi.org/10.1371/journal.pone.0286639.

[7] BRASIL. O que é Atenção Primária?. [S. l.]: [s. d.], 2020. Disponível em: https://aps.saude.gov.br/smp/smpoquee.

[8] BRASIL. Instituto Nacional do Câncer (INCA): Tipos de Câncer. Ministério da Saúde. [S. l.], 2022. Disponível em: https://www.gov.br/inca/pt-br/assuntos/cancer/tipos/. Acesso em: 3 maio 2026.

[9] BRASIL. Instituto Nacional do Câncer (Inca). Ministério da Saúde. *Estimativa 2023: incidência de câncer no Brasil*. [S. l.]: [s. d.], 2023. Disponível em: https://www.inca.gov.br/sites/ufu.sti.inca.local/files//media/document//estimativa-2023.pdf. Acesso em: 11 mar. 2026.

[10] INSTITUTO NACIONAL DE CÂNCER (INCA). *Estimativa 2026: incidência de câncer no Brasil*. Rio de Janeiro: INCA, 2026. 168 p. ISBN 978-65-88517-52-9.

[11] BRASIL. Resolução n.º 466, de 12 de dezembro de 2012: Dispõe sobre diretrizes e normas regulamentadoras de pesquisas envolvendo seres humanos. [S. l.]: [s. d.], 2013.

[12] BRASIL. Resolução n.º 510, de 7 de abril de 2016. Brasília: Ministério da Saúde, 2016.

[13] CAMARGO, Brigido Vizeu; JUSTO, Ana Maria. IRAMUTEQ: um software gratuito para análise de dados textuais. *Temas em Psicologia*, Ribeirão Preto, v. 21, n. 2, p. 513–518, dez. 2013. Disponível em: http://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1413-389X2013000200016&lng=pt&nrm=iso. Acesso em: 2 set. 2026. DOI: https://doi.org/10.9788/TP2013.2-16.

[14] CARVALHO, Lucas; LEIFHEIT, Bhruno Roan; RIVERO, Luis; NAKAMURA, Walter; GUERINO, Guilherme; VALLE, Pedro Henrique Dias; SOUZA, Renato; SILVA, Williamson. What Do We Know About Usability Evaluation for Chatbots?: A systematic mapping study. In: *Proceedings of the XX Brazilian Symposium on Information Systems (SBSI '24)*. Juiz de Fora, Brazil, 2024. ACM, New York, NY, USA. DOI: https://doi.org/10.1145/3658271.3658324.

[15] CSIKSZENTMIHALYI, M.; CSIKSZENTMIHALYI, I. S. (org.). *Beyond boredom and anxiety: The experience of play in work and games*. San Francisco: Jossey-Bass, 1975.

[16] CSIKSZENTMIHALYI, M. *Finding flow: The psychology of engagement with everyday life*. New York: Basic Books, 1997.

[17] DAVIS, Fred D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, v. 13, n. 3, p. 319–340, 1989. DOI: https://doi.org/10.2307/249008.

[18] GERHARDT, Margô; MONTAGNER, Graziela. *Pesquisa em Ciências Sociais: métodos e técnicas*. Porto Alegre: Editora UFRGS, 2009.

[19] INSTITUTO NACIONAL DE CÂNCER (INCA). *O que é câncer?*. [S. l.], 2026. Disponível em: https://www.gov.br/inca/pt-br/assuntos/cancer/o-que-e-cancer/. Acesso em: 4 set. 2026.

[20] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems*. ISO 9241-210:2019. Geneva: ISO, 2019.

[21] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *Ergonomics of human-system interaction – Part 11: Usability: Definitions and concepts*. ISO 9241-11:2018. Geneva: ISO, 2018.

[22] JACKSON, S. A.; EKLUND, R. C. Assessing flow in physical activity: The Flow State Scale-2 and Dispositional Flow Scale-2. *Journal of Sport and Exercise Psychology*, v. 24, n. 2, p. 133–150, 2002.

[23] JACKSON, Susan A.; MARTIN, Andrew J.; EKLUND, Robert C. Long and short measures of flow: the construct validity of the FSS-2, DFS-2, and new brief counterparts. *Journal of Sport and Exercise Psychology*, v. 30, n. 5, p. 561–587, 2008. DOI: https://doi.org/10.1123/jsep.30.5.561.

[24] LOURENÇO, Douglas Fabiano; CARMONA, Elenice Valentim; LOPES, Maria Helena Baena de Moraes. Translation and cross-cultural adaptation of the System Usability Scale to Brazilian Portuguese. *Aquichan*, v. 22, n. 2, e2228, 2022. DOI: https://doi.org/10.5294/aqui.2022.22.2.8.

[25] HOLMES, Samuel et al. Usability testing of a healthcare chatbot: can we use conventional methods to assess conversational user interfaces?. In: *Proceedings of the 31st European Conference on Cognitive Ergonomics*. New York: ACM, 2019. p. 207–214. DOI: https://doi.org/10.1145/3335082.3335094.

[26] MARTINS, I.; LIMA, V. M. R.; AMARAL-ROSA, M. P.; MOREIRA, L. J.; RAMOS, M. G. Handcrafted and software-assisted procedures for Discursive Textual Analysis: analytical convergences or divergences?. In: COSTA, A. P.; REIS, L. P.; MOREIRA, A. (Org.). *Computer Supported Qualitative Research*. Cham: Springer Nature Switzerland, 2020. v. 1068, p. 189–205.

[27] MARTINS, R. M. G.; ALVES, D. A.; ALVES, S. A. A.; MARTINS, F. J. G.; LIMA, L. G. A.; PEREIRA, M. L. D. Desenvolvimento e validação de chatbot para abordagem de lesões da hanseníase. *Ciência & Saúde Coletiva*, v. 31, n. 5, e01402026, 2026. DOI: https://doi.org/10.1590/1413-81232026315.01402026.

[28] NAZARETH, Rodrigo Trisoglino; DE ALMEIDA, José Júlio Gonçalves; BASTOS, Alder Thiago. Utilização do WhatsApp e o parecer CFM n.º 14/2017. *Revista da Universidade Ibirapuera*, 2020. Disponível em: https://www.ibirapuera.br/seer/index.php/rev/article/view/222. Acesso em: 21 maio 2026.

[29] PICOLO, L.; TOMBINI, G. F.; DOS SANTOS, R. A. A.; DE SOUZA, C.; ENZ, L. de G. UMA Proposta de Chatbot para Apoio à Conscientização e Prevenção do Câncer do Colo do Útero no Contexto do Sistema de Saúde Pública do Município de Nova Andradina. *Encontro Internacional de Gestão, Desenvolvimento e Inovação (EIGEDIN)*, v. 7, n. 1, maio 2024. Disponível em: https://periodicos.ufms.br/index.php/EIGEDIN/article/view/20590.

[30] PICOLO, Luiz Fernando; TOMBINI, Guilherme Ferreira; DOS SANTOS, Rafael de Andrade Albuquerque; DE AGUIAR, Gabriely Abreu; LEITE, Guilherme Fracalossi. Chatbot e a saúde: uma revisão bibliográfica sobre o uso de chatbots para o auxílio no tratamento do câncer do colo de útero. *Encontro Internacional de Gestão, Desenvolvimento e Inovação (EIGEDIN)*, v. 6, n. 1, 2022. Disponível em: https://trilhasdahistoria.ufms.br/index.php/EIGEDIN/article/view/16797.

[31] PICOLO, Luiz Fernando et al. *Projeto Fundect – Pictec: projeto-chatbot-saude*. GitHub, 2022–2026. Disponível em: https://github.com/luizpicolo/projeto-chatbot-saude.

[32] PREECE, Jennifer; ROGERS, Yvonne; SHARP, Helen. *Design de interação: além da interação humano-computador*. 3. ed. Porto Alegre: Bookman, 2013.

[33] RAMA, C. et al. Rastreamento anterior para câncer de colo uterino em mulheres com alterações citológicas ou histológicas. *Revista de Saúde Pública*, v. 42, n. 3, p. 411–419, 2008. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102008000300004&lng=pt&tlng=pt. Acesso em: 11 mar. 2026.

[34] RICHARDSON, Roberto J. *Pesquisa social: métodos e técnicas*. 4. ed. São Paulo: Atlas, 1999.

[35] SAURO, Jeff; LEWIS, James R. *Quantifying the user experience: practical statistics for user research*. 2. ed. Cambridge: Morgan Kaufmann, 2016.

[36] SILVA, S. C.; DE CICCO, R.; VLAČIĆ, B.; ELMASHHARA, M. G. Using chatbots in e-retailing: how to mitigate perceived risk and enhance the flow experience. *International Journal of Retail & Distribution Management*, v. 51, n. 3, p. 285–305, 2023. DOI: https://doi.org/10.1108/IJRDM-05-2022-0163.

[37] VALE, Diama Bhadra Andrade Peixoto do et al. Avaliação do rastreamento do câncer do colo do útero na Estratégia Saúde da Família no Município de Amparo, São Paulo, Brasil. *Cadernos de Saúde Pública*, v. 26, n. 2, p. 383–390, 2010. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-311X2010000200017&lng=pt&tlng=pt. Acesso em: 22 jun. 2026.

[38] VENTURA, Felipe. WhatsApp chega a 99% dos celulares no Brasil; Telegram cresce. [S. l.], 2020. Disponível em: https://tecnoblog.net/noticias/2020/02/27/whatsapp-chega-a-99-por-cento-celulares-brasil-telegram-cresce/. Acesso em: 4 jun. 2026.

[39] VENKATESH, Viswanath; DAVIS, Fred D. A theoretical extension of the Technology Acceptance Model: four longitudinal field studies. *Management Science*, v. 46, n. 2, p. 186–204, 2000. DOI: https://doi.org/10.1287/mnsc.46.2.186.11926.

[40] VENKATESH, Viswanath; MORRIS, Michael G.; DAVIS, Gordon B.; DAVIS, Fred D. User acceptance of information technology: toward a unified view. *MIS Quarterly*, v. 27, n. 3, p. 425–478, 2003. DOI: https://doi.org/10.2307/30036540.

[41] DENECKE, K.; MAY, R. Usability Assessment of Conversational Agents in Healthcare: A Literature Review. In: *Challenges of Trustable AI and Added-Value on Health*. Studies in Health Technology and Informatics, 2022. DOI: https://doi.org/10.3233/SHTI220431.

[42] LARANJO, L. et al. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association*, v. 25, n. 9, p. 1248–1258, 2018. DOI: https://doi.org/10.1093/jamia/ocy072.

[43] LE BONNIEC, Alice; SAUVAGET, Catherine; LUCAS, Eric; NASSIRI, Abdelhak; SELMOUNI, Farida. Design and validation of a chatbot-based cervical cancer screening decision aid for women experiencing socioeconomic disadvantage: user-centered approach study. *JMIR Cancer*, v. 11, e70251, 2025. DOI: https://doi.org/10.2196/70251.

[44] MILNE, D. et al. The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review. *Journal of Medical Internet Research*, 2020. DOI: https://doi.org/10.2196/20346.