

Tecnologias de Avaliação da Experiência do Usuário (UX) em Chatbots: Um Mapeamento Sistemático da Literatura

Maria Rocha¹, Pietra Soares¹, Luiz Caldeira^{1,2}, Paulo Souza^{1,2}
Ricardo F. Vilela³, Pedro H. D. Valle⁴, Williamson Silva^{1,2}

¹ Universidade Federal do Pampa - UNIPAMPA (Campus Alegrete), Alegrete, RS, Brasil

²PPGES (UNIPAMPA - Campus Alegrete), Alegrete, RS, Brasil

³Universidade Estadual de Campinas - Unicamp (FT), Limeira, SP, Brasil

⁴Instituto de Matemática e Estatística (IME-USP), São Paulo, SP, Brasil

⁵Universidade Federal do Cariri – UFCA, Juazeiro do Norte, CE, Brasil

marialuiza, pietrasoares, luizcaldeira}.aluno@unipampa.edu.br

paulosilas@unipampa.edu.br, rfvilela@unicamp.br,

pedrohenriquevalle@usp.br, williamson.silva@ufca.edu.br

Abstract. *The popularization of chatbots in areas such as healthcare, education, and customer service has raised a central question: how can we ensure that these interactions truly provide a satisfying user experience? To explore this topic, this study conducts a systematic mapping of the literature, focusing on the most widely used technologies and methods for evaluating User Experience (UX) with conversational agents. Based on the analysis of dozens of studies retrieved from recognized databases, trends and gaps have been identified in how UX has been measured in this context. The research shows, for example, the predominance of quick and objective evaluations, while also highlighting underexplored methods that could transform how we understand emotional engagement in chatbot interactions. The results offer insights that broaden the perspective on the challenges and possibilities for the future of this constantly evolving field.*

Resumo. *A popularização dos chatbots em áreas como saúde, educação e atendimento ao cliente tem levantado uma questão central: como garantir que essas interações realmente ofereçam uma experiência satisfatória ao usuário? Para explorar essa temática, este trabalho realiza um mapeamento sistemático da literatura, com foco nas tecnologias e métodos mais utilizados na avaliação da Experiência do Usuário (UX) com agentes conversacionais. A partir da análise de dezenas de estudos extraídos de bases reconhecidas, foram reveladas tendências e lacunas na forma como a UX tem sido medida nesse contexto. A pesquisa mostra, por exemplo, o predomínio de avaliações rápidas e objetivas, mas também aponta métodos pouco explorados que podem transformar a forma como entendemos o engajamento emocional em interações com chatbots. Os resultados oferecem insights que ampliam a visão sobre os desafios e as possibilidades para o futuro dessa área em constante evolução.*

1. Introdução

O uso de chatbots tem se expandido rapidamente em setores como atendimento ao cliente, saúde e educação, impulsionado pelos avanços em inteligência artificial e processamento de linguagem natural. Estima-se que esse mercado alcance mais de 1,9 bilhão de dólares até 2027, refletindo uma ampla adoção de chatbots por grandes corporações, principalmente devido à sua capacidade de automatizar interações e reduzir custos operacionais [Fortune Business Insights 2025].

No entanto, apesar da eficiência percebida, estudos como os de [Georgescu et al. 2018] e [Gumus and Cark 2021] indicam desafios relevantes relacionados à qualidade da experiência oferecida por esses sistemas, como baixa personalização, respostas genéricas e pouca sensibilidade a aspectos emocionais. Esses fatores impactam negativamente a Experiência do Usuário (UX), tornando a avaliação da interação humano–chatbot um aspecto central para seu sucesso e aceitação.

Ainda que a literatura apresente diferentes métodos e métricas para mensurar a UX, há pouca clareza sobre como esses instrumentos são aplicados, especialmente no que diz respeito à qualidade hedônica da experiência [Rapp et al. 2021, Rogers et al. 2023]. A ausência de um panorama consolidado compromete tanto o avanço científico quanto a prática profissional, dificultando decisões fundamentadas sobre o uso e o aprimoramento dessas tecnologias. Diante disso, este trabalho tem como objetivo identificar e categorizar as principais tecnologias¹. A proposta busca contribuir com a sistematização do conhecimento existente, apoiando desenvolvedores e pesquisadores no uso de estratégias de avaliação mais eficazes, alinhadas às necessidades reais dos usuários e às demandas do design centrado no ser humano [Georgescu et al. 2018].

2. Referencial Teórico e Trabalhos Relacionados

2.1. Experiência do Usuário (UX)

De acordo com [Stati and Sarmiento 2021], a Experiência do Usuário (do inglês, *User Experience* - UX), tornou-se um campo de estudos essencial no desenvolvimento de projetos digitais. A popularização da internet e o avanço das tecnologias móveis (dispositivos e infraestrutura digital) transformaram profundamente a forma como as pessoas interagem com *softwares*, aplicativos e serviços digitais. Interfaces que, inicialmente, eram estáticas e simples, hoje são dinâmicas, responsivas e cada vez mais focadas em proporcionar interações agradáveis, intuitivas e eficientes. Nesse contexto, a UX emerge como uma disciplina que se dedica a projetar experiências que atendam às necessidades, expectativas e emoções dos usuários.

A UX está diretamente relacionada à qualidade em uso definida pela norma ISO/IEC 25010:2011 (ISO/IEC 25010, 2011), que estabelece um modelo de qualidade de software. Segundo essa norma, a qualidade em uso abrange as seguintes características: (i) **Efetividade** - capacidade do usuário atingir seus objetivos com precisão e completude; (ii) **Eficiência** - uso adequado de recursos (como tempo e esforço) para alcançar os objetivos; (iii) **Satisfação** - percepção positiva do usuário em relação ao uso do sistema; (iv)

¹No contexto deste trabalho, o termo tecnologia será utilizado como uma generalização de métodos, técnicas, modelos, ferramentas e outros tipos de propostas elaboradas na área de IHC e ES para o *design* e avaliação de UX, conforme [Silva et al. 2015]

Segurança - minimização de riscos e de impactos negativos; (v) **Cobertura de contexto** - capacidade do sistema atender a diferentes contextos de uso.

A UX não se limita apenas à estética da interface, mas envolve aspectos funcionais, emocionais e contextuais da interação dos usuários com sistemas digitais. A (ISO/IEC 25010, 2011) reforça que produtos de software devem ser avaliados não apenas pela qualidade interna e externa (como desempenho e manutenibilidade), mas também pela qualidade em uso, que reflete diretamente na percepção dos usuários sobre o produto. Portanto, considerar os princípios de UX no desenvolvimento de sistemas, especialmente em soluções baseadas em chatbots, é fundamental para garantir que os usuários interajam de forma eficiente, segura, agradável e satisfatória. Segundo [Crudu and Team 2024], isso reduz os riscos de abandono da solução e contribui para o sucesso do produto no mercado, uma vez que problemas de usabilidade estão entre os principais motivos para o abandono de aplicativos, enquanto boas práticas de UX, como navegação intuitiva e personalização, impactam diretamente na retenção e no engajamento dos usuários.

2.2. Trabalhos Relacionados

O estudo conduzido por [Carvalho et al. 2024] aborda a avaliação de usabilidade em *chatbots*, contextualizado pelo crescente interesse em agentes conversacionais como ferramentas de suporte e interação eficientes. Motivado pela falta de consenso sobre práticas e critérios de avaliação de usabilidade para *chatbots*, o trabalho tem como objetivo caracterizar tecnologias e métodos de avaliação que possam melhorar a qualidade dos *chatbots* sob a perspectiva de usabilidade. A pesquisa foi conduzida por meio de um Mapeamento Sistemático da Literatura, identificando 27 estudos relevantes que incluem técnicas como questionários e métodos de teste de usabilidade. Como resultado, o estudo proporciona uma visão geral das práticas atuais e conclui com sugestões para o aprimoramento de metodologias e ferramentas na avaliação de *chatbots*, promovendo maior qualidade nas interações entre humanos e máquinas.

O estudo realizado por [Carla Tubin and de Marchi 2022] explora a experiência do usuário (UX) com agentes conversacionais, um tema motivado pela crescente popularidade desses agentes em diversas aplicações. O objetivo da pesquisa foi identificar e avaliar os métodos usados para medir a UX em interações com agentes conversacionais. Para isso, foi conduzida uma revisão sistemática da literatura com base em critérios do PRISMA, que resultou na seleção de 27 estudos de UX com agentes. A pesquisa revelou que a maioria dos estudos utiliza questionários próprios em vez de instrumentos validados, com pouca integração de avaliações antes, durante e depois do uso. Os resultados destacam a necessidade de métodos de avaliação mais específicos e validados para UX com agentes conversacionais, sugerindo o uso combinado de métodos para capturar melhor a percepção do usuário ao longo da interação.

O estudo de [Haugeland et al. 2022] investiga a usabilidade de *chatbots*, motivado pelo crescente uso desses agentes em diferentes áreas e pela importância de incorporar princípios de interação humano-computador (IHC) para melhorar a experiência do usuário. O objetivo do estudo é identificar e categorizar as técnicas de usabilidade aplicadas a *chatbots* por meio de um mapeamento sistemático, que analisou 170 referências, resultando em 21 estudos principais. A pesquisa organiza os resultados em quatro critérios principais: técnicas de usabilidade, características de usabilidade, métodos de pesquisa e

tipos de *chatbots*. Conclui-se que a usabilidade de *chatbots* é um campo emergente, onde a maioria dos estudos utiliza métodos informais, como questionários e testes de usabilidade, indicando a necessidade de experimentos mais formais para desenvolver diretrizes de design focadas na usabilidade.

Os estudos de [Carvalho et al. 2024], [Carla Tubin and de Marchi 2022] e [Haugeland et al. 2022] destacam o interesse crescente em métodos de avaliação de *chatbots*, com foco predominante na usabilidade e em aspectos pragmáticos de UX. Embora relevantes, esses trabalhos apontam que, apesar do uso disseminado de questionários e testes de usabilidade, essas técnicas carecem de padronização, validação e, sobretudo, de sensibilidade para captar dimensões emocionais e contextuais da interação.

Outro ponto recorrente é que, na maioria dos casos, as avaliações são pontuais e aplicadas após o uso, sem acompanhar a jornada completa do usuário. Essa limitação restringe o entendimento de como a experiência evolui ao longo do tempo e dificulta o desenho de melhorias iterativas. Também são raros os estudos que combinam abordagens qualitativas e quantitativas de forma integrada, o que poderia fornecer uma visão mais abrangente e profunda. Diferentemente dos trabalhos acima, este mapeamento sistemático coloca ênfase explícita na qualidade hedônica da experiência, investigando não apenas métricas tradicionais de desempenho e usabilidade, mas também técnicas e métodos capazes de capturar engajamento emocional, satisfação estética e prazer no uso. Além disso, este trabalho busca identificar e recomendar abordagens mistas (qualitativas e quantitativas) que possam orientar desenvolvedores e pesquisadores a incorporar, desde as fases iniciais do ciclo de vida, estratégias avaliativas mais completas e sensíveis às dimensões subjetivas da interação humano-*chatbot*.

3. Metodologia

Para abordar essa lacuna, foi conduzido um Mapeamento Sistemático da Literatura (MSL). O MSL tem como objetivo identificar quais tecnologias e métodos têm sido utilizados na literatura para avaliar a UX em *chatbots*, além de permitir que os pesquisadores tenham uma visão geral das tendências de pesquisa nessa área, destacando as principais abordagens e métodos empregados. De acordo com a metodologia estabelecida por [Mayr et al. 2022], o MSL segue um processo estruturado para garantir a cobertura adequada e a relevância dos estudos.

- **Escolha das Bibliotecas Digitais:** O primeiro passo consiste em selecionar as bibliotecas digitais mais adequadas, como *ACM Digital Library*, *Engineering Village*, *IEEE Digital Library*, *Scopus* e *Springer Link*, que foram utilizadas nesse mapeamento e que contêm estudos relevantes sobre a avaliação da experiência do usuário (UX) em *chatbots*.
- **Busca Sistemática da Literatura:** Com as bibliotecas selecionadas, é realizada uma busca detalhada utilizando palavras-chave e termos específicos relacionados à UX e *chatbots*. Essa busca visa garantir uma cobertura abrangente da literatura existente.
- **Seleção dos Estudos:** A partir dos resultados da busca, os estudos são filtrados com base em critérios de inclusão e exclusão, assegurando que apenas artigos que atendam aos requisitos definidos sejam analisados.
- **Extração de Dados:** Após a seleção dos estudos, os dados relevantes são extraídos, como as métricas de UX utilizadas, incluindo metodologias aplicadas e as

tecnologias adotadas.

- **Síntese dos Resultados:** Por fim, os dados coletados são sintetizados e organizados de maneira a identificar tendências, lacunas na literatura e oportunidades para futuras pesquisas sobre a avaliação da UX em *chatbots*.

Podemos ter uma visão melhor desses processos na Figura 1.

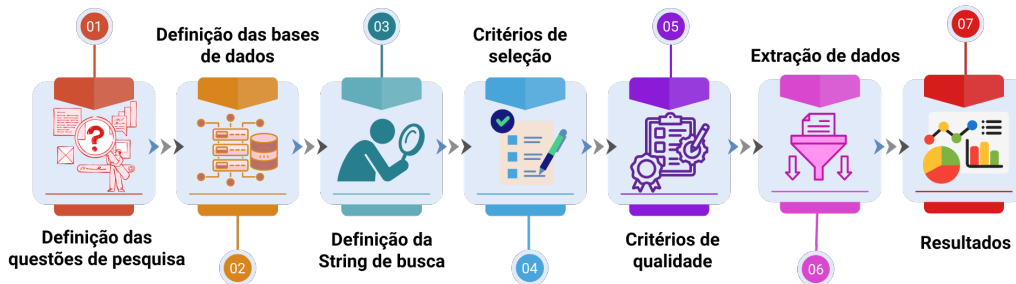


Figura 1. Metodologia do Mapeamento Sistemático

4. Planejamento do Mapeamento Sistemático

O mapeamento foi conduzido em março de 2023, antes da popularização massiva de sistemas baseados em LLMs como o ChatGPT, o que influencia diretamente o perfil dos estudos recuperados. O planejamento deste Mapeamento Sistemático da Literatura (MSL) seguiu as diretrizes propostas por [Kitchenham and Charters 2007], amplamente reconhecidas por sua aplicação em estudos baseados em evidências.

4.1. Definição das Questões de Pesquisa

A Tabela 1 apresenta as questões de pesquisa que orientaram este MSL.

Tabela 1. Questões de Pesquisa

ID	Questão de pesquisa
QP1	Qual a tecnologia relatada no artigo/publicação?
QP2	Qual a temporalidade da tecnologia relatada?
QP3	Em que fase do desenvolvimento a tecnologia é aplicada?
QP4	Qual a forma de interação do <i>chatbot</i> avaliado?

4.2. Processo de Busca

A busca foi realizada nas bases ACM Digital Library (ACM), IEEE Xplore Digital Library (IEEE), Scopus, Springer e Engineering Village (Compendex). A string de busca, apresentada abaixo, foi utilizada para pesquisar automaticamente em bibliotecas digitais, usando o operador booleano OR para combinar grafias alternativas e sinônimos. Simultaneamente, adotou-se o operador booleano AND para unir os três conceitos.

```
("UX"OR "user experience"OR "user-centered evaluation"OR "user centered evaluation") AND ("tool"OR "framework"OR "technique"OR "method"OR "model"OR "process"OR "guideline"OR "pattern"OR "metric"OR "approach"OR "inspection"OR "principle"OR "aspect"OR "requirement"OR "heuristic"OR "methodology"OR "mechanism"OR "evaluation"OR "assessment") AND ("chatbot"OR "chatterbot"OR "artificial conversational entity"OR "chatbots"OR "mobile chatbots"OR "conversational agent"OR "conversational interface"OR "conversational system"OR "dialogue system"))
```

4.3. Critérios de Inclusão e Exclusão

Os Critérios de Inclusão (CI) e Exclusão (CE) foram definidos para assegurar a relevância e qualidade dos estudos selecionados (ver Tabela 2).

Tabela 2. Critérios de Inclusão e Exclusão

ID	Descrição
CI1	Publicações que avaliem UX de agentes conversacionais baseados em texto (<i>chatbots</i>).
CI2	Publicações que apresentem tecnologias aplicáveis à avaliação de UX em <i>chatbots</i> .
CE1	Estudos que não tratem da UX em <i>chatbots</i> .
CE2	Artigos duplicados (considerando apenas a versão mais completa e recente).
CE3	Publicações do tipo livros, teses, dissertações, patentes, tutoriais ou pôsteres.
CE4	Publicações não escritas em inglês.
CE5	Estudos sem texto completo disponível para análise.

4.4. Critérios de Qualidade

Os critérios de qualidade (CQ) foram utilizados para avaliar a relevância dos estudos selecionados. Cada critério foi pontuado como: Sim (1), Parcialmente (0,5) ou Não (0). A soma total classifica os estudos em cinco níveis: Muito Ruim (0–1), Razoável (1,5–2), Bom (2,5–3), Muito Bom (3,5–4) e Excelente (4,5–5).

- CQ1: O estudo apresenta uma tecnologia para avaliação de UX em *chatbots*?
- CQ2: A descrição do uso da tecnologia é clara e detalhada?
- CQ3: O estudo fornece orientações sobre a aplicação da tecnologia em *chatbots*?
- CQ4: Os resultados obtidos com a aplicação da tecnologia são bem apresentados?

5. Resultados

O processo de mapeamento sistemático retornou 1052 estudos provenientes de diferentes bases de dados, como mostrado na Figura 5. Na primeira etapa, 322 estudos duplicados foram identificados e removidos, resultando em 730 artigos únicos. Em seguida, realizou-se a leitura dos títulos e resumos, utilizando os critérios de inclusão e exclusão, o que levou ao descarte de 611 estudos que não estavam alinhados ao objetivo da pesquisa. Após essa triagem, 119 estudos passaram para a análise dos critérios de qualidade. Na aplicação dos critérios de qualidade, 34 estudos foram excluídos por não atenderem às condições estabelecidas, reduzindo o conjunto para 85 artigos. Por fim, após a leitura completa dos textos, 66 estudos foram selecionados como válidos para a pesquisa.

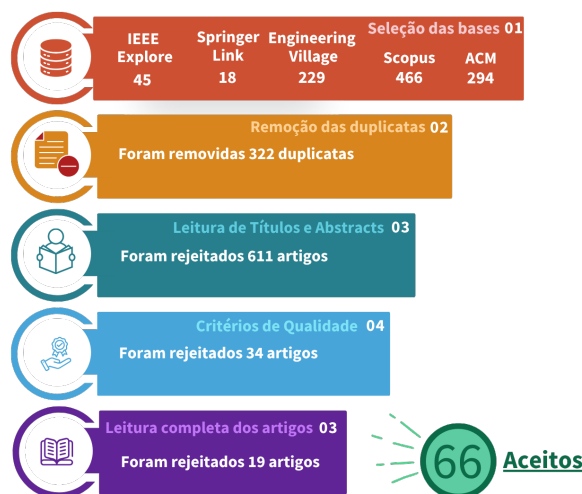


Figura 2. Filtros do Mapeamento Sistemático

5.1. Visão Geral dos Resultados

Nesta seção, são apresentados os resultados obtidos a partir da análise dos 66 estudos selecionados na revisão sistemática da literatura. A Tabela 3 sintetiza os dados coletados com base nas questões de pesquisa, categorizando as principais características metodológicas, tecnológicas e contextuais observadas nos estudos analisados. A seguir, são apresentados os resultados para cada uma das QPs delineadas.

Tabela 3. Resumo dos resultados das subquestões analisadas

Sub-Questão	Respostas	Quant.	%	Artigos
QP1	Scale	61	92,42	S01–S06, S07–S13, S14–S16, S17–S19, S21–S28, S30–S34, S36–S39, S40–S46, S48–S52, S54–S57, S58–S66
	Retrospective Analysis	6	9,09	S01, S14, S23, S31, S47, S66
	Controlled User Monitoring	19	28,79	S01, S06, S08, S10, S15, S24, S28, S29, S30, S31, S35, S39, S41, S45, S52, S53, S58, S60, S64
	Form	18	27,27	S04, S06, S10, S13, S15, S16, S23, S24, S26, S27, S28, S30, S33, S38, S54, S59, S64, S66
	Interview	11	16,67	S01, S05, S15, S26, S27, S30, S33, S47, S63, S64, S65
	Checklist	8	12,12	S07, S08, S16, S18, S20, S31, S39, S62
	Experience Sampling	2	3,03	S14, S23
	Log	6	9,09	S24, S30, S53, S58, S63, S65
	Framework	13	19,70	S06, S10, S15, S16, S20, S24, S27, S28, S30, S34, S38, S63, S65
	Others	2	3,03	S45, S58
QP2	Episódico	54	81,82	S01–S15, S16, S19–S37, S42–S45, S49–S52, S54–S61, S63–S66
	Longitudinal	2	3,03	S38, S53
	Episódico, Longitudinal	3	4,55	S23, S39, S62
	Não informado	7	10,61	S17, S18, S40, S41, S46–S48
QP3	VV&T	47	71,21	S01–S12, S14, S15, S19, S21, S23, S24, S26–S31, S33–S38, S44, S45, S49–S53, S56, S57, S59–S66
	Não informado	13	19,70	S16, S18, S20, S22, S39–S41, S43, S46, S47, S54, S55, S58
	Projeto	4	6,06	S13, S25, S32, S42
	Requisitos	1	1,52	S17
	Codificação	1	1,52	S48
QP4	PLN	39	59,09	S05, S07, S09, S10, S13, S14, S16, S17, S20, S23, S26–S30, S32, S35–S37, S42, S44, S45, S49–S53, S54, S56–S58, S60, S61, S63–S66
	NLU	33	50,00	S03, S07, S14, S18, S25, S27, S29, S31, S32, S34, S36, S38, S41, S42, S44, S45, S47, S48, S50–S53, S57, S58, S60–S66
	Gerador de Resposta	9	13,64	S01, S04, S08, S14, S15, S24, S32, S59, S66
	Gerenciador de diálogos	4	6,06	S03, S32, S38, S65
	Baseado em Regras	1	1,52	S20
	Não informado	11	16,67	S06, S11, S12, S21, S22, S33, S39, S40, S44, S48, S56

5.1.1. QP1 – Tipo de tecnologia

A análise evidencia uma predominância expressiva do uso de escalas padronizadas, como SUS, UEQ, AttrakDiff e MeCue, aplicadas majoritariamente de forma isolada. Essas tecnologias visam à mensuração objetiva da experiência do usuário, com foco em atributos como usabilidade, atratividade e satisfação. No entanto, o predomínio dessas ferramentas revela uma abordagem limitada, que tende a negligenciar a dimensão hedônica da experiência, reduzindo-a a impressões finais. Embora alguns estudos utilizem abordagens complementares, como entrevistas e protocolos qualitativos, esses casos são minoria. Essa concentração em métodos estruturados evidencia uma oportunidade para avanços metodológicos que explorem experiências mais situadas e emocionalmente significativas.

5.1.2. QP2 – Temporalidade

Os resultados revelam uma lacuna significativa nas abordagens longitudinais. A maioria das avaliações é pontual, realizada após o uso do *chatbot*, o que impossibilita compreender a experiência do usuário como um processo em evolução. Estudos como S23 e S52 constituem exceções relevantes, pois adotam coletas repetidas ou monitoramento de reações imediatas, respectivamente. No entanto, esses casos são minoria e, mesmo assim,

carecem de maior profundidade na análise temporal. A ausência de métodos que acompanhem o usuário ao longo do uso do sistema compromete a identificação de mudanças na percepção de usabilidade, satisfação e envolvimento e, por consequência, na proposição de melhorias iterativas.

5.1.3. QP3 – Fase do processo

A concentração de estudos na fase de Validação, Verificação e Testes (VV&T) reforça o caráter somativo das avaliações. Apenas quatro estudos foram realizados na fase de projeto e um na fase de requisitos. A ausência de investigações formativas evidencia uma falta de preocupação com a experiência do usuário ao longo do ciclo de desenvolvimento. Esse fator compromete o aprimoramento progressivo da interface e do comportamento dos *chatbots*, podendo resultar em soluções menos eficazes e menos engajadoras. A maioria dos estudos limita-se à validação final de sistemas já implementados, perdendo a oportunidade de integrar a perspectiva do usuário nas fases decisivas do design.

5.1.4. QP4 – Forma de interação

Nota-se uma adesão majoritária à utilização de Processamento de Linguagem Natural (PLN), presente em 59,09% dos estudos. No entanto, apenas metade dos sistemas analisados utiliza NLU, o que indica um uso potencialmente superficial da linguagem natural, sem compreensão real da intenção do usuário. Poucos estudos (13,64%) mencionam o uso de geradores de resposta, e apenas seis indicam o uso de gerenciadores de diálogo. Esses componentes são fundamentais para criar interações mais dinâmicas e naturais, capazes de sustentar conversas com mínima frustração. A ausência dessas informações compromete a análise da qualidade da experiência. Além disso, 11 estudos não especificam a forma de interação, limitando a reprodutibilidade dos resultados e dificultando a compreensão sobre os fatores que influenciam positivamente ou negativamente a UX.

5.2. Discussão dos Resultados

Os resultados apontam para um modelo de avaliação da UX com *chatbots* ainda centrado em práticas tradicionais e pouco contextualizadas. A predominância de escalas objetivas, aplicadas após a interação, indica um distanciamento entre a avaliação e a experiência real do usuário. Faltam métodos que capturem aspectos subjetivos, emocionais e contextuais da interação, o que compromete a qualidade das análises e a efetividade das melhorias implementadas com base nos resultados. A ausência de abordagens longitudinais também limita a compreensão da UX como processo. Interações com *chatbots* muitas vezes envolvem usos recorrentes, principalmente em contextos de suporte ou intervenção (educação, saúde, assistência). Avaliar apenas a impressão final, e não a jornada do usuário, restringe a capacidade de identificação de mudanças, de pontos críticos e de potencial de engajamento sustentado.

Ademais, a centralização da avaliação na fase de VV&T impede que os feedbacks dos usuários influenciem decisões de design, arquitetura e comunicação dos bots. Esta lacuna sugere que a UX não está, ainda, integrada à engenharia de software de forma iterativa. A baixa presença de componentes como geradores de resposta e gerenciadores

de diálogo reforça o caráter ainda restrito das interações, o que contradiz o potencial dos *chatbots* como sistemas inteligentes e adaptativos.

Observa-se também que, em muitos cenários práticos, os desenvolvedores optam por abordagens episódicas de avaliação por demandarem menos tempo de execução. Além disso, quando existem pressões de prazo, a primeira atividade a ser reduzida ou suprimida costuma ser a etapa de avaliação de qualidade, o que pode comprometer a experiência final do usuário.

Em conjunto, os dados revelam um campo em desenvolvimento, mas ainda carente de maturidade metodológica. O uso predominante de instrumentos clássicos e de aplicação pós-uso demonstra a necessidade de estratégias mais diversificadas, iterativas e sensíveis às dimensões subjetivas da experiência. Além disso, reforçam a relevância de propostas que busquem integrar a avaliação de UX desde os estágios iniciais do desenvolvimento, ampliando as possibilidades de entrega de produtos mais eficazes, agradáveis e centrados no usuário.

6. Conclusão

Este trabalho conduziu um mapeamento sistemático da literatura (MSL) com o objetivo de identificar, categorizar e analisar as tecnologias e métodos utilizados na avaliação da UX em *chatbots*. Com base em 66 estudos, observou-se um cenário dominado por abordagens quantitativas, pontuais e pós-intervenção, com ênfase excessiva em escalas e métricas de fácil aplicação, mas com pouca sensibilidade às dimensões emocionais e contextuais. A análise revelou que, embora o campo da UX em *chatbots* tenha evoluído em termos de ferramentas e *frameworks*, ainda persiste uma lacuna metodológica relevante: a subutilização de abordagens qualitativas, longitudinais e formativas. A experiência do usuário, como processo dinâmico e multifacetado, é muitas vezes reduzida a indicadores estáticos, ignorando momentos críticos da jornada de interação.

Do ponto de vista técnico, observa-se uso restrito de mecanismos conversacionais avançados, como gerenciadores de diálogo, NLU aprofundada e geração dinâmica de respostas, comprometendo a autenticidade e o engajamento das interações. Conclui-se que a avaliação de UX em *chatbots* carece de maior diversidade metodológica, rigor analítico e sensibilidade ao contexto de uso. Recomenda-se ampliar práticas que considerem a temporalidade da experiência, combinem métodos qualitativos e quantitativos e incorporem métricas validadas para capturar não apenas desempenho, mas também emoção, engajamento e valor percebido.

Como trabalhos futuros, pretende-se projetar um sistema de recomendação de tecnologias, métodos e ferramentas para apoiar profissionais e pesquisadores na escolha de abordagens avaliativas, considerando tipo de chatbot, domínio, dados desejados e recursos disponíveis. Assim, busca-se tornar a seleção de técnicas mais acessível, ágil e alinhada às melhores práticas, contribuindo para o desenvolvimento de *chatbots* eficientes, empáticos e centrados nas reais necessidades dos usuários.

Referências

Carla Tubin, J. P. M. R. and de Marchi, A. C. B. (2022). User experience with conversational agent: a systematic review of assessment methods. *Behaviour & Information Technology*, 41(16):3519–3529.

- Carvalho, L. V. d., Valle, P. H. D., Leifheit, B. R., Cabrejos, L. E. R., Nakamura, W., Guerino, G. C., Garcia, R. D. S., and Silva, W. (2024). What do we know about usability evaluation for chatbots?: A systematic mapping study. *ACM*.
- Crudu, A. and Team, M. R. (2024). How ux design influences app retention rates – key insights.
- Fortune Business Insights (2025). Chatbot market size, share, growth & trends analysis, report, 2032. Acesso em: 16 jun. 2025.
- Georgescu, A.-A. et al. (2018). Chatbots for education—trends, benefits and challenges. In *Conference Proceedings of eLearning and Software for Education (eLSE)*, volume 2, pages 195–200. "Carol I" National Defence University Publishing House.
- Gumus, E. and Cark, Z. (2021). Chatbots in customer service: Advantages and challenges in providing quality service. *Journal of Business Research*, 134:13–21.
- Haugeland, I., Følstad, A., Taylor, C., and Bjørkli, C. (2022). Understanding the user experience of customer service chatbots: An experimental study of chatbot interaction design. *International Journal of Human-Computer Studies*, 161:102788.
- International Organization for Standardization (2011). ISO/IEC 25010:2011 - Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models. Available at: <https://iso.org/standard/35733.html>.
- Kitchenham, B. and Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report.
- Mayr, P., Weigand, H., et al. (2022). Systematic mapping studies: A methodological framework for digital transformation research. *Journal of Business Research*, 140:134–145.
- Rapp, A., Curti, L., and Boldi, A. (2021). The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, 151:102630.
- Rogers, Y., Sharp, H., and Preece, J. (2023). *Interaction Design: Beyond Human-Computer Interaction*. Wiley.
- Silva, W., Valentim, N. M. C., and Conte, T. (2015). Integrating the usability into the software development process-a systematic mapping study. In *Proceedings of the 17th International Conference on Enterprise Information Systems*.
- Stati, C. and Sarmiento, C. (2021). *Experiência do usuário (UX)*. Editora Intersaberes.