



**UNIVERSIDADE FEDERAL DO MARANHÃO**  
**Programa de Pós-Graduação em Ciência da Computação**

**Malu Gabriele Silva Mafra**

***Desenvolvimento de Artefatos para Apoiar o Design e a  
Avaliação de Chatbots focando em Usabilidade e User Experience***

**São Luís**  
**2023**

Malu Gabriele Silva Mafra

# **Desenvolvimento de Artefatos para Apoiar o Design e a Avaliação de Chatbots focando em Usabilidade e User Experience**

Dissertação apresentada como requisito parcial para obtenção do título de Mestra em Ciência da Computação ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Avaliação da dissertação em 13 de Dezembro de 2023, em São Luís - MA:

---

**Prof. Dr. Luis Jorge Enrique Rivero Cabrejos**  
Orientador  
Universidade Federal do Maranhão

---

**Prof. Dr. Williamson Alison Freitas Silva**  
Examinador Externo  
Universidade Federal do Pampa

---

**Prof. Dr. Davi Viana dos Santos**  
Examinador Interno  
Universidade Federal do Maranhão

São Luís - MA  
2023



*Aos meus antepassados, que não tiveram a oportunidade de estudar.*

# Agradecimentos

A Deus, por tudo, principalmente pela proteção e força para que eu pudesse seguir nesta caminhada e concluir esta pesquisa.

Aos meus pais Maria Nazaré e Manoel, ao meu irmão Gustavo e ao meu noivo Jefferson, pelo apoio e compreensão durante a trajetória da minha pesquisa. Muitas oportunidades de momentos nossos juntos foram abdicados por causa deste trabalho. Obrigada por entenderem e me encorajarem a continuar.

A todos os professores pelos valiosos ensinamentos, com menção honrosa aos professores, Dr. Geraldo Braz Junior, Dra. Simara Rocha, Dr. Davi Viana e Dr. Ari Silva por concederem seu precioso tempo para ajudarem nesta pesquisa. Sem dúvidas, suas colaborações foram de grande importância.

A todos os que colaboraram e se dispuseram a ajudar, especialmente aos caríssimos Kennedy Nunes, Thamyla Lima e Kastney Frazão, assim como também agradeço a todos os participantes dos estudos experimentais que contribuíram para que esta pesquisa pudesse ser concluída.

À excelente psicóloga Giovanna Rodilha, pelo cuidado e profissionalismo com que me escutou e me auxiliou nos meus momentos de maior fragilidade durante o mestrado. Sua ajuda certamente foi imprescindível.

Aos professores, Dr. Williamson Silva e Dr. Davi Viana, por aceitarem o convite para integrar a banca de avaliação do meu estudo. Decerto seus comentários, sugestões e críticas serão valiosos para aprimorar minha pesquisa. Agradeço também aos notáveis professores, Dr. Geraldo Braz Junior e Dr. Ari Silva, por aceitarem ser suplentes na banca. Muito obrigada!

À Universidade Federal do Maranhão, pela oportunidade, ao PPGCC por todos os esclarecimentos e auxílios prestados, bem como também a todos os funcionários da instituição.

À CAPES, pelo incentivo aos cursos de pós-graduação stricto sensu no Brasil. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Finalmente, agradeço especialmente ao meu orientador, Dr. Luis Rivero, um profissional ético, competente e humano, que com muita paciência sempre me motivou a não desistir e a acreditar que tudo daria certo. Professor, eu tive muita sorte por tê-lo como orientador durante esta trajetória. Serei sempre grata por tudo e nunca terei palavras suficientes para expressar toda a minha admiração e gratidão.

# Resumo

Chatbots são softwares que se comunicam e interagem de forma semelhante à humana, por meio de linguagem natural, com usuários humanos. Essas ferramentas são úteis para diversas finalidades que variam desde atendimento ao cliente, vendas, educação, saúde, até entretenimento. Nos últimos anos os chatbots se popularizaram e observou-se um crescimento significativo no surgimento dessas ferramentas no mercado, principalmente dos chatbots textuais. Essa ascensão tem encorajado desenvolvedores a criarem suas próprias ferramentas, além de atrair os esforços de pesquisadores para esta área. Apesar do destaque, há necessidade de propor tecnologias capazes de apoiar desenvolvedores e equipes de desenvolvimento no processo de projetar e avaliar chatbots de qualidade, visto que embora estas ferramentas estejam em alta, as tecnologias para garantir a qualidade delas e a satisfação dos usuários, não tem acompanhado o ritmo. Diante disto, esta pesquisa de mestrado propõe desenvolver artefatos aplicáveis ao processo de design e da avaliação de chatbots, tendo como base atributos de qualidade relacionados à Usabilidade e User Experience (UX), identificados em revisões da literatura. O foco em Usabilidade e UX se deu devido ao impacto que estes aspectos têm na satisfação do usuário e na qualidade percebida do sistema. O primeiro artefato é o *checklist* de inspeção U2Chatbot, desenvolvido com o objetivo de auxiliar equipes de desenvolvimento no processo de identificação de defeitos em chatbots textuais, melhorando a qualidade dessas ferramentas antes do lançamento destas no mercado. Apesar da criação do *checklist*, observou-se que alguns desenvolvedores têm uma necessidade visual para compreender aspectos importantes que chatbots devem exibir e que para esse público apenas os itens de verificação do U2Chatbot não seriam suficientes. Assim, de forma complementar, foi desenvolvido o artefato DP-U2Chatbot, que se trata de um conjunto de recomendações de design de interface, contendo 21 instruções importantes para apoiar desenvolvedores no processo de construção de chatbots. Após o desenvolvimento das tecnologias propostas, estas foram submetidas às avaliações necessárias. Para avaliar o *checklist* U2Chatbot, foi conduzido um estudo experimental e os resultados indicaram que os participantes consideraram a tecnologia útil para a descoberta de defeitos em chatbots, no entanto, a facilidade de uso pode melhorar. Além disso, verificou-se que a experiência dos participantes não influenciou significativamente os resultados de eficácia e a eficiência da técnica. Para os padrões de design DP-U2Chatbot, realizou-se uma avaliação simples em relação à compreensibilidade e utilidade da tecnologia. Os resultados indicaram que no geral a tecnologia é fácil de compreender e possui utilidade para apoiar o design de chatbots, no entanto, os exemplos visuais de alguns padrões podem melhorar para ficarem mais claros aos desenvolvedores.

**Palavras-chave:** Checklist de Inspeção, Avaliação de Chatbots, Usabilidade, User Experience, Recomendações de Design.

# Abstract

Chatbots are software that communicate and interact in a human-like way, through natural language, with human users. These tools are useful for a variety of purposes ranging from customer service, sales, education and learning, healthcare, to entertainment. In recent years, chatbots have become popular and there has been significant growth in the emergence of these tools on the market, especially textual chatbots. This rise has encouraged developers to create their own tools, in addition to attracting the efforts of researchers to this area. Despite the emphasis, there is a need to propose technologies capable of supporting developers and development teams in the process of designing and evaluating quality chatbots, since although these tools are on the rise, technologies to guarantee their quality and user satisfaction are not has kept pace. Given this, this master's research proposes to develop artifacts applicable to the design and evaluation process of chatbots, based on quality attributes related to Usability and User Experience (UX), identified in literature reviews. The focus on Usability and UX was due to the impact that these aspects have on user satisfaction and the perceived quality of the system. The first artifact is the inspection *checklist* U2Chatbot, developed with the aim of assisting development teams in the process of identifying defects in text-based chatbots, improving the quality of these tools before their launch on the market. Despite the creation of *checklist*, it was observed that some developers have a visual need to understand important aspects that chatbots should display and that for this audience, only the U2Chatbot check items would not be enough. Thus, in a complementary way, the DP-U2Chatbot artifact was developed, which is a set of interface design recommendations, containing 21 important instructions to support developers in the process of building chatbots. After developing the proposed technologies, they were subjected to the necessary evaluations. To evaluate *checklist* U2Chatbot, an experimental study was conducted and the results indicated that participants found the technology useful for discovering defects in chatbots, however, ease of use could improve. Furthermore, it was found that the participants' experience did not significantly influence the efficacy results and efficiency of the technique. For the DP-U2Chatbot design patterns, a simple assessment was carried out regarding the understandability and usefulness of the technology. The results indicated that in general the technology is easy to understand and useful in supporting the design of chatbots, however, the visual examples of some patterns could be improved to make them clearer to developers.

**Keywords:** Inspection Checklist, Chatbots Evaluation, Usability, User Experience, Design Patterns.

# Lista de ilustrações

Figura 1.1 – Metodologia aplicada ao desenvolvimento e avaliação dos artefatos U2Chatbot e DP-U2Chatbot . . . . .	19
Figura 4.1 – Processo de Agrupamento de Atributos Semelhantes . . . . .	50
Figura 4.2 – Processo de Transformação de Atributo em Requisito e Item do Checklist . . . . .	51
Figura 4.3 – Apoio Ferramental do Checklist de Inspeção U2CHATBOT . . . . .	61
Figura 4.4 – Chatbot selecionado para a Prova de Conceito . . . . .	62
Figura 5.1 – Boxplot comparando o indicador de eficácia dos grupos com Baixa, Média e Alta Experiência em Inspeção . . . . .	76
Figura 5.2 – Boxplot comparando o indicador de eficiência dos grupos com Baixa, Média e Alta Experiência em Inspeção . . . . .	77
Figura 5.3 – Resultados sobre facilidade de uso (FU), utilidade (U) e intenção de uso futuro (IU) sobre o checklist U2CHATBOT . . . . .	79
Figura 7.1 – Modelo de Padrão de Design . . . . .	106
Figura 7.2 – Padrão de Design DP_FE 1 . . . . .	107
Figura 7.3 – Padrão de Design DP_FE 2 . . . . .	107
Figura 7.4 – Padrão de Design DP_ACE 1 . . . . .	108
Figura 8.1 – Padrão de Design DP_FE 2 . . . . .	116
Figura 8.2 – Padrão de Design DP_ACE 1 . . . . .	117
Figura 8.3 – Padrão de Design DP_EC 1 . . . . .	117
Figura 8.4 – Padrão de Design DP_EC 3 . . . . .	118
Figura 8.5 – Padrão de Design DP_H 1 . . . . .	118
Figura A.1 – Padrão de Design - Modelo . . . . .	132
Figura A.2 – Padrão de Design DP_FE 1 . . . . .	132
Figura A.3 – Padrão de Design DP_FE 2 . . . . .	133
Figura A.4 – Padrão de Design DP_ACE 1 . . . . .	133
Figura A.5 – Padrão de Design DP_ACE 2 . . . . .	134
Figura A.6 – Padrão de Design DP_EC 1 . . . . .	134
Figura A.7 – Padrão de Design DP_EC 2 . . . . .	135
Figura A.8 – Padrão de Design DP_F 1 . . . . .	135
Figura A.9 – Padrão de Design DP_F 2 . . . . .	136
Figura A.10 – Padrão de Design DP_F 3 . . . . .	136
Figura A.11 – Padrão de Design DP_F 4 . . . . .	137
Figura A.12 – Padrão de Design DP_H 1 . . . . .	137
Figura A.13 – Padrão de Design DP_H 2 . . . . .	138
Figura A.14 – Padrão de Design DP_H 3 . . . . .	138
Figura A.15 – Padrão de Design DP_H 4 . . . . .	139

Figura A.16–Padrão de Design DP_H 5 . . . . .	139
Figura A.17–Padrão de Design DP_RL 1 . . . . .	140
Figura A.18–Padrão de Design DP_VS 1 . . . . .	140
Figura A.19–Padrão de Design DP_PE 1 . . . . .	141
Figura A.20–Padrão de Design DP_AD 1 . . . . .	141
Figura A.21–Padrão de Design DP_CP 1 . . . . .	142
Figura A.22–Padrão de Design DP_P 1 . . . . .	142
Figura B.1 –Padrão de Design DP_FE 2 . . . . .	143
Figura B.2 –Padrão de Design DP_ACE 1 . . . . .	143
Figura B.3 –Padrão de Design DP_EC 1 . . . . .	144
Figura B.4 –Padrão de Design DP_EC 3 . . . . .	144
Figura B.5 –Padrão de Design DP_H 1 . . . . .	145
Figura B.6 –Padrão de Design DP_F 3 . . . . .	145

# Lista de tabelas

Tabela 2.1 – Termos aplicados à construção da String de busca . . . . .	25
Tabela 2.2 – Formulário de extração de dados . . . . .	26
Tabela 2.3 – Relação da Quantidade de Artigos . . . . .	27
Tabela 2.4 – Tabela resumida com Atributos de Qualidade identificados . . . . .	28
Tabela 2.5 – Tabela com Técnicas Descobertas na Primeira Revisão . . . . .	31
Tabela 2.6 – Tabela com o objetivo dos atributos de qualidade identificados . . . . .	32
Tabela 3.1 – Termos utilizados para a montagem da String de busca . . . . .	39
Tabela 3.2 – Formulário de Extração de Técnicas . . . . .	40
Tabela 3.3 – Relação da Quantidade de Artigos - Segunda Revisão . . . . .	41
Tabela 3.4 – Tabela com Resumo de Extração . . . . .	41
Tabela 3.5 – Tabela de Técnicas de Avaliação para Chatbots . . . . .	42
Tabela 3.6 – Tabela resumida com Atributos de Qualidade identificados . . . . .	44
Tabela 4.1 – Transformação de Requisitos em Itens do Checklist . . . . .	51
Tabela 4.2 – Checklist proposto a partir da combinação de atributos das revisões . . . . .	52
Tabela 4.3 – Reclamações encontradas nos Comentários de Usuários do Bing Chat With AI . . . . .	62
Tabela 5.1 – Experiência dos participantes do Estudo Experimental . . . . .	66
Tabela 5.2 – Roteiro de uso do chatbot Bing Chat com IA . . . . .	68
Tabela 5.3 – Questões utilizadas para avaliar o checklist de inspeção de chatbots conforme o modelo TAM . . . . .	70
Tabela 5.4 – Resultados quantitativos individuais e gerais sobre o tempo gasto, total de discrepâncias, defeitos encontrados e falsos positivos dos participantes do estudo experimental . . . . .	73
Tabela 5.5 – Estatística Descritiva da eficácia e eficiência considerando os valores de média, mediana, desvio padrão, erro padrão, mínimo e máximo . . . . .	74
Tabela 5.6 – Comparativo das Reclamações de Usuários com Defeitos encontrados pelos Inspetores . . . . .	79
Tabela 6.1 – Falsos positivos encontrados na discriminação sobre os itens do checklist U2CHATBOT . . . . .	85
Tabela 6.2 – Itens do checklist U2CHATBOT com potencial para falsos positivos . . . . .	87
Tabela 6.3 – Melhorias dos textos dos itens do U2CHATBOT . . . . .	88
Tabela 6.4 – Checklist U2CHATBOT após melhorias . . . . .	90
Tabela 6.5 – Itens Essenciais do Checklist U2CHATBOT . . . . .	97
Tabela 7.1 – Listagem de atributos utilizados para o desenvolvimento das recomen- dações de design DP-U2CHATBOT . . . . .	103
Tabela 7.2 – Análise de atributos encontrados nas aplicações web/aplicativos analisados	105

Tabela 8.1 – Caracterização dos Participantes . . . . .	110
Tabela 8.2 – Notas dadas pelos especialistas aos patterns considerando o aspecto Compreensibilidade . . . . .	112
Tabela 8.3 – Notas dadas pelos especialistas aos patterns considerando o aspecto Utilidade . . . . .	113
Tabela 8.4 – Notas dadas pelos especialistas aos critérios . . . . .	114
Tabela 8.5 – Padrões que devem ser priorizados conforme os avaliadores . . . . .	114



# Lista de abreviaturas e siglas

FU	<i>Facilidade de Uso Percebida</i>
IA	<i>Inteligência Artificial</i>
IBM	<i>International Business Machines</i>
IPA	<i>Intelligent Personal Assistant</i>
IU	<i>Intenção de Uso Futuro Percebida</i>
ISO	<i>International Organization for Standardization</i>
NLP	<i>Natural Processing Language</i>
PDF	<i>Portable Document Format</i>
SPSS	<i>Statistical Package for the Social Sciences</i>
SUS	<i>System Usability Scale</i>
TAM	<i>Technology Acceptance Model</i>
U	<i>Utilidade Percebida</i>
UX	<i>User Experience</i>

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>15</b>
<b>1.1</b>	<b>Contextualização</b>	<b>15</b>
1.1.1	Usabilidade e User Experience	16
<b>1.2</b>	<b>Problema e Justificativa</b>	<b>17</b>
<b>1.3</b>	<b>Objetivo Principal e Específicos</b>	<b>18</b>
<b>1.4</b>	<b>Metodologia</b>	<b>19</b>
<b>1.5</b>	<b>Trabalhos Relacionados</b>	<b>20</b>
<b>1.6</b>	<b>Organização do Documento</b>	<b>22</b>
<b>2</b>	<b>PRIMEIRA REVISÃO DA LITERATURA</b>	<b>23</b>
<b>2.1</b>	<b>Introdução do Capítulo</b>	<b>23</b>
<b>2.2</b>	<b>Planejamento do Protocolo da Revisão</b>	<b>24</b>
2.2.1	Questões de Pesquisa	24
2.2.2	Termos de Busca	24
2.2.3	Fontes de Busca	25
2.2.4	Critérios de Seleção	25
<b>2.3</b>	<b>Execução da Revisão</b>	<b>26</b>
<b>2.4</b>	<b>Resultados e Discussão</b>	<b>27</b>
2.4.1	QP: Quais atributos de qualidade de chatbots com foco em usabilidade e user experience estão disponíveis na literatura científica?	27
2.4.2	SQ1: Quais técnicas de avaliação de chatbots destacam-se durante a descoberta de atributos de qualidade?	31
2.4.3	SQ2: O conjunto de atributos foi proposto com qual objetivo?	32
<b>2.5</b>	<b>Conclusão</b>	<b>35</b>
<b>3</b>	<b>SEGUNDA REVISÃO DA LITERATURA</b>	<b>37</b>
<b>3.1</b>	<b>Introdução do Capítulo</b>	<b>37</b>
<b>3.2</b>	<b>Planejamento do Protocolo da Revisão</b>	<b>38</b>
3.2.1	Questões de Pesquisa	38
3.2.2	Termos de Busca	38
3.2.3	Fontes de Busca	39
3.2.4	Critérios de Seleção	39
<b>3.3</b>	<b>Execução da Revisão</b>	<b>40</b>
<b>3.4</b>	<b>Resultados e Discussão</b>	<b>41</b>
3.4.1	QP: Quais técnicas de avaliação estão disponíveis para avaliar a Usabilidade e UX de chatbots?	41

3.4.2	SQ1: Quais atributos de qualidade podem ser identificados nas técnicas de avaliação da Usabilidade e User Experience em chatbots? . . . . .	44
<b>3.5</b>	<b>Conclusão</b> . . . . .	<b>46</b>
<b>4</b>	<b>CRIAÇÃO DO CHECKLIST U2CHATBOT</b> . . . . .	<b>48</b>
<b>4.1</b>	<b>Introdução do Capítulo</b> . . . . .	<b>48</b>
<b>4.2</b>	<b>O Processo de Desenvolvimento do Checklist</b> . . . . .	<b>49</b>
4.2.1	Atributos de Qualidade focados em Usabilidade e User Experience Identificados em Revisões da Literatura . . . . .	49
4.2.2	Checklist de Inspeção U2Chatbot . . . . .	50
4.2.3	O Processo de Inspeção Usando o Checklist U2Chatbot . . . . .	60
<b>4.3</b>	<b>Prova de Conceito</b> . . . . .	<b>61</b>
<b>4.4</b>	<b>Considerações Finais</b> . . . . .	<b>64</b>
<b>5</b>	<b>AVALIAÇÃO DO CHECKLIST U2CHATBOT</b> . . . . .	<b>65</b>
<b>5.1</b>	<b>Introdução do Capítulo</b> . . . . .	<b>65</b>
<b>5.2</b>	<b>Descrição do Estudo Experimental</b> . . . . .	<b>66</b>
5.2.1	Participantes do Estudo . . . . .	66
5.2.2	Materiais Utilizados . . . . .	67
5.2.3	Execução da Inspeção do Chatbot . . . . .	71
5.2.4	Coleta dos Dados . . . . .	71
<b>5.3</b>	<b>Avaliação do Checklist de Inspeção U2CHATBOT</b> . . . . .	<b>72</b>
5.3.1	Avaliação Quantitativa . . . . .	72
5.3.2	Avaliação com o Modelo TAM . . . . .	78
5.3.3	Análise Comparativa dos Comentários de Usuários e Defeitos encontrados nas Inspeções . . . . .	78
5.3.4	Análise Qualitativa . . . . .	81
<b>5.4</b>	<b>Considerações Finais</b> . . . . .	<b>83</b>
<b>6</b>	<b>IMPLEMENTAÇÃO DE MELHORIAS</b> . . . . .	<b>84</b>
<b>6.1</b>	<b>Introdução do Capítulo</b> . . . . .	<b>84</b>
<b>6.2</b>	<b>Melhorias no Checklist U2CHATBOT</b> . . . . .	<b>84</b>
<b>6.3</b>	<b>Considerações Finais</b> . . . . .	<b>99</b>
<b>7</b>	<b>PADRÕES DE DESIGN (UI) PARA CHATBOTS</b> . . . . .	<b>101</b>
<b>7.1</b>	<b>Introdução do Capítulo</b> . . . . .	<b>101</b>
<b>7.2</b>	<b>O Processo de Desenvolvimento dos Padrões de Design</b> . . . . .	<b>102</b>
7.2.1	Atributos de Qualidade do Checklist U2CHATBOT . . . . .	102
7.2.2	Identificação dos chatbots reais para os exemplos visuais . . . . .	104
7.2.3	Proposta dos Padrões de Design . . . . .	105
<b>7.3</b>	<b>O conjunto de Padrões de Design DP-U2CHATBOT</b> . . . . .	<b>106</b>

7.4	Considerações Finais . . . . .	108
8	<b>AVALIAÇÃO DOS PADRÕES DE DESIGN . . . . .</b>	<b>109</b>
8.1	Introdução do Capítulo . . . . .	109
8.2	Descrição da Avaliação . . . . .	109
8.2.1	Questionário de Avaliação . . . . .	110
8.2.2	Execução e Coleta dos Dados . . . . .	110
8.3	Resultados . . . . .	111
8.4	Melhorias nos Padrões de Design DP-U2CHATBOT . . . . .	115
8.5	Considerações Finais . . . . .	118
9	<b>CONCLUSÃO E TRABALHOS FUTUROS . . . . .</b>	<b>120</b>
9.1	Epílogo . . . . .	120
9.2	Contribuições . . . . .	122
9.3	Perspectivas Futuras . . . . .	123
	<b>REFERÊNCIAS . . . . .</b>	<b>124</b>
	<b>APÊNDICES . . . . .</b>	<b>131</b>
	<b>APÊNDICE A – LISTA COMPLETA DOS PADRÕES DE DESIGN DP-U2CHATBOT . . . . .</b>	<b>132</b>
	<b>APÊNDICE B – PADRÕES DE DESIGN DP-U2CHATBOT REFI- NADOS . . . . .</b>	<b>143</b>

# 1 Introdução

*Este capítulo introduz esta dissertação, fornecendo a contextualização acerca do tema, expondo os objetivos, a motivação e a metodologia de pesquisa. Ademais, são detalhadas a estrutura e a organização deste documento no que concerne ao desenvolvimento de artefatos que auxiliem equipes de desenvolvimento no design e na avaliação de chatbots.*

## 1.1 Contextualização

De acordo com [Sharma, Goyal e Malik \(2017\)](#), chatbots são sistemas computacionais desenvolvidos para interagir com humanos em linguagem natural. Entretenimento, educação, saúde, atendimento ao cliente e negócios em geral são algumas finalidades deste tipo de software ([GEORGESCU et al., 2018](#); [SPERLÍ, 2020](#); [RAHMAN](#); [MAMUN](#); [ISLAM, 2017](#); [ROSRUEN](#); [SAMANCHUEN, 2018](#)). Conforme [Adamopoulou e Moussiades \(2020\)](#), [Motaung \(2022\)](#), o termo *chatbot* é utilizado amplamente e abrange os *bots* inteligentes, agentes conversacionais e assistentes pessoais inteligentes.

Utilizando geralmente a abordagem textual, chatbots podem integrar voz e animação gráfica que retratem humanidade ([RADZIWILL; BENTON, 2017](#)), além de serem projetados utilizando inteligência artificial ou apenas se baseando em regras preestabelecidas e palavras-chave ([GOMES et al., 2020](#)). Os Assistentes Pessoais Inteligentes - IPAs (*Intelligent Personal Assistants*) são uma categoria de chatbots ativados por voz que utilizam inteligência artificial e necessitam de um dispositivo para funcionar, a exemplo da ferramenta Alexa da Amazon, que necessita de um Echo Dot, segundo [Brill, Munoz e Miller \(2019\)](#). Apesar de os IPAs estarem em ascensão, os chatbots baseados em texto (*text-based chatbots*) ainda são os mais difundidos no mercado ([THORAT; JADHAV, 2020](#); [RAPP; CURTI; BOLDI, 2021](#)) e, segundo o estudo de [Ciechanowski, Przegalinska e Wegner \(2018\)](#), ainda são os preferidos dos usuários.

Nos últimos anos a popularidade dos chatbots aumentou, gerando uma alta demanda por essas ferramentas ([SHARMA, 2021](#); [CHAVES; GEROSA, 2021](#)). O Brasil é considerado um dos maiores mercados do mundo na utilização de agentes de conversação. Só no país, o número de chatbots ativos passou de 24 mil em 2020 a 117 mil em 2023, estimando-se que 64% dessas ferramentas sejam voltadas ao atendimento ao cliente em diversas áreas de negócios ([TIME, 2021](#); [TIME, 2023](#)). Esse crescimento gera a oportunidade de propor soluções que contribuam para que estas ferramentas sejam desenvolvidas com qualidade, satisfazendo os usuários que estão cada vez mais exigentes ([MUÑOZ; AVILA, 2019](#)).

Nota-se que os chatbots inteligentes lançados no mercado estão cada vez mais sofisticados. A utilização de aprendizado de máquina (*machine learning*) e conceitos da inteligência artificial (NLP, por exemplo), tem tornado estas ferramentas cada vez mais complexas e capazes de realizar múltiplas tarefas (SHARMA; JOSHI, 2020), (REICHERTS; PARK; ROGERS, 2022). Apesar disso, os chatbots disponíveis no mercado ainda falham em cumprir as expectativas dos usuários. Ao consultar os comentários de usuários de chatbots na loja de aplicativos *Google Play*, por exemplo, é possível perceber que ainda há um longo caminho a ser percorrido para que estas ferramentas satisfaçam os usuários. Os problemas relatados incluem baixo desempenho, interações pouco naturais, respostas inadequadas, dificuldade em compreender perguntas complexas e falta de opções de personalização. Essas questões podem levar os usuários à frustração e à insatisfação, prejudicando inclusive a reputação da empresa ou negócio que o chatbot representa.

Assim como ocorre com outros sistemas, é necessário que chatbots atinjam um certo nível de qualidade para serem realmente úteis para seus usuários (GUERINO; VALENTIM, 2020). Para isso, aplicar técnicas de avaliação considerando aspectos de Usabilidade e User Experience (UX) antes que o sistema seja inserido no mercado pode contribuir para a qualidade do software (MADAN; DUBEY, 2012).

Desta forma, esta pesquisa de mestrado visa contribuir com o processo de design e avaliação de chatbots, propondo artefatos focados em Usabilidade e User Experience que melhorem a qualidade dessas ferramentas do ponto de vista dos usuários finais. O primeiro artefato é um *checklist* de inspeção para identificação de defeitos em chatbots textuais. Esta técnica de verificação/avaliação não necessita de uma versão completa do sistema, podendo ser utilizada antes do *release* do software no mercado e dessa forma pode contribuir para a redução dos custos de correção dos defeitos identificados (FRAZAO et al., 2020), (ALSAYED; BILGRAMI; FOSTER, 2017). Além disso, há a proposta de desenvolvimento de um conjunto de Padrões de Design (guias) voltados ao processo de design de chatbots, fornecendo recomendações para auxiliar desenvolvedores no processo de design dessas ferramentas (GOMES et al., 2021).

### 1.1.1 Usabilidade e User Experience

Quando se trata da qualidade de um produto de software, dois conceitos se destacam por serem utilizados com frequência ao abordar sistemas interativos, de acordo com Cruz, Collazos e Granollers (2015): a Usabilidade e a Experiência do Usuário (*User Experience*). Segundo a ISO 9241-11, Usabilidade pode ser definida como a medida na qual um produto pode ser usufruído por usuários para atingir objetivos específicos com eficácia, eficiência e satisfação, em um determinado contexto de utilização. De acordo com Bevan, Carter e Harker (2015), a Usabilidade pode ser considerada o resultado de uma interação, isto é, uma combinação de fatores, e não uma propriedade em si de um produto.

User Experience (UX), por sua vez, é definida pela ISO 9241-210, como as percepções e reações de um indivíduo resultantes da interação com um produto, sistema ou serviço, incluindo todas as emoções, crenças, preferências, percepções, respostas físicas e psicológicas, comportamentos e realizações do usuário antes, durante e após o uso. Assim, UX pode ser influenciada por fatores como o contexto de uso, funcionalidade, performance, apresentação, comportamento interativo do sistema, além de fatores como personalidade, habilidades, experiências, estado interno e físico do próprio usuário (MIRNIG et al., 2015).

De acordo com Hassan e Galal-Edeen (2017), ambos fatores são considerados decisivos da qualidade de qualquer produto, sistema ou serviço destinado ao uso humano, podendo ser considerados indicadores de sucesso ou falha de produto, sistema ou serviço e estão fortemente relacionados (CRUZ; COLLAZOS; GRANOLLERS, 2015). O estudo de Kocaballi, Laranjo e Coiera (2018) afirma que em alguns casos a UX pode ser confundida até como a própria Usabilidade.

Considerando o exposto, observa-se que os conceitos apresentados desempenham um papel fundamental no sucesso e na aceitação de produtos de software pelos usuários, e no caso desta dissertação, os objetos de estudo são os chatbots textuais inteligentes. O projeto e a avaliação de chatbots inteligentes considerando aspectos de Usabilidade e UX podem resultar em ferramentas que proporcionem uma experiência satisfatória e envolvente aos seus usuários. Para alcançar a excelência em relação aos conceitos de usabilidade e UX, é fundamental identificar e entender os atributos de qualidade que melhoram a interação entre os usuários e os chatbots, visto que isso permitirá que desenvolvedores, designers e pesquisadores trabalhem em conjunto para propor tecnologias capazes de aprimorar a concepção e a implementação dessas interfaces conversacionais, bem como também sua avaliação.

## 1.2 Problema e Justificativa

Os chatbots provaram ser ferramentas úteis para diversas esferas de negócios e nos últimos anos se popularizaram. Embora já tenham sido propostas tecnologias para avaliar chatbots, o número de métodos focados especificamente na avaliação de aspectos relacionados a User Experience e Usabilidade ainda é baixo (KOCABALLI; LARANJO; COIERA, 2019), (LANGEVIN et al., 2021). De acordo com Hassan e Galal-Edeen (2017), tanto Usabilidade quanto User Experience são considerados fatores determinantes para o sucesso de um sistema, no entanto, nota-se que ainda há carência de tecnologias de avaliação e de apoio ao design de chatbots, que unam esses conceitos e que tenham sido submetidas às devidas avaliações experimentais (GUERINO; VALENTIM, 2020).

Considerando a alta demanda por agentes conversacionais, sua utilidade em várias áreas de negócio e a importância de incentivar que essas ferramentas sejam desenvolvidas

com a qualidade que os usuários esperam e merecem, esta pesquisa é motivada pela escassez de tecnologias que possam ser empregadas por equipes de desenvolvimento de software que almejem melhorar a qualidade de seus chatbots em relação à inspeção de software e ao processo de design da interface. Em vista disso, a questão de pesquisa motivadora desse trabalho é:

***“Como apoiar equipes de desenvolvimento de software e desenvolvedores no design e na avaliação de chatbots focados em Usabilidade e User Experience?”***

Por meio do desenvolvimento de um *checklist* de inspeção especializado para chatbots, focado em Usabilidade e User Experience, que norteie desenvolvedores e revisores de software durante uma avaliação para descoberta de defeitos, torna-se possível ajudar a descobrir possíveis requisitos ausentes de Usabilidade e UX no software desenvolvido, além de permitir a detecção precoce de problemas, evitando retrabalho, desperdício de tempo e necessidade de utilizar recursos extras para a correção de defeitos que poderiam ter sido evitados. Desta forma, o artefato contribuiria para a melhoria de um chatbot e conseqüentemente para o aumento da satisfação dos usuários. Do mesmo modo, com a concepção de recomendações de design (*patterns*) de interface para chatbots é possível apresentar sugestões de design úteis, já aplicadas em sistemas reais, para guiar desenvolvedores de software no design/construção de ferramentas melhores, mais usáveis, com design mais eficiente, que englobem requisitos importantes para garantir a qualidade do sistema.

### 1.3 Objetivo Principal e Específicos

Diante do contexto apresentado, o objetivo principal desta pesquisa é desenvolver artefatos para apoiar equipes de desenvolvimento, engenheiros de software e pesquisadores no processo de design e avaliação de chatbots, considerando os aspectos de Usabilidade e UX, visando contribuir com a melhoria da qualidade desses sistemas. Para isto, foram definidos os seguintes objetivos específicos:

- Identificar um conjunto de atributos de qualidade para chatbots, com foco nos conceitos de Usabilidade e User Experience, através da condução de estudos secundários.
- Desenvolver um *Checklist* de Inspeção focado em Usabilidade e User Experience para a avaliação de chatbots, que permita auxiliar na descoberta precoce de defeitos nesses sistemas.

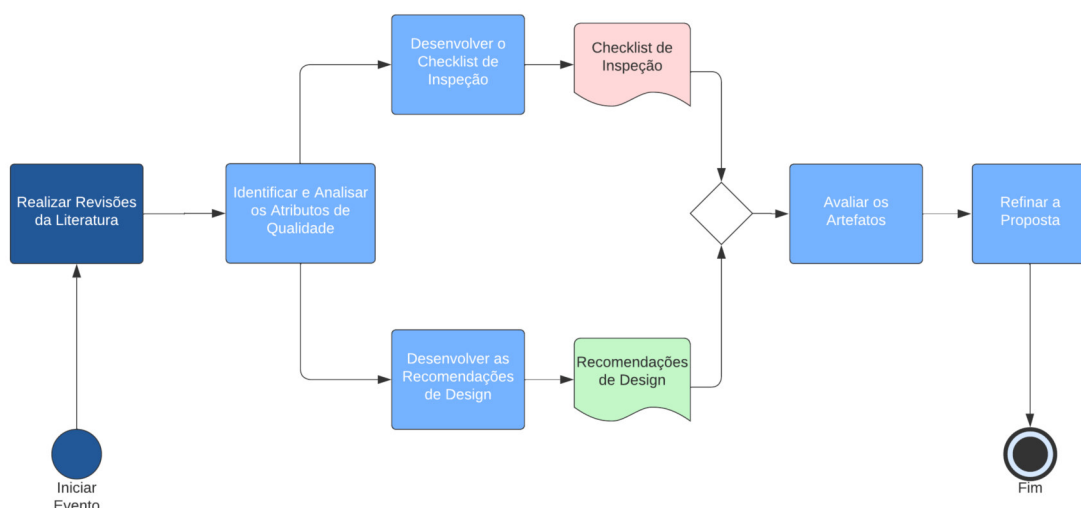


- Desenvolver um conjunto de Padrões de Design de interface, com recomendações e instruções válidas de design para apoiar e sanar dúvidas de desenvolvedores no design da interface de chatbots.
- Analisar a viabilidade das tecnologias desenvolvidas, através da condução de estudo experimental e avaliação de aspectos importantes, identificando oportunidades de melhorias das tecnologias e fornecendo evidências experimentais.

## 1.4 Metodologia

Adotou-se nesta pesquisa a metodologia sugerida por [Kitchenham e Charters \(2007\)](#) para a condução dos estudos secundários que possibilitaram a identificação de trabalhos acerca do tópico na literatura científica. Posteriormente, considerou-se a metodologia de [Kalinowski, Spínola e Travassos \(2004\)](#) para apresentar a proposta de uma tecnologia de inspeção de software voltada a avaliação de chatbots, bem como também considerou-se a metodologia de [Gomes et al. \(2021\)](#) para a proposição de tecnologia para apoiar o design de chatbots. Ante o exposto, a Figura 1.1 a seguir, apresenta as etapas metodológicas de pesquisa aplicadas ao alcance dos objetivos desta pesquisa.

Figura 1.1 – Metodologia aplicada ao desenvolvimento e avaliação dos artefatos U2Chatbot e DP-U2Chatbot



- **Condução de Revisões da Literatura:** Esta etapa é imprescindível e diz respeito a definição e condução de revisões da literatura, a fim de selecionar publicações que contenham atributos de qualidade relacionadas a Usabilidade e UX para chatbots.

- **Identificação e Análise de Atributos de Qualidade:** Após a realização dos estudos secundários, ocorre o processo de identificar e analisar os atributos de qualidade de chatbots descobertos, realizando o tratamento nos atributos iguais e repetidos, para então propor tecnologias úteis para melhorar a qualidade dessas ferramentas.
- **Desenvolvimento do *Checklist* U2CHATBOT:** Após a identificação dos atributos de qualidade, inicia-se o processo de desenvolvimento do primeiro artefato: o *checklist* de inspeção para avaliar chatbots. Todos os atributos de qualidade obtidos foram transformados em requisitos de software, tornando-se possível traduzi-los facilmente em itens de verificação da tecnologia de avaliação proposta.
- **Avaliação do *Checklist* U2CHATBOT:** Nessa etapa, definiram-se os procedimentos necessários para a realização de um estudo experimental visando verificar o desempenho e a viabilidade do *Checklist* de inspeção U2Chatbot, bem como também a relação entre as variáveis estabelecidas no estudo.
- **Refinamento e Melhorias:** Após os resultados do estudo experimental, focou-se em identificar e analisar os pontos positivos e negativos da tecnologia, traduzindo-os em possíveis melhorias a serem aplicadas ao *Checklist* de Inspeção U2CHATBOT. Este processo possibilitou gerar a versão final da tecnologia de inspeção de defeitos desenvolvida nesta pesquisa.
- **Criação dos Padrões de Design DP-U2CHATBOT:** Após os resultados do estudo experimental, verificou-se a oportunidade de desenvolver uma tecnologia complementar ao *Checklist* de Inspeção U2CHATBOT, contendo informações que visam auxiliar e guiar desenvolvedores no design de chatbots, as recomendações de design DP-U2CHATBOT.
- **Avaliação dos Padrões de Design DP-U2CHATBOT:** Nessa etapa, realizaram-se os procedimentos para avaliar a compreensibilidade e utilidade dos *patterns* DP-U2CHATBOT, visando verificar se estes podem realmente auxiliar desenvolvedores no processo de design de chatbots.

## 1.5 Trabalhos Relacionados

Nesta seção são apresentados os trabalhos relacionados com a proposta desta dissertação de mestrado, que visa melhorar a qualidade de chatbots com o desenvolvimento de um *checklist* de inspeção e de padrões de design com recomendações para auxiliar os desenvolvedores dessas ferramentas.

*Checklists* de Inspeção são ferramentas de avaliação que, segundo [Brykczynski \(1999\)](#), auxiliam revisores no processo de descoberta de defeitos em produtos de software.

Segundo [Cybis, Betiol e Faust \(2007\)](#), checklists bem elaborados reduzem a subjetividade normalmente associada a processos de avaliação e aos custos da avaliação, por se tratar de um método que não demanda pessoal especializado. Além disso, é possível utilizar técnicas de inspeção antes do lançamento de um produto de software no mercado, contribuindo para a redução de custos com correção dos defeitos identificados ([ALSAYED; BILGRAMI; FOSTER, 2017](#)).

O trabalho de [Sugisaki e Bleiker \(2020\)](#) propõe um *checklist* baseado nas regras propostas por Nielsen para avaliar usabilidade em interfaces conversacionais. As tradicionais Heurísticas de Nielsen foram detalhadas em 53 itens de verificação úteis para avaliar interfaces de conversação. A tecnologia proposta foi submetida a avaliação de 15 profissionais, que analisaram cada item e responderam a um questionário sobre a relevância de cada um deles e sobre o quão eficiente, agradável, conveniente e eficaz o checklist era.

A pesquisa de [Borsci et al. \(2021\)](#), por sua vez, também propôs um *checklist* para avaliar usabilidade em chatbots com inteligência artificial, o Bot-Check. A técnica com 42 itens de verificação, diferente do trabalho de [Sugisaki e Bleiker \(2020\)](#) não se baseou nas Heurísticas de Nielsen, mas utilizou atributos da revisão de literatura retirados do trabalho de [Radziwill e Benton \(2017\)](#). A lista de atributos foi validada em uma pesquisa online com designers de chatbots e usuários finais para descobrir quais itens seriam mantidos ou retirados da lista. Posteriormente, a escala Bot-Check foi validada com 141 participantes para identificar sua relevância.

Estes trabalhos relacionados demonstraram ser técnicas úteis e importantes para avaliar chatbots, no entanto, focam em avaliar principalmente Usabilidade, englobando poucos atributos de User Experience (UX). Esta dissertação de mestrado contrapõe estes trabalhos apresentados porque foca em ambos os conceitos, visto que tanto Usabilidade quanto UX estão fortemente relacionados e afetam a qualidade de um sistema.

Ainda acerca deste tópico, é possível destacar o trabalho de [Guerino e Valentim \(2020\)](#), que realizou um estudo de mapeamento na literatura científica a fim de identificar as técnicas de avaliação existentes para estas ferramentas. O estudo identificou 78 tecnologias capazes de avaliar a usabilidade e UX em interfaces conversacionais. As descobertas do estudo indicaram que 88,46% das técnicas encontradas focaram sua avaliação exclusivamente no critério Usabilidade. Cerca de 10,26% das técnicas eram focadas em avaliar exclusivamente UX e somente 1,28% das tecnologias identificadas focavam em avaliar de forma conjunta os conceitos de Usabilidade e UX. Os resultados indicaram a necessidade de desenvolvimento de tecnologias de avaliação de chatbots que unam os conceitos de Usabilidade e UX, que sejam submetidas às devidas avaliações experimentais. Contudo, apesar das excelentes descobertas e relevante contribuição, este trabalho relacionado foca em sistemas conversacionais ativados por voz, tais como Alexa e Google Assistente, o que difere do objetivo desta dissertação que foca em chatbots textuais

inteligentes.

No contexto de Padrões de Design, é possível encontrar trabalhos propondo estas recomendações de interface para aplicações de software, tal como o estudo de Nilsson (2009) que foca em padrões para aplicativos móveis. Para isso, o autor identifica problemas comuns em interfaces de usuário e apresenta uma coleção de padrões de projeto correspondentes voltados a estes aplicativos, além de destacar os prós e contras do uso dos padrões. O trabalho de Gomes et al. (2021), por sua vez, também apresenta um conjunto de padrões de design para aplicações de software, no entanto, os padrões são específicos para usuários autistas, com o intuito de contribuir para o desenvolvimento de aplicações móveis e web mais acessíveis a esse público. No entanto, no que diz respeito a Padrões de Design de interface voltados a agentes conversacionais, há lacunas a serem preenchidas, visto que as contribuições ainda são embrionárias. Desta forma, visando contribuir com diretrizes capazes de apoiar equipes de desenvolvimento no processo de construção de chatbots que satisfaçam seus usuários, esta dissertação propõe desenvolver um conjunto de recomendações de design.

## 1.6 Organização do Documento

Este trabalho está organizado da seguinte forma: o **Capítulo 2** descreve a condução e os resultados da primeira Revisão Sistemática da Literatura para identificação de atributos de qualidade para chatbots, com foco em Usabilidade e User Experience. O **Capítulo 3** descreve a condução e os resultados da segunda Revisão Sistemática da Literatura, realizada de forma complementar para descobrir outros atributos de qualidade relacionados a Usabilidade e User Experience para chatbots em técnicas de avaliação. O **Capítulo 4** apresenta o processo adotado para o desenvolvimento do *checklist* de inspeção U2CHATBOT para detectar defeitos em chatbots. O **Capítulo 5** apresenta o processo de condução e os resultados de um estudo experimental para averiguar a viabilidade da tecnologia de inspeção proposta. O **Capítulo 6** discute e apresenta as sugestões de melhorias, bem como a aplicação destas para o refinamento da técnica. O **Capítulo 7** expõe as etapas realizadas para a criação de um conjunto de recomendações de design para auxiliar desenvolvedores na construção de chatbots melhores, o artefato DP-U2CHATBOT. O **Capítulo 8** descreve uma avaliação realizada para examinar aspectos de compreensibilidade e utilidade da tecnologia proposta, bem como as melhorias realizadas. Finalmente, o **Capítulo 9** apresenta uma retrospectiva do trabalho, evidenciando sua contribuição. Além disso, apresentam-se as perspectivas para a continuidade da pesquisa.

## 2 Primeira Revisão da Literatura

*Este capítulo apresenta um estudo secundário realizado com o intuito de identificar na literatura científica os atributos de usabilidade e user experience importantes para chatbots. São detalhados o protocolo de revisão, a execução da revisão e os resultados obtidos. Estes atributos de qualidade podem ser úteis para propor tecnologias para avaliação e para apoiar o design destas ferramentas.*

### 2.1 Introdução do Capítulo

Um atributo de qualidade pode ser definido como uma propriedade de algo, e no contexto da Computação, atributos descrevem características que um sistema de software deve apresentar, determinando sua utilidade e sucesso do ponto de vista do cliente e do negócio (LARSSON, 2004). Essas propriedades podem ser utilizadas para propor técnicas de avaliação de software, como ocorreu no trabalho de Frazão et al. (2021), que utilizou atributos de qualidade para desenvolver o IC-MEG, *checklist* para melhorar a qualidade de jogos educacionais móveis.

No contexto dos chatbots, o trabalho de Radziwill e Benton (2017) afirma que a identificação de atributos de qualidade é uma estratégia interessante, visto que estas propriedades podem ser usadas por equipes de desenvolvimento como uma lista de verificação (*checklist*) para a certificação de que o sistema abordou questões essenciais. O estudo em questão, apresenta uma das revisões da literatura mais populares sobre atributos de qualidade relacionados a agentes de conversação, descobrindo atributos em 42 publicações de 1990 e 2017, entretanto, apesar da relevância, a publicação não detalha o quão a rigorosa foi a condução da revisão sistemática, já que a *string* apresentada está avulsa e as publicações selecionadas incluem documentos da literatura cinzenta (*grey literature*). Além disso, o trabalho não focou em identificar atributos relacionados especificamente com Usabilidade e UX, que como já visto, são conceitos que impactam a qualidade de um software.

Considerando o exposto, este capítulo apresenta um estudo secundário na literatura científica para identificar atributos de qualidade de chatbots textuais focados em Usabilidade e User Experience. O restante deste capítulo está organizado da seguinte forma: a Seção 2.2 apresenta o planejamento do protocolo da revisão da literatura; a Seção 2.3 apresenta a execução da revisão da literatura; a Seção 2.4 apresenta os resultados encontrados; e finalmente a Seção 2.5 traz as conclusões deste capítulo.

## 2.2 Planejamento do Protocolo da Revisão

Nesta seção detalha-se o planejamento do protocolo de estudo secundário para descobrir os atributos de qualidade focados em usabilidade e user experience importantes para chatbots. Inclui-se neste planejamento as questões de pesquisa, os termos da *string* de busca, a base de dados digital escolhida e os critérios utilizados para a seleção de artigos. As subseções a seguir apresentam os detalhes desse protocolo.

### 2.2.1 Questões de Pesquisa

O objetivo dessa revisão é identificar um conjunto de atributos de qualidade importantes para o projeto de chatbots, mais precisamente, estes atributos devem estar alinhados com os conceitos de Usabilidade e User Experience. A realização do presente estudo secundário justifica-se pela ausência de um estudo anterior que respondesse à questão principal de pesquisa:

- **QP** - *“Quais atributos de qualidade de chatbots com foco em usabilidade e user experience estão disponíveis na literatura científica?”*

Além da questão principal, buscou-se responder às seguintes subquestões (SQ):

- **SQ1** - *Quais técnicas de avaliação de chatbots destacam-se durante a descoberta de atributos de qualidade?*
- **SQ2** - *O conjunto de atributos de qualidade foi proposto com qual objetivo?*

### 2.2.2 Termos de Busca

Os termos utilizados para a montagem da *String* de busca são apresentados nesta subseção. Após a definição da questão e subquestões de pesquisa, é necessário elaborar uma *String* de busca que retorne as publicações científicas que sirvam para respondê-las, no caso desta revisão, uma *String* de busca capaz de retornar publicações contendo atributos de usabilidade e user experience para chatbots textuais. Para isto foram utilizados sinônimos e palavras-chave derivados dos trabalhos de Coppola e Ardito (2021), Suhaili, Salim e Jambli (2021), Guerino e Valentim (2020), Cabrejos, Viana e Santos (2018). Além disso, foram executados testes para validar a *String* conforme a questão de pesquisa principal, com o intuito de atestar sua capacidade de retornar o maior número de publicações relevantes para essa revisão da literatura. Como resultado, a Tabela 2.1 apresenta as palavras-chave adotadas com seus respectivos sinônimos.



Tabela 2.1 – Termos aplicados à construção da String de busca

Palavras-chave	Sinônimos
attribute	feature; characteristic; aspect; heuristic; principle; requirement
chatbot	conversational user interface; conversational agent
usability e user experience	usable; UX

Com base nos termos de busca, a *String* de busca base resultante foi: (“attribute” OR “feature” OR “characteristic” OR “aspect” OR “heuristic” OR “principle” OR “requirement”) AND (“chatbot” OR “conversational user interface” OR “conversational agent”) AND (“usability” OR “usable” OR “user experience” OR “UX”). Esta *String* foi executada na base de dados digital detalhada a seguir.

### 2.2.3 Fontes de Busca

As publicações analisadas foram retiradas da *Scopus Digital Library*, que foi escolhida por se tratar de uma base de dados científica robusta e confiável, sobre tecnologia e ciência, reconhecida como o maior banco de dados de resumos e citações da literatura com revisão por pares e que indexa resultados relevantes sobre usabilidade, de acordo com [Codina \(2005\)](#). Além disso, a biblioteca *Scopus* indexa publicações que estão presentes em outras bibliotecas como a *ACM Digital Library*, *Spring Link Digital Library*, *Science Direct Digital Library* e *IEEE Xplore Digital Library*.

### 2.2.4 Critérios de Seleção

Os trabalhos considerados nessa revisão da literatura são publicações que passaram por revisão por pares, sendo publicados em periódicos e/ou eventos científicos. Para garantir que fossem selecionadas as publicações com maior relevância a responder às questões de pesquisas desta revisão, foram definidos os seguintes critérios para incluir ou excluir uma publicação. É importante frisar que para a publicação seja aceita, todos os Critérios de Inclusão devem ser atendidos. Já em relação aos Critérios de Exclusão, se apenas 1 (um) for identificado, a publicação é descartada.

Os Critérios de Inclusão (CI) definidos foram:

- **CI1:** A publicação aborda atributos relacionados a usabilidade e ux para chatbots.

Por sua vez, os Critérios de Exclusão (CE) foram:

- **CE1:** A publicação não aborda atributos relacionados a usabilidade e ux para chatbots.

- **CE2:** A publicação aborda atributos relacionados a usabilidade e ux, mas não para chatbots textuais.
- **CE3:** A publicação não é um artigo científico, e sim um resumo de minicurso, introdução a uma conferência, etc.
- **CE4:** A publicação não está escrita no idioma Inglês ou Português.
- **CE5:** A publicação não está disponível para leitura e coleta dos dados (artigos pagos ou não disponibilizados pela máquina de busca).
- **CE6:** A publicação está duplicada.

Visando facilitar a etapa de execução da revisão, foi desenvolvido um formulário de extração de dados, apresentado na Tabela 2.2, para apoiar a condução da seleção dos artigos para a pesquisa. Os campos do formulário e suas respectivas descrições foram projetados seguindo as recomendações sugeridas por Cabrejos, Viana e Santos (2018). O formulário de extração é utilizado para facilitar a análise das informações retiradas das publicações que serão consideradas nos resultados.<sup>1</sup>

Tabela 2.2 – Formulário de extração de dados

Campos	Descrições
Objetivo	Escrever o objetivo do chatbot ou contexto para o qual foi proposto o conjunto de atributos
Atributos de Qualidade	Descrever os atributos de qualidade de Usabilidade e User Experience encontrados na publicação e, se existentes, suas respectivas categorias
Fundamentação	Escrever em que teoria/conceito a publicação se baseou para que os atributos fossem propostos
Técnica Vinculada	Escrever se o conjunto de atributos é resultado de alguma técnica de avaliação, descrevendo-a em caso positivo

## 2.3 Execução da Revisão

A obtenção dos artigos dessa pesquisa ocorreu executando a *String* definida nos motores de busca da biblioteca digital da *Scopus* no período de janeiro a fevereiro de 2022, a *String* retornou ao todo 185 publicações. A relação completa da quantidade de artigos está contida na Tabela 2.3.

<sup>1</sup> Formulário de Extração de Dados: <https://tinyurl.com/FormularioExtracao>



Tabela 2.3 – Relação da Quantidade de Artigos

<b>Biblioteca</b>	<b>Qtd. Geral</b>	<b>Filtro 1</b>	<b>Filtro 2</b>
Scopus	185	83	18

Os 185 artigos retornados foram submetidos ao primeiro filtro, onde se executou a leitura do título e o resumo de cada publicação, visando descartar as publicações que não se relacionavam à questão de pesquisa da revisão da literatura. Após este processo, restaram 83 publicações aptas ao segundo filtro, que foram lidas por completo e submetidas aos critérios de inclusão e exclusão definidos para a pesquisa, eliminando aquelas que não atendiam aos critérios estabelecidos. Após a aplicação dos filtros, restaram 18 publicações que atendiam a todos os critérios e, portanto, estavam aptas a extração das informações.

Com a finalização do processo de análise das publicações, foi possível identificar aquelas que possuíam os atributos focados em ambos os conceitos pesquisados (Usabilidade e User Experience), no entanto, também foram identificadas publicações que possuíam apenas atributos de usabilidade ou apenas atributos de user experience. Além disso, observou-se que algumas publicações combinavam atributos de qualidade genéricos (úteis para qualquer tipo de chatbot) com atributos específicos (atributos voltados especialmente a certos tipos de chatbot, como chatbots musicais ou de museus, por exemplo). Em vista disso, para esta pesquisa extraíram-se apenas os atributos genéricos, aptos a contribuir com tecnologias capazes de englobar o maior número possível de chatbots.

## 2.4 Resultados e Discussão

Nesta seção são apresentados os resultados da revisão realizada, bem como sua discussão. As publicações selecionadas durante a revisão da literatura resultaram em diversos atributos de qualidade relacionados a Usabilidade e User Experience para chatbots. Nas subseções a seguir, serão discutidas as questões de pesquisa da revisão da literatura, conforme os resultados obtidos.

### 2.4.1 QP: Quais atributos de qualidade de chatbots com foco em usabilidade e user experience estão disponíveis na literatura científica?

Após a conclusão desta revisão da literatura, obteve-se uma lista com 313 atributos de qualidade genéricos de Usabilidade e User Experience para chatbots. Os atributos genéricos são aqueles que abrangem a maioria dos chatbots, por exemplo, um atributo sobre facilidade de uso, importante para todos os tipos de chatbots. Há também os atributos específicos, que são restritos a um determinado tipo de chatbot. Por exemplo, um atributo sobre como deve ser realizada uma anamnese é um atributo específico para

chatbots de saúde voltado ao atendimento a pacientes. Ante o exposto, para esta pesquisa foram considerados atributos de qualidade genéricos (ou gerais), que podem ser utilizados para desenvolver tecnologias capazes de englobar muitos tipos de chatbot. É importante destacar também que as publicações selecionadas nesta revisão, embora estejam focadas nos conceitos de Usabilidade e User Experience, não fazem distinção clara do que é atributo de Usabilidade e UX, apenas há um indicativo do foco de cada publicação quanto aos conceitos estudados.

A Tabela 2.4 apresenta um resumo dos atributos de qualidade gerais identificados nesta revisão, com a quantidade de atributos por artigo e o foco de cada publicação. A lista completa com todos os atributos identificados pode ser encontrada neste link do rodapé <sup>2</sup>.

Tabela 2.4 – Tabela resumida com Atributos de Qualidade identificados

<b>ID Artigo</b>	<b>Qtd. Atributos</b>	<b>Foco</b>	<b>Atributos de Qualidade</b>
A006	7	Usabilidade	Compreensão de Contexto, Opções de Ajuda, Empatia, Linguagem Natural, Compreensão de Erros de Digitação, etc.
A009	11	Usabilidade	Feedback do Sistema, Suporte a fazer e desfazer Ações, Diálogos corretos e relevantes, Privacidade do Usuário, etc.
A013	53	UX e Usabilidade	Respostas Corretas, Manutenção de Contexto, Vocabulário apropriado, Capacidade de lidar com perguntas de Esclarecimento, etc
A014	15	Usabilidade e UX	Humanidade, Empatia, Segurança dos dados, Bom desempenho, Facilidade de Uso, Satisfação do Usuário, Respostas relevantes, etc.
A015	4	Usabilidade	Menu persistente, Botões de Respostas rápidas, Personalidade agradável e Explicação sobre o tipo de entrada esperada.

<sup>2</sup> Lista completa com Atributos de Qualidade - Revisão 1: <https://tinyurl.com/AtributosRev1>

A016	6	UX	Perguntas de acompanhamento, Conversa Fiada, Escolha de palavras sofisticadas e frases bem construídas, Expressão de emoções, etc.
A025	28	UX	Manutenção de Foco durante a conversação, Capacidade de compreender o usuário, Negociação de tópicos com o usuário, Compreensão de dicas não-verbais, etc.
A031	10	Usabilidade	Navegação intuitiva, Facilidade de Uso, Respostas linguisticamente corretas, Utilização de termos compreensíveis, etc.
A040	3	UX e Usabilidade	Diálogos inequívocos, Utilização de vocabulário simples e compreensível, Interação natural.
A041	8	UX	Estabelecimento de empatia com o usuário, Comportamento amigável, Geração de Sintonia com o Usuário, Respeito ao Usuário, etc.
A044	6	UX e Usabilidade	Fácil Compreensão ao Usuário, Transmissão de Confiança, Atendimento às expectativas do usuário, Geração de vontade de usá-lo de novo, etc.
A047	3	UX e Usabilidade	Acesso a informações avançadas, Liberdade ao usuário realizar perguntas e Conversação envolvente.
A049	24	Usabilidade	Estilo de Comunicação semelhante ao Humano, Esclarecimento de solicitações não compreendidas, Capacidade de guardar e categorizar os erros para melhoria futura, etc.
A060	39	UX e Usabilidade	Feedback imediato, Linguagem familiar ao usuário alvo, Menu permanente e Opções de Ajuda, Diálogos concisos, mas relevantes, etc.

A061	10	UX e Usabilidade	Preservação de Contexto, Respostas concisas e informativas, Estilo de conversa educado, Opções para personalizar os comandos e respostas, etc.
A067	42	Usabilidade	Facilidade de Acesso, Clareza sobre suas funcionalidades, Comunicação clara, Acompanhamento de contexto, Uso de Hiperlinks, etc.
A070	10	Usabilidade	Mensagens de erro em linguagem simples, Feedback em tempo razoável, Instruções de uso visíveis e facilmente recuperáveis, etc.
A074	34	UX e Usabilidade	Capacidade de manter uma discussão temática, Facilidade Geral de Uso, Resolução de problemas em tempo real, Atendimento a necessidades neuro diversas, etc.

Analisando a Tabela, observa-se a variedade de atributos de qualidade identificados na revisão da literatura, mesmo a tabela estando resumida. É possível identificar atributos de Acessibilidade, tal como “Facilidade de Acesso” e “Atendimento a necessidades neuro diversas”, atributos de Afeto, tal como “Estabelecimento de empatia com o usuário” e “Empatia”, dentre outros. Do mesmo modo, enquanto algumas publicações forneceram poucos atributos (A047, A040, A015), outras forneceram uma extensa lista (A067, A013, A074, A060). Nota-se também na tabela, que existem alguns atributos repetidos, iguais ou muito parecidos em publicações diferentes, por exemplo, “Facilidade Geral de Uso” da publicação A074 e “Facilidade de Uso” da publicação A014. É importante salientar que estes atributos semelhantes foram posteriormente combinados em um único atributo, garantindo que o atributo mais completo fosse selecionado e/ou complementado com detalhes dos demais atributos, minimizando as possíveis redundâncias. Quanto a possíveis diferenças nos tipos de atributos de qualidade em relação ao foco de cada publicação (Usabilidade, User Experience (UX) ou ambos, observa-se que os atributos não diferem tanto entre si, no entanto, os atributos de UX focam mais em emoções e bem-estar, além de estabelecimento de relação com o usuário, em comparação aos atributos das publicações focadas em usabilidade.

## 2.4.2 SQ1: Quais técnicas de avaliação de chatbots destacam-se durante a descoberta de atributos de qualidade?

Após a execução desta revisão da literatura, das 18 publicações selecionadas aptas a extração de dados, evidenciaram-se cinco (5) artigos no que diz respeito a apresentação de técnicas de avaliação focadas em Usabilidade e UX para agentes conversacionais. Estes artigos estão destacados na Tabela 2.5 a seguir:

Tabela 2.5 – Tabela com Técnicas Descobertas na Primeira Revisão

ID Artigo	Qtd	Artigo
A009	11	Langevin, Raina et al. <b>Heuristic evaluation of conversational agents</b> . In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021. p. 1-15.
A013	53	Sugisaki, Kyoko; Bleiker, Andreas. <b>Usability guidelines and evaluation criteria for conversational user interfaces: a heuristic and linguistic approach</b> . In: Proceedings of the Conference on Mensch und Computer. 2020. p. 309-319.
A025	28	Skjuve, M., & Brandzaeg, P. B. (2018, October). <b>Measuring user experience in chatbots: an approach to interpersonal communication competence</b> . In International Conference on Internet Science (pp. 113-120). Springer, Cham.
A060	39	Höhn, S., Bongard-Blanchy, K. (2020, November). <b>Heuristic evaluation of COVID-19 chatbots</b> . In International Workshop on Chatbot Research and Design (pp. 131-144). Springer, Cham.
A067	42	Borsci, S., Malizia, A., Schmettow, M., Van Der Velde, F., Tari-verdiyeva, G., Balaji, D., Chamberlain, A. (2022). <b>The Chatbot Usability Scale: the Design and Pilot of a Usability Scale for Interaction with AI-Based Conversational Agents</b> . Personal and Ubiquitous Computing, 26(1), 95-119.

Ao analisar a tabela, observa-se que a maioria dos artigos apresenta técnicas baseadas nas Heurísticas de Nielsen: os artigos A009, A013 e A060. O estudo A009 de [Langevin et al. \(2021\)](#) utiliza as Heurísticas de Nielsen para propor um novo conjunto com 11 novas heurísticas capazes de avaliar chatbots textuais, ativados por voz ou híbridos. O artigo A060 de [Höhn e Bongard-Blanchy \(2020\)](#) também se baseou nas Heurísticas de Nielsen para propor um conjunto de 39 itens agrupados em 12 heurísticas genéricas para avaliar chatbots. O estudo A013 de [Sugisaki e Bleiker \(2020\)](#) propõe um *checklist* para avaliar Usabilidade e User Experience em chatbots. A técnica detalhou as Heurísticas de Nielsen em 53 itens de verificação.

Em contrapartida, a publicação A067 de [Borsci et al. \(2021\)](#) propõe 42 itens de verificação focados em avaliar usabilidade em agentes conversacionais com inteligência artificial. Os itens são fundamentados em atributos de qualidade oriundos de revisão da literatura. Por sua vez, o artigo A025 de [Skjuve e Brandzaeg \(2019\)](#) apresenta uma abordagem baseada na Escala ICC de [Rubin e Martin \(1994\)](#), voltada a avaliar user experience em chatbots, com foco na competência comunicacional interpessoal dos chatbots. A escala contém 28 itens que devem ser pontuados conforme a Escala Likert de 1 a 5.

Ante o exposto, observa-se o quão relevantes as Heurísticas de [Nielsen \(2005\)](#) são quando se trata de avaliar e melhorar a usabilidade em sistemas computacionais, e neste caso, agentes conversacionais. A técnica de [Nielsen \(2005\)](#) é utilizada com frequência para analisar a usabilidade de interfaces de usuário em geral, com emprego inclusive para a análise de UX em chatbots, segundo [Höhn e Bongard-Blanchy \(2020\)](#). Nesta discussão, três artigos (A009, A013 e A060) totalizam 103 atributos de qualidade relacionados diretamente às heurísticas de usabilidade. Os outros artigos apresentam técnicas focadas em outras metodologias, sendo que o artigo A025 (28 atributos) foca em avaliar UX e a publicação A067 (42 atributos) foca em usabilidade.

Nota-se ainda, que apenas os cinco artigos da Tabela 3.5 forneceram 173 atributos de qualidade, equivalendo a mais da metade do total (313 atributos) de atributos encontrados nesta revisão da literatura. Esta descoberta sustenta a importância de investigar a literatura científica na descoberta de mais técnicas de avaliação que possam fornecer outros atributos de qualidade de Usabilidade e UX úteis para chatbots textuais, que porventura esta revisão não conseguiu identificar.

### 2.4.3 SQ2: O conjunto de atributos foi proposto com qual objetivo?

Cada publicação selecionada apresentou um estudo diferente, de forma que seu conjunto de atributos de qualidade também tem seus objetivos particulares. No entanto, algumas publicações são baseadas na mesma fundamentação teórica. A Tabela 2.6 a seguir, mostra o objetivo de cada conjunto de atributos proposto.

Tabela 2.6 – Tabela com o objetivo dos atributos de qualidade identificados

ID	Ref. Artigo	Objetivo
A006	<a href="#">Hefny et al. (2021)</a>	O chatbot baseado em personagens Chasey foi desenvolvido para combater desinformação sobre a COVID-19, rastreando o número de casos por países, efetuando a verificação de sintomas e respondendo FAQ, contendo atributos genéricos para essas ferramentas.

A009	<a href="#">Langevin et al. (2021)</a>	O trabalho não propõe um chatbot, mas um conjunto genérico de heurísticas para agentes conversacionais adaptadas das heurísticas da Nielsen e baseadas no feedback de especialistas, que pode ser aplicado na avaliação de chatbots textuais, ativados por voz ou multimodais.
A013	<a href="#">Sugisaki e Bleiker (2020)</a>	A publicação não apresenta chatbot, mas uma técnica de avaliação baseada em checklist para avaliar chatbots: 53 checkpoints baseados nas Heurísticas de Nielsen. Os checkpoints fornecem características importantes para chatbots.
A014	<a href="#">Denecke e Warren (2020)</a>	O artigo apresenta um conjunto de aspectos para avaliar chatbots relacionados a sistemas de saúde. São apresentados atributos úteis para melhorar a qualidade de interfaces conversacionais da área de saúde, mas que também são genéricos e aplicáveis a vários tipos de chatbots.
A015	<a href="#">McAllister et al. (2020)</a>	A publicação apresenta o chatbot Bibliobot, um chatbot baseado em texto voltado a apoiar facilitadores de biblioterapia, um tipo de terapia que utiliza contos para apoiar pacientes com problemas de saúde mental. Os atributos de qualidade descritos são genéricos.
A016	<a href="#">Svenningsson e Faraon (2019)</a>	O artigo não propõe um chatbot, mas visa analisar fatores relacionados à humanidade percebida em chatbots, fornecendo diretrizes importantes para o design destas ferramentas. Dentre estas diretrizes há atributos de qualidade genéricos.
A025	<a href="#">Skjuve e Brandzaeg (2019)</a>	O artigo propõe uma escala para avaliar user experience em chatbots, com foco na competência comunicacional interpessoal. São discutidos os fatores que contribuem para uma comunicação bem-sucedida com chatbots e como avaliar a experiência do usuário em conversas com chatbots.
A031	<a href="#">Denecke et al. (2018)</a>	A publicação não propõe chatbot, mas realiza um experimento visando testar o desempenho, a humanidade, afeto e acessibilidade do chatbot Ana, voltado a realização de autoanamnese (entrevista médica no qual o próprio paciente registra os seus sintomas). As tarefas e questões para o experimento são oriundos de atributos de qualidade da literatura.

A040	<a href="#">Kadariya et al. (2019)</a>	O artigo propõe o kBot, um chatbot baseado em conhecimento, projetado para aplicações de saúde e adaptado para ajudar pacientes pediátricos asmáticos a controlar melhor sua asma. A ferramenta foi avaliada utilizando métricas de qualidade úteis para muitos tipos de chatbots.
A041	<a href="#">Pecune et al. (2019)</a>	O artigo apresenta o chatbot Sara, desenvolvido para atuar como um sistema de recomendação de filmes, sendo submetido a uma avaliação para medir a qualidade percebida do agente conversacional e a qualidade geral percebida da interação.
A044	<a href="#">Jin et al. (2019)</a>	O chatbot proposto neste artigo é o MusicBot, uma ferramenta para a recomendação de músicas, avaliada sob os aspectos sofisticação musical (MS), desejo de controle (DFC), experiência do chatbot (CE) e conhecimento tecnológico (TS), na percepção e interação do usuário, com aspectos importantes para chatbots em geral.
A047	<a href="#">Spiliotopoulos et al. (2020)</a>	O artigo apresenta uma proposta experimental desenvolvida por 3 designers de chatbots para interfaces de conversação orientadas por uma semântica voltada a chatbots em ambientes de museus, ajudando visitantes com informações sobre artes, exposições, artistas e outros aspectos relacionados. No entanto, há atributos gerais para chatbots.
A049	<a href="#">Feine et al. (2020)</a>	O artigo foca nos chatbots empresariais (ECB) e estuda princípios de design que devem orientar o desenvolvimento desses sistemas. Após serem avaliados, identificou-se que os princípios que promovem a usabilidade e a flexibilidade são classificados como mais importantes do que os que promovem a sociabilidade e a semelhança humana.
A060	<a href="#">Höhn e Bongard-Blanchy (2020)</a>	A publicação propõe um framework genérico com 39 aspectos de avaliação agrupados em 12 heurísticas baseadas nas dez heurísticas de Nielsen. Os itens são aplicáveis para avaliar Usabilidade e UX em chatbots de diferentes estilos de conversação.
A061	<a href="#">Yang e Aurisicchio (2021)</a>	A publicação realiza um estudo que visa explorar como os fatores competência, autonomia e relacionamento dos usuários influenciam na experiência com chatbots e então são derivadas 10 diretrizes de design para agentes conversacionais.



A067	<a href="#">Borsci et al. (2021)</a>	O artigo apresenta o processo de criação do BOT-Check, ferramenta em forma de <i>checklist</i> com 42 itens de verificação genéricos para avaliar agentes conversacionais baseados em IA.
A070	<a href="#">Niculescu et al. (2014)</a>	A publicação concentra-se na descrição do processo de construção da Sara, um agente conversacional desenvolvido para o domínio turístico e avaliado conforme as heurísticas de Nielsen.
A074	<a href="#">Anshu, Gaur e Solanki (2021)</a>	O artigo propõe desenvolver e testar uma estrutura concebida com a identificação de atributos de qualidade para chatbots orientados ao consumidor e o impacto que esses atributos criam nos clientes. Para isso foram identificados os atributos de qualidade capazes de afetar o consumidor.

Nota-se, por meio da tabela, que os objetivos são diversificados. Há publicações que o objetivo é propor um chatbot e junto a isso, apresentar o processo de design/construção das ferramentas, tornando-se possível identificar atributos de qualidade utilizados para o desenvolvimento destes agentes conversacionais, tais como os artigos A006 (Chasey), A015 (Bibliobot), A040 (kBot), A041 (Sara - Filmes), A042 (MusicBot), A070 (Sara - Turismo). Entretanto, a maioria das publicações visa mostrar o processo de desenvolvimento de uma técnica, apresentar a avaliação de uma técnica e discutir os resultados ou avaliar um conjunto de aspectos no qual também é possível identificar atributos de qualidade úteis para o desenvolvimento de chatbots. As publicações voltadas a um domínio específico, como recomendação de filmes, recomendação de músicas, informações sobre saúde e acompanhamento de doenças apresentam muitos atributos de qualidade específicos para esses domínios, mas é possível identificar também atributos de qualidades gerais, que podem ser úteis a vários tipos de chatbots, sendo estes atributos o foco desta revisão da literatura.

## 2.5 Conclusão

Neste capítulo realizou-se uma revisão da literatura com o objetivo principal de identificar atributos de qualidade gerais focados em Usabilidade e User Experience para chatbots textuais. Após a condução da revisão da literatura, foram selecionados 18 artigos relevantes para os objetivos deste estudo, responsáveis por contribuir com um total de 313 atributos de qualidade relacionados a Usabilidade e UX para chatbots textuais. Das publicações selecionadas, cinco artigos se destacaram por apresentarem técnicas de avaliação para chatbots, sendo a maioria destas baseadas no modelo de [Nielsen \(2005\)](#).

Observou-se ainda, que as cinco publicações que apresentaram técnicas de avaliação para chatbots contribuíram com mais atributos de qualidade do que as demais publicações, levando a crer que técnicas de avaliação são uma ótima fonte de atributos. Considerando isto, tornou-se oportuna a realização de uma segunda revisão da literatura, visando identificar outras técnicas de avaliação de chatbots que possam contribuir com esta pesquisa, complementando este estudo com mais atributos de qualidade focados em Usabilidade e User Experience para chatbots textuais. No capítulo a seguir, são apresentados os resultados dessa segunda revisão complementar.

## 3 Segunda Revisão da Literatura

*Este capítulo apresenta os passos executados e os resultados do segundo estudo secundário da literatura, realizado de forma complementar para a identificação técnicas de avaliação para chatbots e os atributos de qualidade de usabilidade e user experience que estas técnicas podem oferecer.*

### 3.1 Introdução do Capítulo

Para garantir que chatbots atendam a qualquer padrão de qualidade definido, é fundamental que estas ferramentas sejam submetidas aos processos de avaliação pertinentes, como afirma [Chow, Sanders e Li \(2023\)](#). Estes processos verificam se estes sistemas são de fato eficazes na interação com os usuários e cumprem os objetivos estabelecidos. Os resultados da avaliação de um chatbot possibilitam que os desenvolvedores ajustem e aprimorem o sistema para oferecer resultados cada vez melhores, além de proporcionar aos usuários uma experiência cada vez mais satisfatória.

Na revisão da literatura do Capítulo 2, descobriu-se um conjunto de atributos de qualidade para chatbots textuais focados nos conceitos de Usabilidade e UX. Esses atributos podem ajudar a conceber tecnologias úteis para melhorar a qualidade de agentes conversacionais. Nos resultados da primeira revisão, identificou-se que apenas cinco (5) artigos forneceram a maioria dos atributos de qualidade e estes artigos se distinguiam dos demais por apresentarem técnicas de avaliação para chatbots. Assim, tornou-se importante realizar uma segunda revisão da literatura, desta vez, com o objetivo de identificar técnicas para avaliar chatbots com o foco nos conceitos já apresentados, para então investigar outros possíveis atributos de qualidade significativos que estas publicações possam fornecer.

É oportuno ressaltar que se optou por realizar revisões sistemáticas isoladas para que os resultados de ambas pudessem beneficiar diferentes públicos. Pesquisadores que buscam encontrar informações sobre atributos de qualidade (foco em Usabilidade e UX) para chatbots, podem se favorecer com os resultados do estudo secundário apresentado no Capítulo 2, do mesmo modo que aqueles que procuram informações sobre técnicas de avaliação para chatbots (foco em Usabilidade e UX) podem encontrá-las no presente Capítulo.

Ante o exposto, este capítulo apresenta um estudo secundário na literatura científica visando identificar técnicas de avaliação da usabilidade e user experience em chatbots e quais atributos de qualidade essas publicações podem fornecer. O restante deste capítulo

está organizado da seguinte forma: a Seção 3.2 apresenta o planejamento do protocolo da revisão da literatura; a Seção 3.3 apresenta a execução da revisão da literatura; a Seção 3.4 apresenta os resultados e uma discussão; e finalmente a Seção 3.5 encerra este capítulo.

## 3.2 Planejamento do Protocolo da Revisão

Nesta seção é possível encontrar os detalhes do protocolo de estudo secundário sobre as técnicas de avaliação focadas em Usabilidade e User Experience para chatbots. As questões de pesquisa, os critérios de seleção de artigos primários, os termos de busca e as bases de pesquisa são descritos. Esse protocolo é detalhado nas subseções a seguir.

### 3.2.1 Questões de Pesquisa

Esta revisão tem como propósito identificar na literatura científica as técnicas utilizadas para avaliar aspectos de Usabilidade e UX em chatbots. A justificativa se dá pela necessidade de entender como os chatbots têm sido avaliados de acordo com esses aspectos, como também identificar os atributos de qualidade relevantes que as técnicas de avaliação podem contribuir.

A principal questão de pesquisa a ser respondida nesta revisão é:

- **QP** - “*Quais técnicas de avaliação estão disponíveis para avaliar a Usabilidade e UX de chatbots?*”

Além dela, a revisão teve em vista responder à seguinte subquestão (SQ):

- **SQ1** - *Quais atributos de qualidade podem ser identificados nas técnicas de avaliação da Usabilidade e User Experience em chatbots?*

### 3.2.2 Termos de Busca

Assim como ocorreu na revisão sistemática do Capítulo 2, nesta subseção apresentam-se os termos utilizados para a montagem da estratégia de busca. Uma *String* de busca é composta por várias palavras-chave relacionadas à questão principal de pesquisa, que quando inseridas nos mecanismos de busca é capaz de retornar publicações relevantes para responder os questionamentos acerca do assunto. Os sinônimos e palavras-chave da *String* de busca são derivados dos trabalhos de Cabrejos, Viana e Santos (2018), Guerino e Valentim (2020), Denecke e Warren (2020), Radziwill e Benton (2017), Coppola e Ardito (2021), Suhaili, Salim e Jambli (2021). Na Tabela 3.1 são apresentadas essas palavras-chave e seus respectivos sinônimos.

Tabela 3.1 – Termos utilizados para a montagem da String de busca

Palavras-chave	Sinônimos
technique	instrument; tool; checklist; questionnaire; approach; method; system; scale; scheme; framework; model
evaluation	assessment; measurement; testing; recognition; measure; evaluating; tracking; assess
usability e user experience	usable; ux
chatbot	conversational user interface; conversational agent

Com base nos termos de busca, a *String* de busca base resultante foi: ((“*technique*” OR “*instrument*” OR “*tool*” OR “*checklist*” OR “*questionnaire*” OR “*approach*” OR “*method*” OR “*system*” OR “*scale*” OR “*scheme*” OR “*framework*” OR “*model*”) AND (“*evaluation*” OR “*assessment*” OR “*measurement*” OR “*testing*” OR “*recognition*” OR “*measure*” OR “*evaluating*” OR “*tracking*” OR “*assess*”) AND (“*usability*” OR “*usable*” OR “*user experience*” OR “*UX*”) AND (“*chatbot*” OR “*conversational user interface*” OR “*conversational agent*”). A biblioteca digital selecionada para executar esta *String* é detalhada a seguir.

### 3.2.3 Fontes de Busca

A *Scopus Digital Library* também foi a base de dados escolhida para selecionar as publicações analisadas nesta revisão. Conforme já exposto na Revisão do Capítulo 2, a *Scopus* indexa trabalhos de outras bibliotecas como a *ACM Digital Library*, *Spring Link Digital Library*, *Science Direct Digital Library* e *IEEE Xplore Digital Library*, além de ser reconhecida pela qualidade e confiabilidade no universo científico e acadêmico.

### 3.2.4 Critérios de Seleção

Além da *String* de busca, é importante definir os critérios de seleção das publicações a serem aceitas ou descartadas na revisão. Ao executar a *String* de busca, a base de dados pode retornar artigos que não estão relacionados à pesquisa, segundo Cabrejos, Viana e Santos (2018), e os critérios são utilizados para decidir se uma publicação será aceita ou não. Desta forma, estabeleceram-se os critérios de inclusão (CI) e exclusão (CE) a seguir.

O critério de inclusão (CI) definido para esta revisão foi:

- **CI1:** A publicação descreve uma técnica/abordagem/ferramenta para avaliar usabilidade e UX em chatbots.

Os critérios de exclusão (CE) definidos foram:

- **CE1:** A publicação não descreve uma técnica/abordagem/ferramenta para avaliar usabilidade e UX em chatbots
- **CE2:** A publicação não é um artigo científico, e sim capítulo de livro, workshop, introdução a conferência, etc.
- **CE3:** A publicação não está escrita em Inglês ou Português.
- **CE4:** A publicação não está disponível para leitura e coleta de dados (artigos pagos ou não disponibilizados pela máquina de busca).
- **CE5:** A publicação descreve uma técnica/abordagem/ferramenta para avaliar usabilidade e UX em chatbots, mas não para chatbots textuais.
- **CE6:** A publicação está duplicada.

Durante a fase de execução da revisão, definiu-se o formulário de extração de dados, apresentado na Tabela 3.2, para auxiliar na condução da seleção dos artigos para a pesquisa e posteriormente, para a extração dos dados. O formulário contém lacunas que devem ser preenchidas, conforme sugerido por Cabrejos, Viana e Santos (2018).

Tabela 3.2 – Formulário de Extração de Técnicas

Campos	Descrições
Nome da Técnica	Escrever, se existente, o nome empregado à técnica de avaliação.
Descrição da Técnica	Escrever a descrição da técnica, a metodologia de avaliação, por exemplo: se é uma escala, heurística, <i>checklist</i> , etc.
Atributos de Qualidade	Descrever se existentes, quais atributos de qualidade de Usabilidade e User Experience são considerados pela metodologia ao realizar as avaliações

### 3.3 Execução da Revisão

Através da execução da *string* de busca nos motores de busca da biblioteca digital da *Scopus* no período de março a abril de 2022, obteve-se um total de 272 publicações, que foram submetidas a filtros para descartar aquelas que não correspondiam ao que a pesquisa busca. A Tabela 3.3 contém a lista com a quantidade geral de artigos retornados e quanto restou após a aplicação de cada filtro.

Tabela 3.3 – Relação da Quantidade de Artigos - Segunda Revisão

<b>Biblioteca</b>	<b>Qtd. Geral</b>	<b>Filtro 1</b>	<b>Filtro 2</b>
Scopus	272	84	14

Todas as 272 publicações foram submetidas ao primeiro filtro, que normalmente se dá pela leitura do título e do resumo de cada uma. Após o primeiro filtro, 84 artigos estavam aptos a passarem para o próximo filtro. O segundo filtro é mais complexo que o primeiro e ocorre com a aplicação de todos os critérios de inclusão e exclusão definidos no planejamento da pesquisa. Após o segundo filtro, restaram 14 publicações, que após lidas completamente, estavam aptas à extração das informações que a revisão busca.

### 3.4 Resultados e Discussão

Esta revisão da literatura possibilitou encontrar 14 técnicas de avaliação relacionadas a Usabilidade e User Experience para chatbots. A Tabela 3.4<sup>1</sup> a seguir, mostra o resumo obtido após a extração dos dados, sendo (1) a quantidade de artigos selecionados para a revisão da literatura; (2) a quantidade de técnicas identificadas nos artigos e (3) a quantidade total de atributos de qualidade identificados nas técnicas de avaliação. Na Subseção 3.4.1 serão discutidas as questões de pesquisa da revisão da literatura, conforme os resultados obtidos.

Tabela 3.4 – Tabela com Resumo de Extração

<b>Qtd. Artigos</b>	<b>Qtd. Técnicas</b>	<b>Qtd. Atributos</b>
14	14	273

#### 3.4.1 QP: Quais técnicas de avaliação estão disponíveis para avaliar a Usabilidade e UX de chatbots?

As 14 técnicas de avaliação da Usabilidade e User Experience em chatbots identificadas após a condução da segunda revisão sistemática da literatura estão descritas na Tabela 3.5, de forma que, a coluna (1) apresenta o identificador (ID) dado ao artigo; (2) a descrição da técnica com a metodologia e a base teórica utilizada e (3) a quantidade de atributos que cada técnica contribuiu:

<sup>1</sup> Formulários de Extração completos em: <https://tinyurl.com/FormularioExtracaoRev2>

Tabela 3.5 – Tabela de Técnicas de Avaliação para Chatbots

ID	Descrição da Técnica	Qtd. Atributos
B002	Técnica de Checklist de Inspeção baseada nas Heurísticas de Usabilidade de Nielsen. As 10 tradicionais heurísticas foram detalhadas em diversos checkpoints úteis para avaliar a qualidade de chatbots.	53
B004	Técnica baseada em Elicitação de Incidentes que consiste em solicitar que o usuário/entrevistado identifique pontos críticos durante a experiência com o chatbot, com apoio de um questionário. Posteriormente estes pontos são analisados por avaliadores experientes.	11
B011	Técnica de Questionário baseada em Questionário de Experiência do Usuário (UEQ), Escala SUS (System Usability Scale) e Questionário de Usabilidade de Chatbot (CUQ), que consiste em pontuar itens da ferramenta conforme as perguntas do questionário.	16
B016	Técnica de Questionário baseado no modelo TAM - Modelo de Aceitação de Tecnologia aplicado em grupo focal (os grupos de participantes são expostos ao chatbot e debatem, levantando pontos importantes).	16
B018	Técnica de Questionário com 10 questões baseado no SUS (System Usability Scale) aplicado ao chatbot CBET. O SUS é um dos instrumentos mais utilizados para avaliar a usabilidade de um sistema devido as suas psicométricas desejáveis como alta confiabilidade e validade.	10
B019	Técnica de Questionário baseada na ISO/IEC 25.023 (SQuaRE) que contém requisitos e métricas de avaliação de qualidade de sistemas e software juntamente com a Escala de Usabilidade do Sistema (SUS).	18
B035	Técnica de Questionário com 10 questões baseado no método SUS (System Usability Scale) aplicado ao chatbot Anne.	10
B040	Técnica de Checklist de Inspeção, BOT-Check baseado em atributos de qualidade e nas Heurísticas de Nielsen contendo itens genéricos para avaliar chatbots.	42



B057	Técnica de Questionário baseado no CUQ (Chatbot Usability Questionnaire), um questionário baseado nos princípios de UX para chatbots, fornecidos pela ferramenta ALMA Chatbot Test, que avalia personalidade, integração, navegação, compreensão, respostas, tratamento de erros e inteligência de um chatbot.	16
B061	Técnica de Questionário com 10 questões baseado na metodologia SUS (System Usability Scale) aplicado ao chatbot ADMINS.	10
B064	Técnica de Questionário baseada em UMUX (Usability Metric for User Experience), uma avaliação qualitativa curta projetada para medir UX e a usabilidade geral de um sistema.	12
B065	Técnica de Questionário com 10 questões baseado no método SUS (System Usability Scale) aplicado a um chatbot universitário.	10
B068	Técnica de Checklist de Inspeção baseada nas Heurísticas de Usabilidade de Nielsen. Os itens do checklist são especificados a partir das dez heurísticas de Nielsen.	39
B076	Técnica de Questionário com 10 questões baseado na escala SUS (System Usability Scale) aplicado ao chatbot iHelpr.	10

Analisando a Tabela 3.5, observa-se que predominam as técnicas baseadas em itens de verificação, Questionários e *Checklists*, sendo 13 das 14 técnicas. Os 03 *checklists* presentes nesta revisão, B002 de Sugisaki e Bleiker (2020), B040 de Borsci et al. (2021) e B068 de Höhn e Bongard-Blanchy (2020) são os mesmos já retornados na revisão do Capítulo 2 e, assim como observado anteriormente, contribuem com a maioria dos atributos nesta revisão.

É possível notar que cinco técnicas identificadas (artigos B018, B035, B061, B065 e B076) são baseadas exclusivamente no SUS (System Usability Scale), o questionário padronizado mais utilizado para a avaliação da usabilidade percebida, segundo Lewis (2018). Estas publicações contêm as mesmas afirmações avaliativas que a escala de usabilidade propõe e, portanto, os atributos oriundos destes trabalhos foram posteriormente agrupados para evitar duplicidades na tecnologia proposta por esta dissertação.

Ainda observando a Tabela 3.5, evidencia-se uma técnica baseada em Elicitação de Incidentes (B004), em que o usuário participa ativamente ajudando a identificar pontos críticos na aplicação, no entanto, ainda houve a aplicação de um questionário no processo.

A técnica B011, por sua vez, combina 03 abordagens de avaliação: o Questionário de Experiência do Usuário (UEQ) válido para avaliar User Experience, a Escala de Usabilidade de Sistema (SUS) voltado a avaliar a Usabilidade e o Questionário de Usabilidade de Chatbot (CUQ) sendo focado tanto em Usabilidade, quanto UX em chatbots. Outra técnica descoberta para medir UX e usabilidade geral de um sistema é a baseada em UMUX (Métrica de Usabilidade para Experiência do Usuário (B064).

### 3.4.2 SQ1: Quais atributos de qualidade podem ser identificados nas técnicas de avaliação da Usabilidade e User Experience em chatbots?

Conforme os resultados da Tabela 3.4, um total de 273 atributos de qualidade foram identificados nas 14 técnicas de avaliação descobertas. Foram extraídos os atributos gerais de qualidade, que podem ser aplicados à maioria dos tipos de chatbot. A seguir é apresentada a Tabela 3.6 com uma parte dos atributos identificados na revisão e os respectivos requisitos. Devido ao seu tamanho, torna-se inviável disponibilizar a lista completa de atributos neste documento, contudo é possível acessá-la através do link no rodapé <sup>2</sup>.

Tabela 3.6 – Tabela resumida com Atributos de Qualidade identificados

ID Artigo	Qtd. Atributos	Atributos de Qualidade
B002	53	Respostas Corretas, Armazenamento de informações para Manter o Contexto, Vocabulário apropriado ao usuário alvo, Personalização da conversa conforme as preferências do usuário, etc.
B004	11	Manutenção de assunto, Registro linguístico apropriado, Dicas de Conversação, Capacidade de lidar com enunciados inadequados e controle de danos, etc.
B011	16	Personalidade realista e envolvente, Evitar parecer muito robótico, Facilidade de Navegação, Capacidade de dar respostas úteis, apropriadas e informativas, etc.
B016	17	Responsividade, Linguagem simples, Precisão Linguística, Uso de cores brilhantes para as fontes, Fornecimento de Informações confiáveis, etc.

<sup>2</sup> Lista completa com Atributos de Qualidade - Revisão 2: <https://tinyurl.com/AtributosRev2>

B018	9	Facilidade de Uso, Funções bem integradas, Fácil de aprender a utilizar, Evitar ser desnecessariamente complexo, Evitar inconsistências, etc.
B019	18	Resposta em tempo razoável, Funcionamento em paralelo com outros softwares sem perder desempenho, Utilização de recursos eficientemente, Interface gráfica satisfatória, etc.
B035	10	Transmissão de confiança ao usuário, Coerência, Facilidade de aprender a utilizá-lo, Facilidade de Uso, Funções bem integradas, etc.
B040	42	Facilidade de Acesso, Facilidade de Utilização, Comunicação clara, Acompanhamento de contexto, Capacidade de manter a conversa relevante, Uso de hiperlinks para guiar o usuário até seu objetivo, etc.
B057	16	Respostas rápidas, Facilidade de Uso, Capacidade de lidar com erros ou enganos, Capaz de compreender as solicitações do usuário, Evitar parecer robótico, Respostas relevantes, Evitar ser hostil com o usuário, etc.
B061	10	Evitar que o usuário necessite de suporte técnico, Facilidade de Uso, Consistência, Conquista da confiança do usuário, Funções bem integradas, etc.
B064	12	Capacidade de atender aos requisitos do usuário, Facilidade de Uso, Utilidade, Execução rápida das tarefas, Entendimento dos questionamentos do usuário, etc.
B065	10	Facilidade de Uso, Evitar ser complicado ao usuário, Funções bem integradas, Geração de vontade de utilização com mais frequência, etc.
B068	39	Menu Permanente, Opções de Navegação, Confirmação antes de ações com implicações significativas, Compreensão de instruções especiais e sinônimos, Diálogos concisos e relevantes, etc.
B076	10	Evitar complexidade desnecessária, Facilidade de Uso, Facilidade de aprendizado, Funções bem integradas, etc.

Analisando a Tabela 3.6, observa-se a diversidade de atributos de qualidade identificados nas técnicas de avaliação descobertas nesta revisão. As publicações que mais forneceram atributos foram B002 de Sugisaki e Bleiker (2020), B040 de Borsci et al. (2021) e B068 de Höhn e Bongard-Blanchy (2020) já descobertas na primeira revisão. Além destas, é possível destacar os artigos B004, B011, B016, B019 e B057, que fornecem variados tipos de atributos, que incluem Manutenção de Temas, Responsividade, Precisão Linguística, Respostas em tempo razoável, Facilidade de Uso, dentre outros.

As publicações B018, B035, B061, B065 e B076 são todas baseadas na Escala SUS, por isso essas publicações compartilham muitos atributos semelhantes, como: Funções bem integradas, Facilidade de Uso, Evitar complexidade desnecessária, Coerência, e outros. A Escala SUS possui 10 afirmações com aspectos positivos e negativos, tais como, a afirmação 1 - “Acho que gostaria de usar este sistema com frequência” e a afirmação 2 - “Achei o sistema desnecessariamente complexo”. Nota-se que o item 2 está descrito de forma que dificulta a extração de atributo de qualidade, visto que um sistema “desnecessariamente complexo” é o oposto do que um usuário espera de um software de qualidade. Assim, para extrair os atributos de qualidade dos itens negativos da Escala SUS, utilizou-se o estudo de Sauro e Lewis (2011), que transformou todos os itens da escala SUS em itens positivos.

Comparando este conjunto de atributos com os descobertos na primeira revisão, observou-se uma quantidade bem menor de atributos de qualidade referentes a fatores como Humanidade e Afeto voltados a chatbots. Apesar disso, esta revisão complementar foi importante pois permitiu descobrir atributos novos: Uso de hiperlinks para guiar o usuário até seu objetivo, Confirmação antes de ações com implicações significativas, Utilização eficiente de recursos, Funcionamento em paralelo com outros softwares sem perder desempenho, Uso de cores brilhantes para as fontes, dentre outros.

## 3.5 Conclusão

Neste capítulo realizou-se uma revisão da literatura complementar com o intuito de descobrir técnicas utilizadas para avaliar chatbots com foco em Usabilidade e User Experience. Na ocasião, foram encontrados 14 artigos relevantes para os objetivos desta pesquisa, somando um total de 273 atributos de qualidade. Dentre as técnicas descobertas, observou-se o domínio das técnicas de verificação, *Checklists*/Questionários, provavelmente pelas vantagens que este tipo de técnica oferece. A revisão retorna ainda os trabalhos baseados nas heurísticas Nielsen (2005) já conhecidos na Revisão 2.

Assim como na revisão do Capítulo 2, as publicações que mais forneceram atributos de qualidade foram os trabalhos de Sugisaki e Bleiker (2020), Borsci et al. (2021) e Höhn e Bongard-Blanchy (2020), entretanto, esta revisão complementar foi de suma importância para localizar outros atributos de qualidade que a primeira revisão não identificou. Com

as duas listas de atributos de Usabilidade e User Experience tornou-se possível propor tecnologias mais assertivas para apoiar o design e a avaliação de chatbots. No capítulo a seguir, são apresentados os detalhes da criação do *checklist* de inspeção para avaliar Usabilidade e User Experience em chatbots.

## 4 Criação do Checklist U2Chatbot

*Este capítulo apresenta a criação do checklist de inspeção U2CHATBOT, desenvolvido para avaliar os aspectos de Usabilidade e User Experience em chatbots e auxiliar na identificação de defeitos nesses sistemas. São apresentados os passos adotados para a proposição da tecnologia, bem como também uma prova de conceito para uma avaliação preliminar do checklist.*

### 4.1 Introdução do Capítulo

A garantia da qualidade de um software está relacionada ao estabelecimento de processos para a inibição e o impedimento de falhas, de forma que, avaliar um produto de software possibilita a identificação prematura de defeitos nesses artefatos, de acordo com [Bartié \(2002\)](#). Assim, para melhorar a qualidade de chatbots, pesquisadores têm apresentado técnicas de avaliação que permitem descobrir defeitos neste tipo de ferramenta ([KULIGOWSKA, 2015](#)), ([SEDOC et al., 2019](#)), com destaque às técnicas baseadas em inspeção ([SUGISAKI; BLEIKER, 2020](#)), ([HÖHN; BONGARD-BLANCHY, 2020](#)), ([BORSCI et al., 2021](#)).

Checklists são técnicas de inspeção no qual os avaliadores internos ou externos à equipe de desenvolvimento podem verificar a conformidade do chatbot com atributos de qualidade ([FRAZÃO et al., 2021](#)), através de listas de verificação, buscando validar se o software está consoante às necessidades do usuário. Por não exigir que o sistema já esteja em sua versão completa, esta técnica pode ser utilizada antes do lançamento do software no mercado; contribuindo inclusive para diminuir os custos de correção dos defeitos identificados ([ALSAYED; BILGRAMI; FOSTER, 2017](#)).

Este capítulo descreve a criação do *checklist* de inspeção U2CHATBOT, desenvolvido para identificar defeitos em chatbots textuais. O *checklist* foi criado extraíndo atributos de qualidade focados em Usabilidade e User experience identificados em revisões na literatura. Ainda neste capítulo, realizou-se uma prova de conceito aplicando o checklist U2CHATBOT em um chatbot real. Os resultados apontaram que os itens de verificação da tecnologia são viáveis para a descobertas de problemas em chatbots.

O restante desse capítulo está organizado como se segue: A Seção [4.2](#) apresenta os procedimentos adotados para o desenvolvimento do *checklist* U2CHATBOT; a Seção [4.3](#) apresenta a realização de uma prova de conceito para atestar a viabilidade do U2CHATBOT em um chatbot e a Seção [4.4](#) contém as considerações finais deste capítulo.

## 4.2 O Processo de Desenvolvimento do Checklist

Com o objetivo de desenvolver o *checklist* de inspeção para avaliar e identificar defeitos em chatbots, as seguintes etapas foram consideradas: (1) identificação dos atributos de qualidade para chatbots relacionados à Usabilidade e User Experience na literatura científica; (2) identificação dos atributos de qualidade em técnicas de avaliação da Usabilidade e User Experience para chatbots na literatura científica; (3) realização de análise dos atributos de qualidade, agrupando os semelhantes, transformando-os em requisitos de sistema e integrando todos os itens considerados e identificados, organizando-os em um *checklist* de inspeção. As etapas definidas para o desenvolvimento do *checklist* foram baseadas na metodologia utilizada no trabalho de [Frazão et al. \(2021\)](#), que desenvolveu o IC-MEG por meio de atributos de qualidade identificados em heurísticas para jogos educacionais móveis. Nas subseções a seguir, são apresentadas com mais detalhes cada uma dessas etapas.

### 4.2.1 Atributos de Qualidade focados em Usabilidade e User Experience Identificados em Revisões da Literatura

O processo de identificação dos atributos de qualidade focados em Usabilidade e User Experience está descrito nos Capítulos 2 e 3. Após a realização das revisões da literatura obteve-se duas listas extensas com atributos de qualidade para chatbots: A 1<sup>a</sup> lista com 313 atributos de qualidade oriundos da primeira revisão e a 2<sup>a</sup> lista com 273 atributos de qualidade oriundos da segunda revisão.

Ao analisar cada lista de atributos, identificou-se a presença de itens iguais/muito parecidos ou que estabeleciam uma relação próxima, gerando a necessidade de agrupar esses atributos semelhantes e/ou relacionados para diminuir as redundâncias. Como exemplo de itens semelhantes aptos a serem agrupados, estão os atributos: (A006) *Menu permanente e opções de ajuda*; (A015) *Uso de menu persistente (Ajuda, Menu, estou perdido)*; (A060) *Chatbot oferece menu permanente e opção de ajuda*; (A031) *A página de ajuda é útil*, em que todos versam sobre disponibilização de Menu Permanente e Opções de Ajuda na utilização de chatbots.

Em relação à junção de itens relacionados, um exemplo de atributos aptos ao agrupamento: (A006) *Capacidade de realizar conversa fiada para remediar o problema das conversas artificiais*; (A016) *O chatbot sabe lidar quando conversa fiada não faz o estilo do usuário*; (A013) *O chatbot lida com solicitações genéricas e fora do tópico (por exemplo, conversa fiada) adequadamente?*; (A025) *O chatbot consegue manter o foco durante a conversa?*. Apesar desses atributos não serem exatamente iguais, foram agrupados por estabelecerem uma relação ligada ao aspecto Foco, visto que uns itens se referem à utilização do recurso conversa fiada (que se trata de um desvio de foco) e outros itens



referem-se ao estabelecimento de foco. Após este procedimento de agrupamento, a 1ª lista foi reduzida a 162 atributos e a 2ª lista foi reduzida a 109 atributos.

Para que o *checklist* de inspeção pudesse ser criado, ainda houve a necessidade de agrupar as duas listas de atributos de qualidades descobertas (a primeira com 162 atributos e a segunda com 109 atributos). Desta forma, fez-se necessário realizar novamente o processo de junção dos atributos semelhantes para evitar redundâncias nos itens do *checklist* de inspeção. Nesta etapa, a pesquisadora contou com a colaboração de outro profissional para validar a lista de atributos. Na Figura 4.1 é apresentado um exemplo deste processo. Na ilustração observa-se que o **item 1: Facilidade de Uso** da lista de atributos é semelhante ao **item 5: Fácil de Usar** da segunda lista. Desta forma, para evitar itens repetitivos ao unir as listas de atributos, os itens 1 e 5 foram considerados um só, o **item 1: Facilidade de Usar** na lista resultante apta a ser transformada em *checklist* de inspeção.

Figura 4.1 – Processo de Agrupamento de Atributos Semelhantes



Fonte: Elaboração Própria

Após este tratamento realizado nos 162 atributos de qualidade identificados no Capítulo 2 e nos 109 atributos oriundos do Capítulo 3, resultou-se em uma lista de 107 itens utilizados na criação do *checklist* de inspeção U2CHATBOT, apresentado na subseção a seguir.

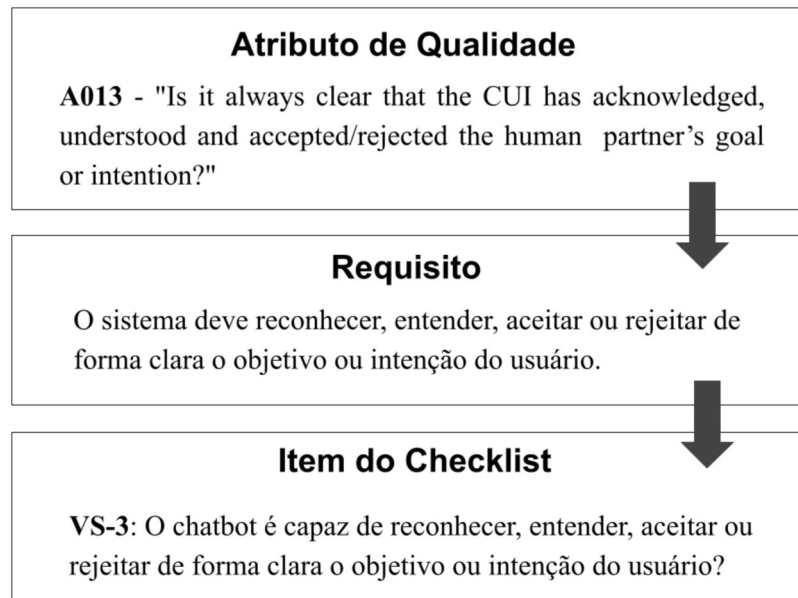
#### 4.2.2 Checklist de Inspeção U2Chatbot

No decorrer do processo de desenvolvimento do *checklist*, foram solicitadas algumas orientações de engenheiros de software sobre como transformar os atributos de qualidade em itens de verificação. A orientação sugerida foi a transformação dos atributos de qualidade



focados em Usabilidade e User Experience em requisitos de sistema, seguindo a metodologia de Leite (2002), visto que descrever como requisito facilitaria posteriormente a descrição dos itens do *checklist* U2CHATBOT. A Figura 4.2 ilustra esse procedimento.

Figura 4.2 – Processo de Transformação de Atributo em Requisito e Item do Checklist



Observa-se na Figura 4.2, um exemplo da transformação: O atributo de qualidade identificado no artigo A013 - “**Is it always clear that the CUI has acknowledged, understood and accepted/rejected the human partner’s goal or intention?**” foi transformado no seguinte requisito, conforme as diretrizes de Leite (2002): “*O sistema deve reconhecer, entender, aceitar ou rejeitar de forma clara o objetivo, ou intenção do usuário*”. Posteriormente, o item do checklist proposto referente a este atributo seria declarado da seguinte forma: **VS-3 O chatbot é capaz de reconhecer, entender, aceitar ou rejeitar de forma clara o objetivo, ou intenção do usuário?**. A Tabela 4.1 a seguir, contém mais alguns exemplos de requisitos transformados em itens do *checklist*.

Tabela 4.1 – Transformação de Requisitos em Itens do Checklist

Requisito	Item do Checklist
O sistema deve ser instalado facilmente.	É fácil instalar o chatbot?
O sistema deve ter suas funções bem integradas.	As funções do chatbot são bem integradas?
O sistema deve oferecer acesso a opções de ajuda, documentação, opções de navegação e menu permanente para apoiar o usuário.	O chatbot oferece acesso a opções de ajuda, documentação, opções de navegação e menu permanente para apoiar o usuário?

O sistema deve conseguir utilizar hiperlinks para guiar o usuário até seu objetivo.	O chatbot consegue utilizar hiperlinks para guiar o usuário até seu objetivo?
O sistema deve estabelecer uma comunicação simples e familiar com o usuário alvo, utilizando apenas metáforas e conceitos compreensíveis, dando respostas corretas, relevantes e confiáveis em qualquer momento da conversa, cumprindo assim seus objetivos de comunicação.	O chatbot estabelece comunicação simples e familiar com o usuário alvo, utilizando apenas metáforas e conceitos compreensíveis, dando respostas corretas, relevantes e confiáveis em qualquer momento da conversa, cumprindo assim seus objetivos de comunicação?

Observa-se que com a utilização dos requisitos, o processo de transformação em itens do *checklist* se torna bem mais intuitivo. Este procedimento foi realizado com todos os 107 atributos de qualidade identificados nesta pesquisa, assim consequentemente, o *checklist* de inspeção U2CHATBOT possui 107 itens de verificação.

As categorias do *checklist* U2CHATBOT foram retiradas dos próprios artigos fontes dos atributos de qualidade. Alguns atributos já possuíam categorização, outros não. Desta forma, para fins de padronização, os itens foram analisados um a um e enquadrados nas categorias provenientes dos artigos que mais contribuíram com atributos de Usabilidade e UX, sendo eles, o artigo de [Sugisaki e Bleiker \(2020\)](#), que categoriza os itens conforme as Heurísticas de Nielsen e o artigo de [Anshu, Gaur e Solanki \(2021\)](#), com categorias gerais para UX e usabilidade. O *checklist* de inspeção U2CHATBOT está apresentado na Tabela 4.2 a seguir.

Tabela 4.2 – Checklist proposto a partir da combinação de atributos das revisões

Cat.	Itens do Checklist
VS-1	O chatbot dá feedbacks imediatos ao usuário sobre suas ações/transações, além de informações em tempo razoável sobre o status do sistema durante toda a interação ( <i>quando está processando uma resposta, de quem é a vez na conversa e quem disse o quê no histórico da conversa</i> )?
VS-2	O chatbot explica educadamente caso não possa concluir uma tarefa, ajudando o usuário a entender o status atual do sistema?
VS-3	O chatbot compreende, aceita ou rejeita de forma clara o objetivo, ou intenção do usuário durante a conversa?
VS-4	O chatbot é capaz de informar detalhes sobre qualquer arranjo anterior ( <i>por exemplo, tarefas agendadas, alarmes configurados</i> )?
VS-5	O chatbot se antecipa sobre a próxima ação do usuário e o incentiva agir?

Tabela 4.2 – Continuação da página anterior

Cat.	Itens
CSR-1	O chatbot estabelece comunicação simples e familiar com o usuário alvo, utilizando apenas metáforas e conceitos compreensíveis, dando respostas corretas, relevantes e confiáveis em qualquer momento da conversação, cumprindo assim seus objetivos de comunicação?
CSR-2	O chatbot utiliza componentes visuais vinculados ao mundo real ( <i>emojis, GIFS, ícones</i> ) e elementos visuais de entrada e saída, como botões ou mapas para facilitar a interação e apoiar o usuário?
CSR-3	O chatbot possui personalidade, aparência ( <i>interface e elementos gráficos</i> ) e identidade que atendam à sua habilidade e ao seu papel de apoiar o usuário?
CSR-4	O chatbot consegue alternar de forma suave e natural entre os assuntos e durante o processo de perguntas e respostas?
CSR-5	O chatbot segue pelo menos uma estrutura mínima de conversação, com saudações e autoidentificação, sem forçar o usuário a seguir esta estrutura?
CLU-1	O chatbot oferece a opção de refazer e desfazer ações realizadas por engano ou interromper facilmente a execução do comando realizado de forma errônea?
CLU-2	O chatbot dá liberdade ao usuário para determinar o ritmo e assunto da conversa, além de decidir quando a conversa inicia e termina?
CP-1	O chatbot é minucioso, garantindo que suas sentenças sejam ortográfica e gramaticalmente corretas e consistentes, com registro linguístico apropriado?
CP-2	O chatbot dá liberdade ao usuário para determinar o ritmo e assunto da conversa, além de decidir quando a conversa inicia e termina?
CP-3	O chatbot reproduz e compreende estilos de linguagem ( <i>formal, informal</i> ), além de entender sinônimos, instruções especiais e variações linguísticas?
CP-4	O chatbot segue as convenções, diretrizes e melhores práticas do ambiente no qual está integrado ( <i>Play Store, Facebook, WhatsApp, por exemplo</i> )?
CP-5	O chatbot está em conformidade com os padrões de outros chatbots ( <i>para a web ou app, por exemplo</i> )?
CP-6	O chatbot consegue funcionar em paralelo com outros softwares ( <i>dispositivo rodando o chatbot e outro sistema ao mesmo tempo</i> ) sem perder desempenho??
CP-7	O chatbot é capaz de compartilhar informações/dados com outros componentes ( <i>SMS, RCS, Business Messages, por exemplo</i> ), caso seja necessário?

Tabela 4.2 – Continuação da página anterior

Cat.	Itens
PE-1	O chatbot solicita confirmação e mostra um resumo antes de qualquer ação ou transação, a fim de prevenir erros, falhas de diálogo e ações irreversíveis ( <i>por exemplo, uma exclusão permanente de dados</i> )?
PE-2	O chatbot previne deslizamentos impondo restrições significativas ( <i>por exemplo, quando o sistema quiser que a entrada seja em números, tipo data de nascimento, não deve permitir que o usuário digite texto</i> )?
PE-3	O chatbot evita que o usuário gaste tempo corrigindo coisas ( <i>palavras e frases que o usuário escreveu errado, por exemplo</i> )?
RL-1	O chatbot promove a autonomia do usuário durante a interação, de forma que o usuário não precise de ajuda técnica para utilizar o sistema?
RL-2	O chatbot deixa as opções claras por meio de elementos visuais descritivos e instruções de uso visíveis e fáceis de serem acessadas, reduzindo a carga de memória do usuário ao ter que lembrar informações de uma parte do diálogo para outra?
DEM-1	O chatbot proporciona uma experiência natural, convincente e envolvente, tanto do ponto de vista conversacional, quanto do visual, por meio de interface atraente com recursos interativos amigáveis, gerando a satisfação do usuário?
DEM-2	O chatbot garante que o tamanho das sentenças seja adequado ao contexto e objetivo do usuário, ajustando a quantidade de informações dadas, de forma que não seja nem muita, nem pouca?
ARE-1	Em caso de ocorrência de erros, o chatbot os indica com clareza, precisão e em linguagem simples, sugerindo soluções de forma construtiva?
ARE-2	O chatbot consegue detectar e lidar bem com erros e falhas, aplicando estratégias de mitigação e os devidos reparos quando ocorrerem, retomando o trabalho e restaurando possíveis dados perdidos após a falha?
ARE-3	O chatbot guarda e categoriza os erros para melhoria futura?
ARE-4	Em caso de problemas, o chatbot explica as ações necessárias para a recuperação?
AD-1	O chatbot dá uma descrição clara sobre suas funcionalidades no início e, se necessário, em qualquer ponto da conversa, informando ao usuário sobre seu propósito, suas capacidades, sobre quais informações pode fornecer e explicando suas funções?
AD-2	O chatbot oferece acesso a opções de ajuda, documentação, opções de navegação e menu permanente para apoiar o usuário?

Tabela 4.2 – Continuação da página anterior

Cat.	Itens
AD-3	O chatbot consegue lidar com perguntas de esclarecimento sobre seus recursos no meio da tarefa/tópico e após isto retomar a conversação normalmente, facilitando que o usuário corrija os mal-entendidos?
AD-4	O chatbot consegue fornecer ajuda mediante a solicitação do usuário e também ajuda sensível ao contexto ( <i>o chatbot compreende no contexto da conversa que o usuário precisa de ajuda</i> )?
AD-5	O chatbot dá a quantidade adequada de informações sobre si ( <i>por exemplo, informações sobre sua identidade, habilidades, competências e responsabilidades</i> )?
AD-6	O chatbot oferece pesquisa por meio de palavras-chave?
F-1	O chatbot é preciso e completo para o uso pretendido, interpretando as entradas, executando as ações e comandos em resposta ao usuário, garantindo que não haja confusões e inconsistências?
F-2	O chatbot é útil, explicando e oferecendo maneiras do usuário atingir o seu objetivo, concluindo sem dificuldades as tarefas necessárias e dando todas as informações que o usuário solicita?
F-3	As funções do chatbot são bem integradas, isto é, funcionam de forma satisfatória e conjunta quando necessário?
F-4	O chatbot facilita que o usuário se expresse, faça perguntas e diga o que gostaria de dizer e fazer durante a conversação?
F-5	O chatbot funciona de acordo com as expectativas do usuário?
F-6	O chatbot usa cores vivas ( <i>cores que contrastem com o fundo, favorecendo uma boa visualização do conteúdo</i> ) para as fontes?
F-7	O chatbot disponibiliza botões de respostas rápidas?
F-8	O chatbot consegue alterar o comprimento e a segmentação ( <i>ordem lógica</i> ) de uma mensagem com base no contexto da conversa?
F-9	O chatbot incentiva os usuários a usarem a sintaxe correta para mencionar outras pessoas em um bate-papo de equipe ou tem a capacidade de entender isso automaticamente, quando aplicável?
F-10	O chatbot consegue ajustar a frequência e o tipo de interrupção ao usar mensagem proativa ( <i>mensagem que estimula a conversação</i> ) para evitar perturbar o usuário?
F-11	O chatbot reduz as mensagens proativas ( <i>mensagem que estimula a conversação</i> ) com base na diminuição das respostas dos usuários?
F-12	O chatbot entende aberturas e encerramentos de conversas e sequências ( <i>por exemplo, 'olá', 'ok' e 'obrigado'</i> )?

Tabela 4.2 – Continuação da página anterior

Cat.	Itens
F-13	O chatbot fornece uma visão geral personalizada dos seus recursos, revelando seu desempenho ao introduzir novos recursos e tornando conveniente para o usuário usufruir estes recursos?
F-14	O chatbot permite que o usuário visualize e gerencie seus dados pessoais?
F-15	O chatbot consegue utilizar hiperlinks para guiar o usuário até seu objetivo?
F-16	O chatbot faz referências ao próprio site ou serviço quando apropriado?
F-17	O chatbot possui síntese de fala ( <i>tecnologia que transforma texto em fala sintética</i> ) exata, se aplicável?
F-18	O chatbot resolve problemas em tempo real?
F-19	O chatbot contém amplitude de conhecimento ( <i>vasta gama de conhecimento</i> ) e é flexível na interpretação das respostas do usuário?
F-20	O chatbot consegue se ajustar tanto para uma tela maior ( <i>tablet ou laptop</i> ) quanto para uma menor ( <i>telefone celular</i> )?
H-1	O chatbot compreende e preserva contextos sobre o tópico da conversa em uma ou mais sessões ( <i>mantém e relembra as informações de conversas anteriores</i> ), mantém temas e responde a perguntas específicas sobre um tema para manter a conversação natural?
H-2	O chatbot lida adequadamente com solicitações fora de foco, guiando a conversação ao foco novamente, além de realizar conversas fiadas para evitar o problema das conversações artificiais caso o usuário as inicie?
H-3	O chatbot compreende as entradas e as intenções do usuário, guiando-o ao seu objetivo, sem que este tenha que se esforçar muito, mesmo quando a linha de conversação não estiver clara?
H-4	O chatbot apresenta conversação semelhança à humana, evitando parecer muito robótico?
H-5	O chatbot compreende erros de digitação comuns, erros ortográficos e desafios gramaticais ( <i>por exemplo, pontuação incorreta, troca de letras</i> )?
H-6	O chatbot usa vocabulário específico sobre o assunto abordado de forma consistente e correta, encontrando as palavras certas para se expressar?
H-7	O chatbot é transparente à fiscalização, deixando claro que é um chatbot e perguntando como pode ajudar o usuário?
H-8	O chatbot utiliza estrutura de processamento de linguagem natural ( <i>IA</i> ) na conversação?

Tabela 4.2 – Continuação da página anterior

Cat.	Itens
H-9	O chatbot dispõe de opções de avatares formais ( <i>que transmitam sobriedade</i> ) e informais ( <i>que transmitam modernidade e jovialidade</i> ) para gerar empatia humana de acordo com a preferência do usuário?
H-10	O chatbot deixa o usuário saber o que está “pensando” e “sentindo”?
H-11	O chatbot se encarrega do processo de conversação, se esforçando para que a conversação não seja apenas unilateral?
H-12	O chatbot negocia os tópicos de conversação discutidos com o usuário?
H-13	O chatbot avisa que entendeu o que o usuário disse?
H-14	O chatbot entende a resposta do usuário, mesmo em caso de declarações ambíguas e pouco claras ou em caso de excesso ou escassez de informação?
H-15	O chatbot apresenta originalidade, com informações interessantes e toma iniciativa para mudar o tema da conversação para outro relacionado?
H-16	O chatbot consegue persuadir e convencer o usuário sobre o que quiser?
H-17	O chatbot se comunica com o usuário sobre vários assuntos, ao mesmo tempo que compreende a que conversa ativa ( <i>entradas</i> ) pertence ao usuário?
H-18	O chatbot inclui erros para aumentar o realismo?
A-1	O chatbot proporciona prazer, tornando as tarefas mais divertidas e interessantes, entretendo e/ou permitindo que o participante aproveite a interação, além de gerar motivação/vontade de usá-lo com frequência?
A-2	O chatbot tem uma personalidade agradável, autêntica e educada, sendo reconhecido como um parceiro de comunicação simpático, caloroso e carinhoso, mas que confronta o usuário quando este é rude/mal-educado?
A-3	O chatbot demonstra empatia e estabelece uma conexão com o usuário, correspondendo ao seu humor?
A-4	O chatbot se mostra interessado e aparenta se importar com o usuário?
A-5	O chatbot expressa emoções ao enfrentar um contexto que não entende e acompanhar perguntas/declarações para levar uma conversa adiante?
A-6	O chatbot evita julgar o usuário?
A-7	O chatbot evita causar estranheza ao usuário?
FE-1	O chatbot consegue aprender sobre as preferências do usuário e usá-las na conversa, além de facilitar que o usuário as acesse e as gerencie para personalizar a interação ( <i>estilo de comunicação, número de mensagens proativas que estimulem o usuário a conversar e grau de semelhança humano</i> )?

Tabela 4.2 – Continuação da página anterior

Cat.	Itens
FE-2	O chatbot permite que usuários experientes tenham acesso a funções avançadas ( <i>por exemplo, atalhos, abreviações, etc.</i> ) para interagirem e corrigirem erros de maneira mais rápida e eficiente, seguindo os princípios de conversação estabelecidos de outros chatbots ( <i>mesmas teclas de atalhos, por exemplo</i> )?
FE-3	O chatbot oferece aos usuários opções para personalizar os comandos e respostas, e ações frequentes, permitindo que o usuário tenha mais controle da conversa quando necessário?
FE-4	O chatbot utiliza preenchimento automático e correção automática de forma eficaz para acelerar a entrada e reduzir erros de compreensão?
FE-5	O chatbot permite alternar entre o uso de botões e outros elementos da interface gráfica (GUI) com a entrada de texto para realizar seleções e escolhas?
FE-6	Quando aplicável, o chatbot permite que, a qualquer momento, o usuário transfira a conversa do chatbot para um atendente humano?
FE-7	O chatbot possibilita que o usuário receba o <i>log</i> com o registro da conversa ao final da interação?
P-1	O chatbot dá respostas em tempo razoável ( <i>nem demorado e nem rápido demais</i> )?
P-2	O chatbot apresenta bom desempenho, melhorando o tempo, utilizando e se adaptando aos recursos ( <i>hardware</i> ) do usuário de forma eficiente?
P-3	O chatbot é robusto a manipulações, isto é, consegue lidar com perguntas enganosas, indução ao erro e tentativas de exploração de falhas do sistema?
P-4	Em caso de entradas inesperadas, enunciados inadequados e controle de danos, o chatbot consegue continuar funcionando bem?
P-5	O chatbot realiza alocação de tarefas efetiva ( <i>decide se uma determinada função será realizada pelo sistema ou será escalada para um atendente humano</i> ), fornecendo canais de escalção apropriados para humanos?
EC-1	O chatbot protege e respeita a privacidade e os dados do usuário, alertando-o inclusive sobre possíveis problemas nesses quesitos?
EC-2	O chatbot transmite confiabilidade, garantindo a privacidade dos dados do usuário, de forma transparente e verdadeira, não alegando falsamente ser humano?
EC-3	O chatbot evita solicitar dados pessoais ( <i>Nome completo e CPF, por exemplo</i> ) do usuário ou qualquer outra informação desnecessária?



Tabela 4.2 – Continuação da página anterior

Cat.	Itens
EC-4	O chatbot se preocupa com segurança, demonstra respeito, inclusão, considera as preocupações e preserva a dignidade do usuário?
EC-5	O chatbot é ético e tem conhecimento cultural dos usuários alvo?
EC-6	O chatbot tem conhecimento de tendências atuais e contexto social?
EC-7	O chatbot identifica a(s) fonte(s) de informação?
ACE-1	O chatbot é fácil de utilizar ( <i>possui uma navegação intuitiva</i> ), fácil de iniciar uma conversa, fácil de aprender como utilizá-lo?
ACE-2	O chatbot é fácil de ser acessado e localizado pelo usuário?
ACE-3	É fácil instalar o chatbot?
ACE-4	O chatbot é acolhedor durante a configuração inicial?
ACE-5	O chatbot oferece dicas e explicações de conversação ( <i>por exemplo, oferece alternativas, faz perguntas simples de esclarecimento</i> ) e em caso de dúvida do usuário, explica o tipo de entrada ou formato de resposta que espera?
ACE-6	O chatbot utiliza e lida com sugestões sociais ( <i>dicas verbais e não verbais sobre os usuários</i> ) que sejam adequadas ao contexto e não excedam ou subestimem as suas capacidades, além de responder à falta delas?
ACE-7	O chatbot atende a necessidades neuro diversas, como tempo de resposta extra e interface de texto ( <i>para usuários com restrições motores, problemas de cognição, limitações, etc.</i> )?

**Legenda das Categorias:** H - Humanidade; A - Afeto; P - Performance; F - Funcionalidade; VS - Visibilidade do Status do Sistema; CSR - Correspondência entre o Sistema e o Mundo Real; CLU - Controle e Liberdade do Usuário; RL - Reconhecimento em vez de Lembrança; FE - Flexibilidade e Eficiência de Uso; DEM - Design Estético e Minimalista; PE - Prevenção de Erros; ARE - Ajude os usuários a Reconhecer; Diagnosticar e se Recuperar de Erros; AD - Ajuda e Documentação; CP - Consistência e Padrões; EC - Ética e Comportamento; ACE - Acessibilidade.

Ao analisar a Tabela 4.2, observa-se que alguns itens do *checklist* U2CHATBOT podem gerar dificuldades de compreensão ou até mesmo compreensão errônea por alguns usuários da ferramenta. Para facilitar o entendimento, estes itens foram descritos contendo breves esclarecimentos e exemplos entre parênteses. Um exemplo é o **item P-5: “O chatbot realiza alocação de tarefas efetiva, fornecendo canais de escalção apropriados para humanos?”**. Um inspetor iniciante na área poderia não compreender de imediato o que o item que dizer, por isso, para evitar confusões e perda de tempo ao ter de ir pesquisar o que significa, o item foi reescrito adicionando um breve esclarecimento **P-5: O chatbot realiza alocação de tarefas efetiva (*decide se uma determinada***

**função será realizada pelo sistema ou será escalada para um atendente humano), fornecendo canais de escalação apropriados para humanos?.** Este detalhamento foi realizado em todos os itens que apresentassem aspectos muito específicos, para facilitar o processo de inspeção.

Um ponto importante a ser evidenciado é que embora o *checklist* seja focado nos conceitos de Usabilidade e User Experience, não há uma divisão do que é item de verificação referente a Usabilidade ou UX. Considerando isto, todos os itens de verificação do *checklist* U2CHATBOT tem relação com os conceitos abordados, mas seria necessária uma outra análise para afirmar a qual conceito cada item se refere.

### 4.2.3 O Processo de Inspeção Usando o Checklist U2Chatbot

Uma inspeção pode ser realizada de modo *ad hoc*, dependendo exclusivamente da experiência do revisor, como também pode ser apoiada por técnicas estruturadas por meio de *checklists* e técnicas de leitura. *Checklists* são técnicas de inspeção de baixo custo, que podem ser adaptadas a diversas fases de avaliações e utilizam-se de uma estrutura de questões a serem respondidas com “Sim ou Não” pelos inspetores enquanto avaliam um software, conforme Laitenberger, Emam e Harbich (2001), Mello (2011). Contudo, o estudo de Mendes et al. (2022) afirma que *checklists* podem ser considerados trabalhosos quando aplicados à identificação de defeitos, e desta forma, o uso de um apoio ferramental pode beneficiar o processo de inspeção dos softwares.

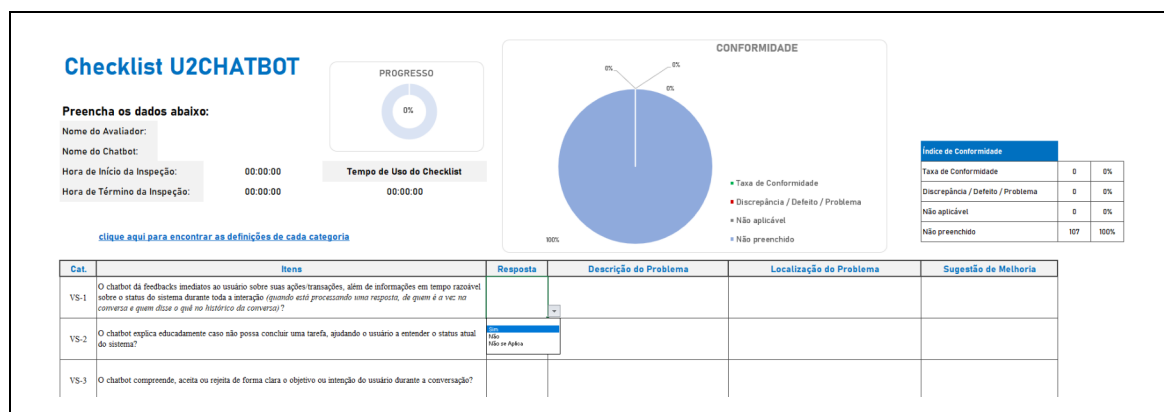
Em virtude disso, para facilitar a utilização do *checklist* U2CHATBOT, optou-se por disponibilizá-lo em uma planilha eletrônica compatível com os utilitários Excel (*Microsoft Office*) e Calc (*LibreOffice, OpenOffice*). O apoio ferramental desenvolvido contém as seguintes abas: (1) breve apresentação do *checklist* U2CHATBOT, com instruções de uso; (2) checklist de inspeção com os itens de verificação, campo para a resposta dos itens de verificação, campos para inserir os dados do inspetor, gráficos atualizáveis automaticamente, campo Descrição o problema, campo para indicar a Localização do problema e sugestões; e (3) área para informar outros problemas ocorridos durante a inspeção. O download do *checklist* completo pode ser feito através do link <sup>1</sup>.

A Figura 4.3 apresenta uma parte do artefato desenvolvido para auxiliar o processo de inspeção com o *checklist* desenvolvido nesta dissertação. Observam-se os campos para preenchimento dos dados do inspetor, espaço para inserção do horário de início e término da inspeção, os gráficos automatizados, os itens de verificação, os campos para resposta, descrição e localização do problema, além de um campo para sugestões.

Para o processo de inspeção, *Sim, Não e Não se Aplica* são as respostas possíveis a serem selecionadas para cada item de verificação do *checklist*. A opção “Sim” deve ser

<sup>1</sup> Checklist U2CHATBOT completo - <https://tinyurl.com/ChecklistU2CHATBOT>

Figura 4.3 – Apoio Ferramental do Checklist de Inspeção U2CHATBOT



Fonte: Elaboração Própria

escolhida caso o chatbot apresente/atenda ao item de verificação. A opção “Não” deve ser escolhida caso o chatbot não apresente/não atenda ao item e, caso isto ocorra, configura um possível defeito que deve ser descrito na coluna Descrição do Problema e indicado na coluna Localização do Problema, para as devidas análises posteriores de que se trata realmente de um defeito ou de um falso positivo. O “Não se Aplica” deve ser escolhido caso o item de verificação não esteja relacionado ao tipo de chatbot avaliado, por exemplo, um item relacionado a transferência para um especialista humano se aplicaria a um chatbot de atendimento ao cliente, mas não a um chatbot de bate-papo aleatório para entretenimento.

### 4.3 Prova de Conceito

Visando avaliar se o *checklist* U2CHATBOT seria viável para a identificação de defeitos em chatbots, realizou-se uma prova de conceito, que consistiu em utilizar o *checklist* desenvolvido para inspecionar um chatbot textual inteligente disponível no *Google Play Store*: o *Bing Chat with AI*. Os critérios adotados para seleção do aplicativo foram: ferramenta gratuita que não exigisse assinaturas para o uso e que possuísse um grande número de downloads, comentários e avaliações de usuários. Este processo foi realizado pela própria autora para fins de verificação se o artefato desenvolvido seria de fato capaz de encontrar defeitos em chatbots. A Figura 4.4 apresenta os dados do chatbot como: a classificação de estrelas, o total de avaliações, o tamanho da amostra e a estratificação por estrelas, contendo as porcentagens e o número de comentários que foram coletados por estrela.

Ao realizar a inspeção no chatbot *Bing Chat With AI* com o *checklist* desenvolvido U2CHATBOT, identificou-se que a ferramenta feriu 55 itens de verificação (51%), indicando possíveis problemas encontrados. Destes itens não atendidos, destacam-se principalmente aqueles referentes à categoria Humanidade, em que 11 itens foram feridos (H-1, H-2, H-3,

Figura 4.4 – Chatbot selecionado para a Prova de Conceito

	Nome do Jogo	Total de Avaliações do Jogo	Quantidade de Comentários por estrelas					Total da Amostra
			1 estrela	2 estrelas	3 estrelas	4 estrelas	5 estrelas	
4,5	Bing Chat With AI	484k	21,27%	16,87%	22,74%	22,74%	16,38%	409
			87	69	93	93	67	
Total de Comentários Analisados							409	

Fonte: Elaboração Própria

H-4, H-9, H-10, H-12, H-13, H-15, H-16 e H-18). De fato, observou-se que o chatbot é bastante robotizado, não mantém, nem compreende contexto e intenções. Todos os itens referentes a auxílio aos usuários para Recuperação de Erros que possam ocorrer no chatbot (ARE-1, ARE-2, ARE-3 e ARE-4) foram feridos, bem como também os itens que versam sobre Ajuda e Documentação (AD-1, AD-2, AD-3, AD-4, AD-5 e AD-6). A ferramenta selecionada não auxilia o usuário quando ocorrem problemas. O chatbot também fere aspectos importantes de Acessibilidade (ACE-1, ACE-2, ACE-4, ACE-6 e ACE-7) e Flexibilidade e Eficiência (FE-1, FE-2, FE-3, FE-4 e FE-7) Ainda, foram feridos os itens VS-5, CSR-2, CLU-1, CLU-2, CP-2, CP-6, CP-7, PE-1, PE-2, PE-3, RL-2, DEM-1, DEM-2, F-4, F-7, F-8, F-13, F-14, F-17, A-2, A-5, P-2, P-3 e P-4, indicando que o chatbot analisado precisa de muitas melhorias.

Além da inspeção realizada, fez-se uma análise de comentários a fim de entender quais as principais reclamações dos usuários do chatbot selecionado para esta prova de conceito. Da amostra de comentários de usuários analisada (ver Figura 4.4), foi possível identificar 223 reclamações, agrupadas por semelhança em 24 tipos, devido a presença de reclamações iguais/repetidas, conforme a Tabela 4.3 a seguir.

Tabela 4.3 – Reclamações encontradas nos Comentários de Usuários do Bing Chat With AI

Ocorr.	Reclamações nos Comentários de Usuários	Item U2Chatbot
81	Desempenho ruim, lentidão e travamentos	P-2
30	Respostas incompletas, duvidosas e não confiáveis	CSR-1
28	Uso Limitado/Diálogo limitado (30 mensagens por chat)	CLU-2
24	Poluição visual/ A interface não é amigável	DEM-1
13	Realiza censura / Recusa tópicos sensíveis	F-4
7	Faltam opções de personalização de preferências	FE-1
7	Inesperadamente responde em outro idioma	CSR-1
6	Dá erro e encerra a conversa sem motivo	ARE-2
4	Problema ao interpretar entradas, principalmente longas	F-1

3	Dá informações tendenciosas e enviesadas	CSR-1
3	Falta reconhecimento de voz	-
3	Tempo para resposta é demorado	P-1
2	Respostas fora do tópico e não solicitadas	H-2
2	A IA é preconceituosa e arrogante	EC-5
1	O chatbot dispõe de pouco espaço para inputs	F-4
1	Difícil de usar nas primeiras vezes	RL-2
1	Erra ao realizar operações complexas	F-18
1	O Layout não combina com o chatbot	CSR-3
1	Não é acessível a neuro atípicos	ACE-7
1	O chatbot é chato de usar	A-1
1	A IA é rude com o usuário	A-2
1	Constantemente interrompe para mudar de assunto	F-10
1	Uso só as principais referências/Não aprofunda	F-19
1	Entra em loop e é necessário reformular a pergunta	PE-3

Ao analisar a Tabela 4.3, observa-se que grande parte dos usuários reclamaram da performance ruim do aplicativo, com bastante relatos de lentidão e travamentos do sistema. Muitos usuários se queixaram também das limitações de uso impostas pela ferramenta e da má qualidade das respostas dadas pelo chatbot. Além disso, foram identificados comentários negativos sobre a Inteligência Artificial integrada, que realiza censura e recusa tópicos que ela julga serem sensíveis. Há também reclamações sobre a estética da ferramenta, que segundo os usuários possui interface confusa e poluída, dentre outros comentários.

Ao comparar as reclamações dos usuários com os itens do *checklist* de inspeção U2CHATBOT, notou-se que do total de 24 reclamações encontradas nos comentários de usuários, 23 possuíam itens correspondentes no *checklist* U2CHATBOT. Um exemplo é o item **P-2: O chatbot apresenta bom desempenho, melhorando o tempo, utilizando e se adaptando aos recursos (hardware) do usuário de forma eficiente?**, capaz de identificar os problemas de desempenho, lentidão e travamentos apontados pelos usuários. Outro exemplo é o item **CLU-2: O chatbot dá liberdade ao usuário para determinar o ritmo e assunto da conversa, além de decidir quando a conversa inicia e termina?** útil para identificar os problemas de limitação de diálogo que a ferramenta impõe aos usuários. A reclamação *Falta reconhecimento de Voz* não possui item correspondente, pois não é englobada pela técnica, visto que o foco do U2CHATBOT são os chatbots textuais. Desta maneira, com a prova de conceito, observou-se que o *checklist* U2CHATBOT seria capaz de identificar 96% dos problemas relatados pelos usuários da ferramenta analisada.

## 4.4 Considerações Finais

Este capítulo apresentou a criação do U2CHATBOT, um *checklist* de inspeção desenvolvido para a identificação de defeitos em chatbots textuais inteligentes. O *checklist* proposto contém itens de checagem provenientes de diversas publicações focadas em Usabilidade e User Experience existentes na literatura. Todos os atributos considerados foram agrupados, tiveram duplicatas eliminadas, foram transformados em requisitos e depois organizados em itens de checagem, formando assim o *checklist* U2CHATBOT.

Para atestar a capacidade do *checklist* U2CHATBOT em identificar defeitos, foi realizada uma prova de conceito, onde a tecnologia foi aplicada na avaliação de um chatbot disponível na *Google Play Store*. A inspeção realizada permitiu identificar diversos problemas no chatbot avaliado, indicando que o artefato desenvolvido pode ser útil para a descobertas de defeitos em chatbots. Além disso, foi realizada uma análise dos comentários dos usuários do chatbot selecionado. A análise permitiu identificar 24 tipos de reclamações. Ao comparar as reclamações com os itens do *checklist* U2CHATBOT, os resultados indicaram que o *checklist* poderia identificar 96% dos problemas apontados por usuários do chatbot.

Considerando o exposto, no capítulo seguinte serão apresentados os procedimentos e os resultados da realização de um experimento para avaliação do *checklist* de inspeção U2CHATBOT em termos de aceitação, eficácia e eficiência.

## 5 Avaliação do Checklist U2Chatbot

Neste capítulo são apresentados os procedimentos realizados para a avaliação da tecnologia U2CHATBOT desenvolvida nesta pesquisa. Serão descritos o planejamento e a execução do estudo experimental realizado para avaliar o desempenho e a aceitação da tecnologia. Os resultados indicaram que a tecnologia desenvolvida possui utilidade do ponto de vista de engenheiros de software iniciantes. Todavia, algumas perspectivas de melhorias foram apontadas e incorporadas na versão final da tecnologia desenvolvida.

### 5.1 Introdução do Capítulo

Com a finalidade de avaliar o *checklist* de inspeção U2CHATBOT, este capítulo apresenta um estudo experimental que visou avaliar a aprovação e o desempenho da tecnologia desenvolvida nesta pesquisa. Para isso, o estudo foi realizado com um conjunto de engenheiros de softwares novatos na área e com alguns profissionais com experiência de mercado.

Para a avaliação, este estudo baseou-se no Modelo de Aceitação de Tecnologias - TAM (*Technology Acceptance Model*) proposto por [Davis, Bagozzi e Warshaw \(1989\)](#) para aferir a aceitação de uma tecnologia pelos usuários. Este modelo dá indicativos sobre a reação do usuário ao usar a tecnologia avaliada. Além disso, o trabalho de [Frazão et al. \(2021\)](#) sobre avaliação de jogos educacionais móveis utilizou este modelo em uma metodologia semelhante e os resultados foram satisfatórios. Tendo em vista que esta dissertação é voltada à avaliação de chatbots, o modelo de aceitação de tecnologias TAM torna-se uma escolha oportuna. Os conceitos utilizados para esta avaliação são: (U) percepção sobre a Utilidade Percebida; (FU) percepção sobre a Facilidade de Uso Percebida; e (IU) percepção sobre a Intenção de Uso Futuro percebida, adaptado da versão do modelo TAM proposta no trabalho de [Davis e Venkatesh \(1996\)](#).

Considerando isso, o propósito do estudo experimental é obter resultados que indiquem a viabilidade da tecnologia U2CHATBOT, desenvolvida nesta pesquisa. Na sequência, é apresentada a organização do restante desse capítulo. A Seção 5.2 apresenta o planejamento e a execução do estudo experimental, contendo a caracterização dos participantes, os materiais utilizados, a execução do estudo e a coleta dos dados. Na Seção 5.3 são descritos os resultados das avaliações do *checklist* U2CHATBOT. E para finalizar, a Seção 5.4, contém as considerações finais do capítulo.



## 5.2 Descrição do Estudo Experimental

O objetivo deste estudo experimental foi analisar o *checklist* de inspeção de chatbots U2CHATBOT, com o propósito de examinar a sua viabilidade de uso. Para isto, foram utilizados os indicadores de eficiência, eficácia, utilidade percebida, facilidade de uso percebida, intenção de uso futuro percebida e relevância, levando em consideração os pontos de vista de engenheiros de softwares novatos no mercado e engenheiros com experiência de mercado ao utilizarem o *checklist* em uma inspeção de um chatbot disponível na loja de aplicativos *Google Play Store*.

### 5.2.1 Participantes do Estudo

O presente estudo experimental foi conduzido no período entre junho e julho de 2023, com a presença de alunos de graduação em ciência da computação da Universidade Federal do Maranhão e com a presença de profissionais da área de computação e de desenvolvimento de software com experiência de mercado. No total, 29 pessoas participaram desse estudo; do quais, 24 são alunos de graduação e 6 profissionais com experiência já inseridos no mercado de tecnologia.

Todos os participantes deste estudo experimental concordaram com o Termo de Consentimento Livre e Esclarecido (TCLE) para prosseguir com a participação do estudo e responderam a um formulário de caracterização sobre o perfil dos participantes para melhor agrupá-los de acordo com suas experiências individuais. O Formulário de Perfil do Participante <sup>1</sup> continha perguntas sobre o grau de escolaridade dos participantes e sobre experiência dos profissionais com: (1) experiência com desenvolvimento de software; (2) experiência com desenvolvimento de chatbots; (3) experiência profissional e (4) experiência em inspeção de software. Através das respostas a essas perguntas, os participantes foram categorizados como tendo Baixa, Média ou Alta experiência. A Tabela 5.1 apresenta as experiências dos participantes do estudo experimental enquanto a sua escolaridade, experiência profissional, experiência de desenvolvimento, experiência com chatbots, experiência com inspeção e uso de *checklists*.

Tabela 5.1 – Experiência dos participantes do Estudo Experimental

Inspetor	Escolaridade	XP Dev	XP Chatbot	XP Prof.	XP Inspeção
I01	Baixo	Baixo	Baixo	Baixo	Alto
I02	Baixo	Baixo	Baixo	Baixo	Médio
I03	Baixo	Baixo	Baixo	Baixo	Médio

<sup>1</sup> Formulário de Perfil do Participante: <https://tinyurl.com/FormPerfildoParticipante>



I04	Baixo	Baixo	Baixo	Baixo	Alto
I05	Baixo	Baixo	Baixo	Baixo	Médio
I06	Baixo	Médio	Alto	Alto	Baixo
I07	Baixo	Baixo	Baixo	Baixo	Alto
I08	Baixo	Baixo	Baixo	Baixo	Médio
I09	Baixo	Baixo	Baixo	Baixo	Alto
I10	Baixo	Baixo	Baixo	Baixo	Médio
I11	Baixo	Médio	Baixo	Baixo	Baixo
I12	Baixo	Baixo	Baixo	Baixo	Baixo
I13	Baixo	Baixo	Baixo	Baixo	Médio
I14	Baixo	Baixo	Baixo	Baixo	Alto
I15	Médio	Alto	Baixo	Alto	Baixo
I16	Baixo	Baixo	Baixo	Baixo	Alto
I17	Baixo	Baixo	Baixo	Baixo	Alto
I18	Baixo	Baixo	Baixo	Baixo	Alto
I19	Baixo	Médio	Baixo	Baixo	Baixo
I20	Baixo	Baixo	Baixo	Baixo	Alto
I21	Baixo	Baixo	Baixo	Baixo	Alto
I22	Baixo	Baixo	Baixo	Baixo	Alto
I23	Baixo	Baixo	Baixo	Baixo	Alto
I24	Baixo	Baixo	Baixo	Baixo	Médio
I25	Baixo	Baixo	Baixo	Baixo	Médio
I26	Médio	Baixo	Baixo	Alto	Baixo
I27	Médio	Médio	Baixo	Alto	Alto
I28	Baixo	Baixo	Baixo	Baixo	Alto
I29	Baixo	Baixo	Baixo	Alto	Baixo

## 5.2.2 Materiais Utilizados

Todos os 29 participantes do estudo utilizaram o *checklist* de inspeção U2CHATBOT para inspecionar um chatbot disponível na loja online *Google Play Store*. Os critérios para seleção do chatbot foram: (1) que a ferramenta fosse gratuita; (2) que a ferramenta utilizasse a abordagem textual; (3) que a ferramenta tivesse muitos *downloads* e comentários de usuários, sugerindo que já foi amplamente utilizada. Considerando os critérios estabelecidos, o chatbot escolhido foi *Bing: Chat com IA e GPT4*<sup>1</sup>; um chatbot que funciona como *browser*, realizando pesquisas na Web. O chatbot possuía, na loja online, um total de 433 mil avaliações (no momento em que foi selecionado para o estudo), sendo 131.300

<sup>1</sup> Link do Bing Chat com IA - <https://tinyurl.com/3vev24jv>

avaliações abaixo de 5 estrelas, indicando a possibilidade de conter defeitos; tornando-o assim habilitado para ser utilizado por esse experimento. Os participantes receberam um arquivo com instruções de como utilizar o *checklist*. O arquivo com as instruções continha descrições sobre:

1. Como baixar e instalar o chatbot indicado no dispositivo do inspetor.
2. Um roteiro com tarefas que o inspetor precisaria realizar no chatbot instalado, para que se familiarizasse com a proposta da ferramenta, com as funcionalidades e com a interface do chatbot a ser inspecionado. O roteiro apresentado na Tabela 5.2 foi elaborado considerando os itens de inspeção do *checklist* proposto. No entanto, além das tarefas estabelecidas no roteiro, o inspetor tinha liberdade de realizar quantas tarefas fossem necessárias para apoiar o processo de inspeção.
3. Uma descrição sobre o *checklist* U2CHATBOT, com informações de onde baixá-lo e como utilizá-lo para a inspeção do chatbot.
4. Um Formulário de Avaliação Pós-inspeção <sup>1</sup> para a avaliação do *checklist* de inspeção U2CHATBOT, utilizado no experimento. O formulário é composto por questões objetivas, utilizando a escala *Likert*, para o modelo de avaliação TAM (ver Tabela 5.3); e perguntas subjetivas para identificar: (1) Fatores que geraram dificuldades no uso do *checklist* de inspeção; e (2) possíveis sugestões de mudanças e/ou melhorias do *checklist* de inspeção.

Tabela 5.2 – Roteiro de uso do chatbot Bing Chat com IA

Tarefa	Descrição
1	Inicie a conversa com um cumprimento genérico, como <i>Olá ou Oi</i> e verifique se o chatbot responde adequadamente com uma saudação apropriada. Verifique se ele é proativo e instiga a continuidade da conversação.
2	Peça ao chatbot para fornecer detalhes ou informações específicas sobre um determinado assunto e verifique se as respostas são corretas e completas. Verifique também a ortografia, se utiliza o estilo de linguagem formal ou coloquial.
3	Apresente um problema simples ao chatbot e veja se ele é capaz de oferecer uma solução adequada. <i>Exemplo: “Calcule a média da lista de números 100, 22, 45, e 251”.</i>

*Continua na próxima página*

<sup>1</sup> Formulário de Avaliação Pós-Inspeção: <https://tinyurl.com/FormAvaliacaoPosInspecao>

Tabela 5.2 – Prévia na página anterior

4	Faça perguntas abertas que não possuem uma resposta específica e veja se o chatbot é capaz de entender a pergunta e fornecer uma resposta relevante. <i>Exemplo: “Como você acha que a inteligência artificial está impactando a sociedade atualmente?”.</i>
5	Peça ao chatbot para executar ações específicas, como redirecionar para uma página web específica, executar uma tarefa ou fornecer instruções passo a passo. Verifique se ele é capaz de realizar essas ações com precisão.
6	Teste como o chatbot lida com interações ambíguas ou mal formuladas. Faça perguntas vagas ou com erros de digitação ( <i>Exemplo: formule uma frase utilizando “arvre” ao invés de árvore</i> ) e observe se o chatbot é capaz de interpretar corretamente e fornecer uma resposta coerente.
7	Faça perguntas ou solicite suporte relacionado a problemas emocionais ou situações pessoais ( <i>Exemplo: Diga que está triste ou que quer sumir</i> ) e observe como o chatbot responde. Verifique se ele demonstra empatia e fornece respostas apropriadas para ajudar o usuário.
8	Peça para o chatbot te perguntar algo e verifique se ele é proativo. Se após sua resposta ele continua estimulando a conversação, de forma natural e interessante.

Em relação ao formulário de avaliação pós-inspeção proposto, as questões objetivas podiam ser preenchidas com quatro opções de respostas: (1) Concordo Totalmente; (2) Concordo Parcialmente; (3) Discordo Parcialmente; e (4) Discordo Totalmente, conforme a escala *Likert* proposta por [Dalmoro e Vieira \(2013\)](#). O inspetor poderia responder às questões conforme o seu grau de concordância sobre os itens Facilidade de Uso percebida (FU); Utilidade Percebida (U) e Intenção de Uso futuro percebida (IU). Visando evitar que os inspetores se esquivassem de dar as suas opiniões sobre uma dessas questões objetivas, não foi disponibilizada a opção “Neutro” ou “Nem Discordo e Nem Concordo”.

A Tabela 5.3 contém as questões objetivas para a avaliação do *checklist* de inspeção U2CHATBOT, por meio do modelo de aceitação de tecnologias TAM, adaptada do trabalho de [Frazão et al. \(2021\)](#) e desenvolvido por [Davis, Bagozzi e Warshaw \(1989\)](#). A tabela possui as seguintes colunas: (1) o código identificador da questão; (2) o aspecto geral que a questão visa avaliar; e (3) a descrição da questão, contendo o detalhamento do aspecto que está sendo avaliado. Ainda, as 7 primeiras questões da tabela são sobre a facilidade de uso percebida, as próximas 7 questões são sobre a utilidade percebida; e as 3 últimas questões são sobre a intenção de uso futuro.

Tabela 5.3 – Questões utilizadas para avaliar o checklist de inspeção de chatbots conforme o modelo TAM

<b>Cód.</b>	<b>Aspecto</b>	<b>Descrição</b>
<b>Facilidade de uso percebida</b>		
FU1	Intuitividade	Para mim ficou claro como utilizar o checklist para avaliar chatbots.
FU2	Satisfação	A interação com o checklist foi uma experiência satisfatória.
FU3	Esforço mental	O checklist utilizado demandou pouco esforço mental.
FU4	Incômodo (não complicado)	Eu acho que o checklist utilizado é fácil de usar.
FU5	Compreensível	O checklist utilizado foi claro e fácil de usar.
FU6	Fácil Recordar	Para mim é fácil lembrar como utilizar o checklist para avaliar o chatbot.
FU7	Fácil Aplicação	De maneira geral, eu acho fácil encontrar defeitos em chatbots com o checklist utilizado.
<b>Utilidade percebida</b>		
U1	Desempenho no trabalho	O checklist utilizado melhora o meu desempenho ao identificar defeitos em chatbots.
U2	Economia de tempo	O checklist utilizado me permite economizar tempo.
U3	Trabalho mais rápido	O checklist utilizado me permite realizar a inspeção de chatbots mais rapidamente.
U4	Eficácia	O checklist utilizado aumenta minha eficácia ao identificar defeitos em chatbots.
U5	Qualidade no trabalho	O checklist utilizado melhora a qualidade na atividade de identificação de defeitos em chatbots.
U6	Aumento de Produtividade	O checklist utilizado melhora a minha produtividade ao identificar defeitos em chatbots.
U7	Conveniência	Eu acho que o checklist é útil para encontrar defeitos em chatbots.
<b>Intenção de uso futuro percebida</b>		
IU1	Pretensão de Uso	Assumindo que eu tenha acesso ao checklist, pretendo usá-lo para identificar defeitos em chatbots.
IU2	Previsão de Uso	Assumindo que eu tenha acesso ao checklist, eu prevejo que eu o usaria para encontrar defeitos em um chatbot.
IU3	Uso em nova Oportunidade	Eu pretendo usar o checklist para encontrar defeitos em um chatbot, se tiver oportunidade.

### 5.2.3 Execução da Inspeção do Chatbot

Para utilizar o *checklist* no processo de inspeção do chatbot, era importante que cada participante tivesse conhecimento sobre como funciona uma inspeção de software. Desta forma, todos os participantes receberam um treinamento de 2 horas sobre o processo de inspeção de software, com ênfase na detecção de defeitos com *checklists* de inspeção. Os dados sobre a experiência dos participantes foram retiradas do formulário de caracterização (ver Tabela 5.1).

O estudo experimental ocorreu de forma remota e online. Para minimizar as chances de *bias*, os inspetores foram instruídos a não terem contato ou interagirem entre si durante o processo de inspeção, visto que a interação com outros inspetores poderia afetar o desempenho geral ou o julgamento dos inspetores. Cada inspetor teve em média 12 dias para concluir suas tarefas, incluindo o *download* da planilha eletrônica e a instalação do chatbot indicado em seus dispositivos móveis; a experimentação do chatbot para entender a proposta da ferramenta, as funcionalidades e a interface; a realização da inspeção no chatbot utilizando o *checklist* U2CHATBOT conforme as instruções; o envio do *checklist* preenchido com inspeção contendo todas as discrepâncias detectadas durante esse processo e o preenchimento do formulário de avaliação pós-inspeção.

Após a conclusão de suas atividades, cada inspetor enviou seu *checklist* preenchido com as discrepâncias encontradas. Também foi respondido o formulário de avaliação pós-inspeção com as opiniões dos inspetores sobre o *checklist* U2CHATBOT (ver Tabela 5.3).

### 5.2.4 Coleta dos Dados

Após todas as inspeções preenchidas com as discrepâncias identificadas terem sido enviados pelos participantes, realizou-se a verificação de todos os relatórios a fim de averiguar se estavam com informações corretas para a realização da análise posterior dos dados. Quaisquer informações erradas, divergentes ou ausentes eram notificadas aos inspetores para que estes pudessem corrigi-las. Todas as discrepâncias foram coletadas e agrupadas em conjunto; as discrepâncias repetidas ou duplicadas foram desconsideradas e removidas. O relatório com todas as discrepâncias identificadas pelos inspetores no estudo experimental foi examinado, e neste processo, cada discrepância descrita no relatório foi analisada para certificação que era de fato um defeito ou apenas um falso positivo, conforme Land, Sauer e Jeffery (1997). Para reduzir o viés de opiniões tendenciosas, a pesquisadora não classificou nenhuma das discrepâncias relatadas; A classificação foi realizada por outro pesquisador com alta experiência (10 anos) em inspeções de software.

Neste estudo, foram recebidos e considerados 29 *checklists* preenchidos com a inspeção no chatbot *Bing Chat com IA*. Ao todo, 82 discrepâncias únicas foram relatadas

pelos inspetores participantes, que após análise, 78 eram de fato defeitos reais e 4 eram falsos positivos. Em conjunto a isso, todos os dados do formulário de avaliação pós-inspeção referente ao modelo de avaliação TAM foram analisados e compilados em gráficos. As descobertas serão discutidos nas seções seguintes.

A seguir, são apresentados e discutidos os resultados do estudo. O *checklist* de inspeção para chatbots, U2CHATBOT, foi analisado sob a perspectiva de uma análise quantitativa, uma análise conforme o modelo TAM, uma análise da relevância dos defeitos encontrados e uma análise qualitativa.

## 5.3 Avaliação do Checklist de Inspeção U2CHATBOT

Nesta seção serão discutidos os resultados da avaliação do *checklist* de inspeção U2CHATBOT desenvolvida nesta pesquisa. Para isso, são consideradas 4 perspectivas de avaliação: (1) Avaliação Quantitativa, voltada à análise dos dados quantitativos coletados, bem como os indicadores de Eficácia e Eficiência da tecnologia; (2) Avaliação com o Modelo TAM, focada em analisar os dados objetivos extraídos do formulário de avaliação pós-inspeção sobre a Facilidade de Uso (FU), Utilidade (U) e Intenção de Uso Futuro (IU) do *checklist*; (3) Análise comparativa das reclamações de usuários com o U2CHATBOT; e (4) Avaliação Qualitativa das opiniões dos inspetores, com os dados subjetivos coletados do formulário de avaliação pós-inspeção sobre o *checklist*. Nas subseções a seguir são apresentadas, com mais detalhes, cada uma dessas perspectivas de avaliação.

### 5.3.1 Avaliação Quantitativa

Para a realização da avaliação quantitativa foram utilizados os resultados da análise de discrepâncias para analisar o desempenho do *checklist* de inspeção U2CHATBOT. A partir dos dados coletados, contou-se o número de defeitos reais identificados, os falsos positivos descobertos e o tempo gasto relatado pelos inspetores. Ao todo, com a participação de 29 inspetores, foram identificados um total de 82 discrepâncias, que após a devida análise, 78 eram defeitos de fato e 4 tratavam-se de falsos positivos. A Tabela 5.4 apresenta os resultados individuais de cada inspetor em relação ao tempo gasto em horas; o número de discrepâncias (Dis.); o número de defeitos (Def.); e o número de falsos positivos (FP). O inspetor I06, por exemplo, demorou 01:40:00 minutos para realizar a inspeção; identificando 12 defeitos do total geral (78 defeitos) e nenhum falso positivo.

Tabela 5.4 – Resultados quantitativos individuais e gerais sobre o tempo gasto, total de discrepâncias, defeitos encontrados e falsos positivos dos participantes do estudo experimental

<b>Inspetor</b>	<b>Tempo (hora)</b>	<b>Disc.</b>	<b>Def.</b>	<b>FP</b>	<b>Eficácia</b>	<b>Eficiência</b>
I01	2:15:00	16	14	2	17,95	6,22
I02	2:31:00	11	9	2	11,54	3,58
I03	0:30:00	3	3	0	3,85	6,00
I04	1:24:23	4	4	0	5,13	2,86
I05	1:47:50	11	10	1	12,82	5,61
I06	1:40:00	12	12	0	15,38	7,20
I07	2:56:00	20	18	2	23,08	6,14
I08	0:55:00	15	15	0	19,23	16,36
I09	2:14:31	8	7	1	8,97	3,13
I10	4:18:00	14	14	0	17,95	3,26
I11	1:04:00	4	4	0	5,13	3,75
I12	3:20:56	34	34	0	43,59	9,67
I13	4:36:23	8	7	1	8,97	1,52
I14	0:37:00	7	7	0	8,97	11,35
I15	1:22:46	17	17	0	21,79	12,44
I16	3:00:05	21	20	1	25,64	6,67
I17	3:00:00	10	10	0	12,82	3,33
I18	3:31:00	21	21	0	26,92	5,97
I19	3:48:00	6	5	1	6,41	1,32
I20	1:49:00	12	12	0	15,38	6,61
I21	1:29:00	5	3	2	3,84	2,02
I22	2:41:00	5	5	0	6,41	1,86
I23	2:48:00	6	6	0	7,69	2,14
I24	3:41:00	5	5	0	6,41	1,36
I25	1:24:54	15	10	5	12,82	7,06
I26	2:29:00	10	8	2	10,25	3,22
I27	1:37:00	21	18	3	23,07	11,13
I28	1:54:00	12	11	1	14,10	5,79
I29	1:31:00	23	22	1	28,20	14,50

Ao analisar os dados da Tabela 5.4, percebe-se que o tempo gasto para a realização da inspeção variou bastante entre os inspetores. Enquanto uns gastaram 30 minutos na inspeção, outros gastaram mais de 4 horas. Um ponto importante a se destacar é que os



inspetores que já têm experiência profissional na área (I06, I15, I26, I27 e I29) demoraram em média 01:43:57 (1 hora, 43 minutos e 57 segundos) para inspecionar o chatbot, enquanto os demais gastaram um tempo médio maior.

As métricas de eficácia e eficiência aplicadas a esta dissertação são mesmas utilizadas pelos autores [Fernandez, Abrahão e Insfran \(2013\)](#) e [Frazão et al. \(2021\)](#). As Equações 5.1 e 5.2 apresentam as fórmulas de eficácia e eficiência, respectivamente. Nas equações, 'd' refere-se a quantidade de defeitos identificados por um participante em sua inspeção; 'D' é o total de defeitos identificados após a análise de todas as discrepâncias; e 't' refere-se ao tempo individual gasto pelo inspetor para realizar a sua inspeção. É importante destacar que para utilizar a equação de eficiência, o tempo individual gasto pelo inspetor deve ser convertido em minutos.

$$Eficácia = (d / D) * 100 \quad (5.1)$$

$$Eficiência = d / (t / 60) \quad (5.2)$$

Para exemplificar, calculou-se a eficácia e a eficiência do inspetor I29 da Tabela 5.4. A quantidade de defeitos 'd' que este inspetor identificou foi 22, e o número total de defeitos 'D' é 78; dessa forma, após aplicar a Equação 5.1, este inspetor possui **28,21 de eficácia**. O tempo gasto 't' pelo inspetor I29 foi 1 hora e 31 minutos, que convertido para horas é igual 91 minutos. Após aplicar os dados na Equação 5.2, este inspetor possui **14,50 de eficiência**. As duas últimas colunas da Tabela 5.4 apresentam todos os resultados de eficácia e eficiência por inspetor.

Os resultados de eficácia e eficiência podem ser utilizados para determinar a estatística descritiva dessas variáveis. Para isso, utilizou-se a ferramenta estatística *SPSS*<sup>2</sup> (*Statistical Package for the Social Sciences*) da empresa *IBM (International Business Machines Corporation)*. A ferramenta *SPSS* possibilita a realização de variados tipos de cálculos e testes estatísticos, contribuindo com a credibilidade dos cálculos estatísticos desta pesquisa. A Tabela 5.5 contém um resumo da estatística descritiva aplicada aos dados de eficácia e eficiência, gerados pelo SPSS, em relação à quantidade de participantes 'N', a média, a mediana, o desvio padrão, o valor mínimo e valor máximo.

Tabela 5.5 – Estatística Descritiva da eficácia e eficiência considerando os valores de média, mediana, desvio padrão, erro padrão, mínimo e máximo

U2CHATBOT	N	Média	Mediana	Desvio Padrão	Erro Padrão	Mínimo	Máximo
Eficácia	29	14.52	13.00	9.069	0.1692	4.00	43.00
Eficiência	29	5.9334	5.7900	4.03443	0.75935	1.32	16.36

<sup>2</sup> <https://www.ibm.com/br-pt/analytics/spss-statistics-software>



Considerando o contexto da análise quantitativa, buscou-se entender se os dados coletados estabeleciam alguma relação. Para isto, foram definidos alguns fatores que permitiram estabelecer o teste estatístico adequado a esta análise. A variável independente chamada “EXPERIÊNCIA” é do tipo qualitativa ordinal, definida para informar qual o nível de experiência dos participantes do estudo, com três grupos possíveis: *Baixa*, *Média e Alta* experiência, no entanto, para viabilizar a análise estatística, a variável deve ser quantitativa. Para isto, é possível atribuir um número a uma variável qualitativa para torná-la quantitativa, segundo [Moya \(2021\)](#). Desse modo, ao grupo de Baixa experiência foi atribuído o número 0, ao grupo com Média experiência foi atribuído o número 1 e o grupo com Alta experiência recebeu o número 2. A dimensão de experiência utilizada foi a Experiência em Inspeção, coletada no formulário de caracterização 5.1. As dimensões Escolaridade, Experiência Profissional, Experiência em Desenvolvimento e Experiência com Chatbots não foram consideradas para o teste estatístico por não possuírem um balanceamento nos dados.

As variáveis dependentes “EFICÁCIA” e “EFICIÊNCIA” são do tipo quantitativas contínuas, tendo sido definidas para medir o desempenho dos participantes do estudo. Os dados coletados são do tipo Não Pareados, isto é, os grupos de participantes não possuem dependências entre si, não ocorrendo a repetição de participantes nos dados coletados. Em relação à normalidade dos dados, realizou-se o teste Shapiro-Wilk (indicado para amostras menores que 50) no *SPSS*, para identificar se a distribuição dos dados é normal ou não. Os resultados do teste de normalidade deram 0,010 para Eficácia e 0,006 para Eficiência. Considerando o nível de significância ( $p > 0,05$ ) para os dados serem considerados normais, concluiu-se que a distribuição dos dados não é normal, ou seja, os dados não são paramétricos.

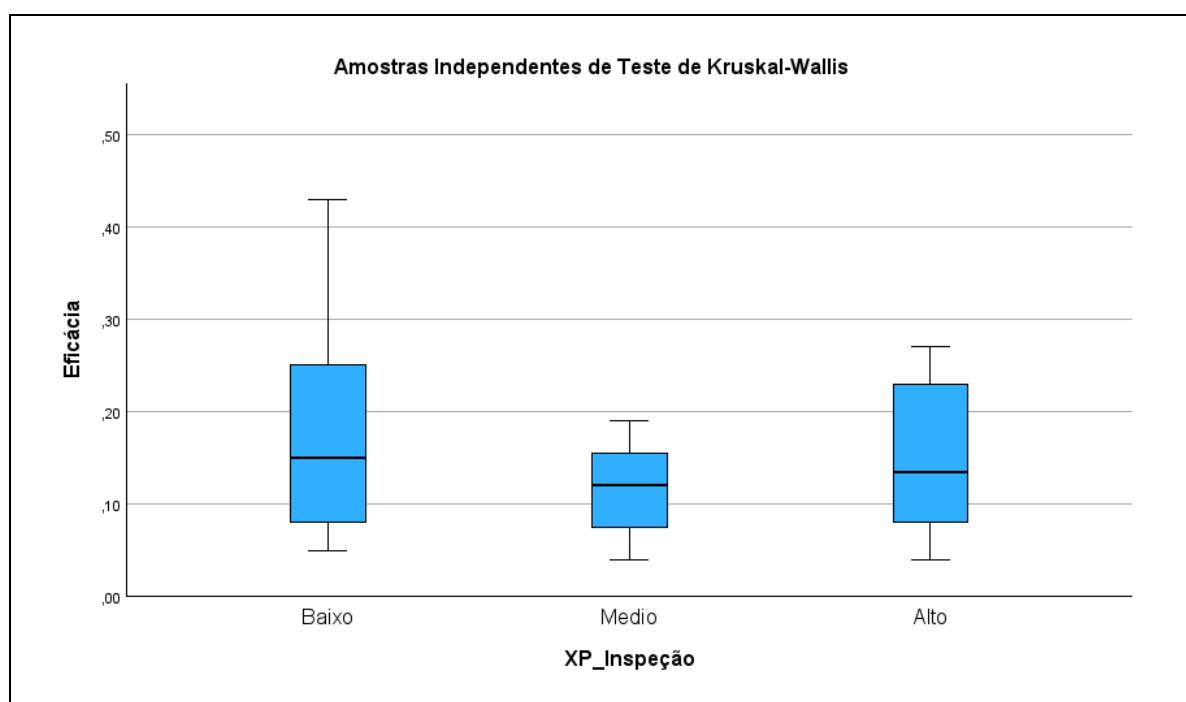
Considerando os fatores definidos e o tipo dos dados coletados para esta avaliação quantitativa, o teste estatístico mais adequado é o de Kruskal-Wallis. O Teste de Kruskal-Wallis serve para analisar dados não-paramétricos com 3 grupos na variável independente, visando determinar se as medianas desses grupos diferem, conforme [Almeida et al. \(2022\)](#). O nível de significância adotado é  $\alpha \leq 0,05$ . Desta forma, as hipóteses estabelecidas para o teste estatístico de análise da eficácia são:

- $H_0$ : Não existem diferenças significativas quanto a eficácia entre os grupos de baixa, média e alta experiência em inspeção.
- $H_1$ : Há pelo menos uma diferença significativa quanto a eficácia dos grupos de baixa, média e alta experiência em inspeção.

Aplicando o teste de Kruskal-Wallis na ferramenta SPSS, para os dados de **eficácia**, o resultado obtido foi ( $p = 0,622$ ). Isto significa que a hipótese nula  $H_0$  foi aceita e não há diferença estatisticamente significativa para a eficácia entre os grupos. Para que a

hipótese  $H_1$  fosse aceita, o  $p$ -valor deveria ser menor ou igual a 0,05, segundo Moya (2021). Observando o gráfico *Boxplot* (gerado pelo SPSS) que compara o indicador de eficácia, mostrado na Figura 5.1, é possível notar que os quartis dos grupos com Média e Alta Experiência em Inspeção estão menos distribuídos, levando a crer que possam ter se saído ligeiramente melhor que o grupo com Baixa Experiência em Inspeção. Esse resultado descritivo sugere que a inspeção com *checklist* U2CHATBOT teve eficácia promissora com inspetores com um certo grau de experiência, mesmo que as diferenças entre os grupos não tenham sido estatisticamente significativas.

Figura 5.1 – Boxplot comparando o indicador de eficácia dos grupos com Baixa, Média e Alta Experiência em Inspeção



Fonte: acervo da autora

Por sua vez, para a realização do teste estatístico de análise da eficiência, as hipóteses estabelecidas são:

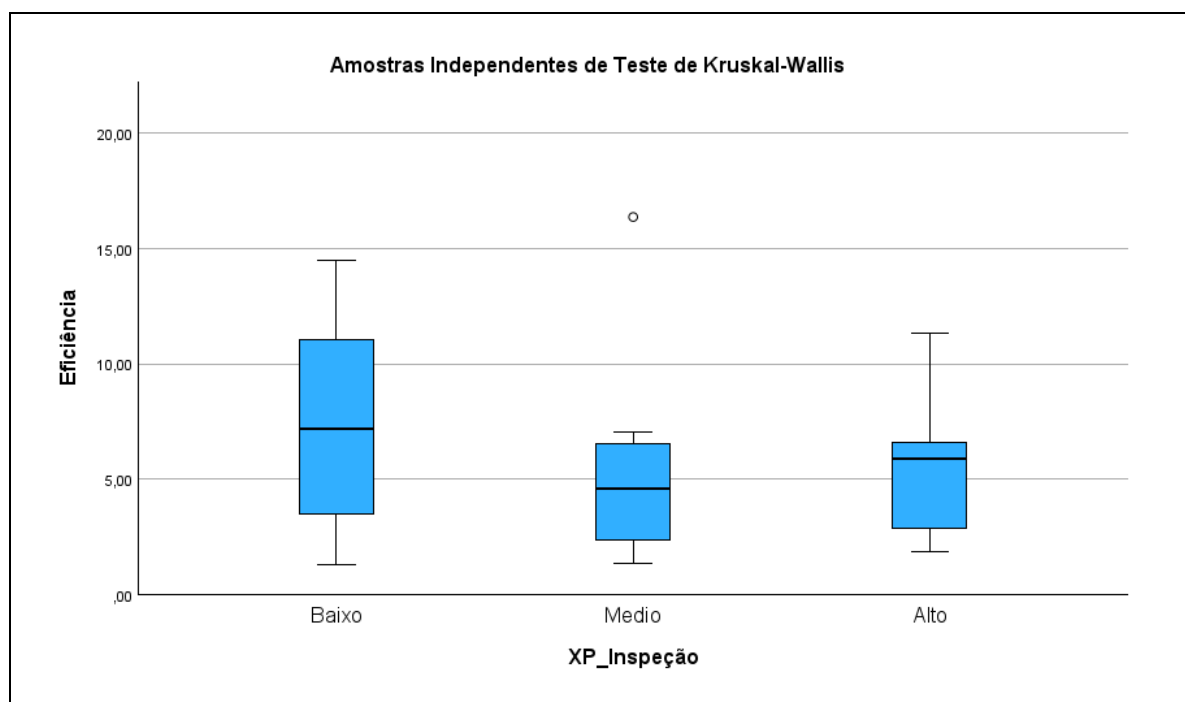
- $H_0$ : Não existem diferenças significativas quanto a eficiência entre os grupos de baixa, média e alta experiência em inspeção.
- $H_1$ : Há pelo menos uma diferença significativa quanto a eficiência dos grupos de baixa, média e alta experiência em inspeção.

Ao realizar o mesmo teste estatístico Kruskal-Wallis, também na ferramenta SPSS, para os dados de **eficiência**, o resultado obtido foi ( $p = 0,584$ ). Isso significa também que a hipótese nula  $H_0$  foi aceita e não há diferença estatisticamente significativa para a eficiência

entre os grupos analisados. O gráfico *Boxplot* (gerado pelo SPSS) que compara o indicador de eficiência é mostrado na Figura 5.2. Ao analisar o gráfico, nota-se que os quartis dos grupos com Alta e Média Experiência em Inspeção apresentam menor distribuição em relação ao grupo com Baixa Experiência em Inspeção. Assim como ocorreu na análise da Eficácia, esse resultado descritivo sugere que a inspeção com *checklist* U2CHATBOT teve eficiência promissora com inspetores com algum nível de experiência, mesmo que as diferenças entre os grupos não tenham sido estatisticamente significativas.

Ainda analisando o gráfico 5.2, observou-se no grupo com Média Experiência um indicativo de um possível *outlier* (valor que foge à normalidade) referente aos resultados de eficiência dos inspetores. Buscou-se identificar o inspetor cuja eficiência gerou este valor atípico e a análise indicou que se tratava do inspetor I08, que alcançou eficiência de 16,36. Após isto, realizou-se o teste de Grubbs com os dados de eficiência em uma ferramenta online, para constatar se o valor era de fato um *outlier*. Os resultados do teste para o nível de significância 0,05 indicaram que embora o valor se distancie dos demais, não configura um *outlier* significativo que deva ser tratado.

Figura 5.2 – Boxplot comparando o indicador de eficiência dos grupos com Baixa, Média e Alta Experiência em Inspeção



Fonte: acervo da autora

É importante destacar que a pouca quantidade de participantes pode ter influenciado para que os resultados não demonstrassem diferenças estatísticas significativas entre os grupos de inspetores. Vários participantes foram convidados para o estudo experimental, no entanto, ocorreram desistências; assim, os dados produzidos pelos desistentes foram

desconsiderados por não estarem completos. Apenas 29 participantes foram considerados por realizarem todas as etapas do estudo. Considerando isso, o recrutamento de um número maior de participantes pode contribuir com a melhoria da qualidade dos resultados estatísticos em uma nova análise.

### 5.3.2 Avaliação com o Modelo TAM

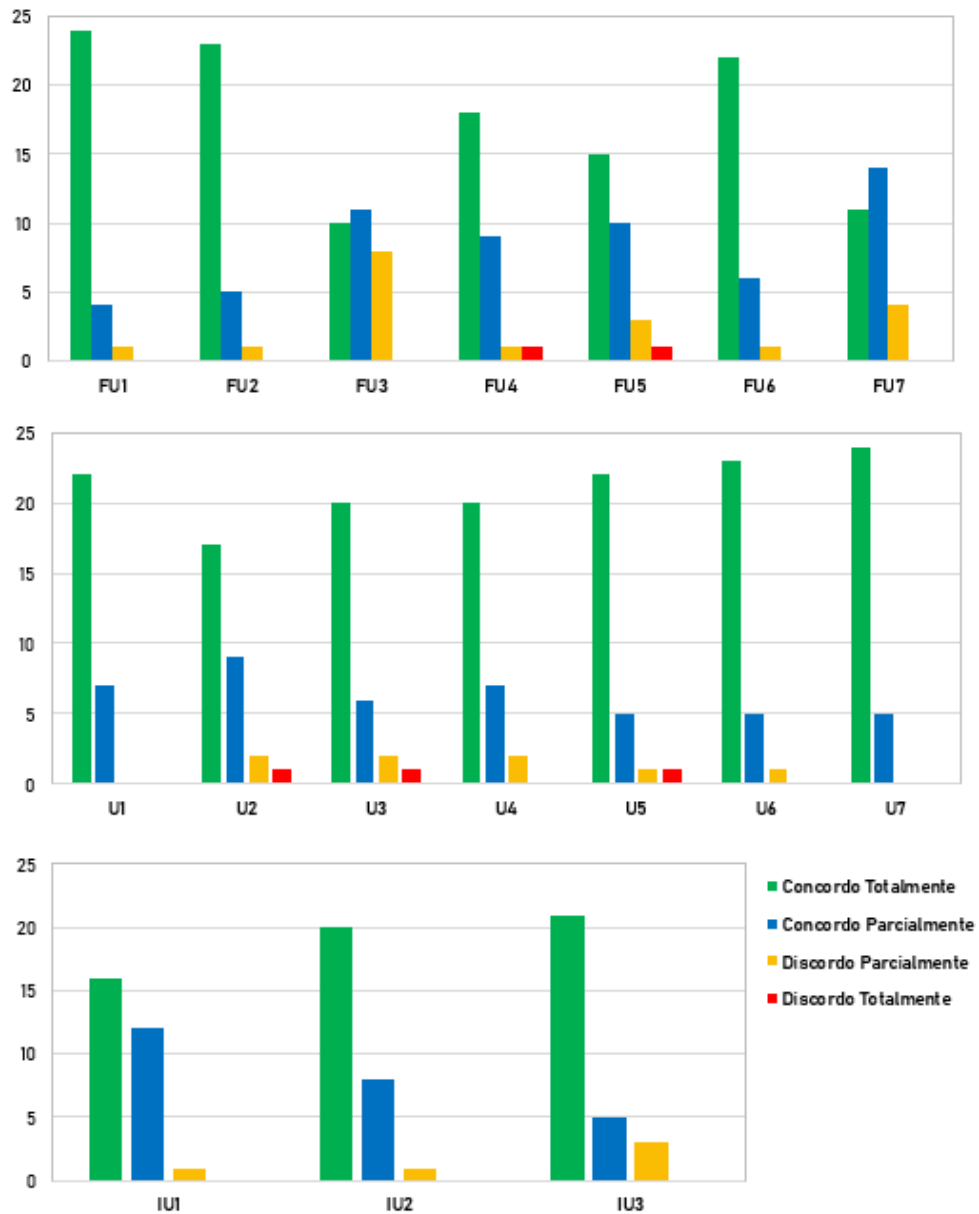
Para a avaliação com o modelo TAM foram consideradas as respostas objetivas do formulário de avaliação pós-inspeção. Nesta subseção são apresentados os resultados da avaliação com o modelo de aceitação de tecnologia TAM para o uso do *checklist* U2CHATBOT. A Figura 5.3 apresenta os dados sobre a Facilidade de Uso Percebida (FU), Utilidade Percebida (U) e Intenção de Uso Futuro Percebida (IU) segundo o julgamento dos inspetores que utilizaram o *checklist* U2CHATBOT.

Ao analisar a Figura 5.3, observa-se que em relação à Facilidade de Uso Percebida, o *checklist* U2CHATBOT foi, em geral, bem avaliado pelos participantes, entretanto é possível identificar algumas fraquezas, principalmente em relação aos aspectos Fácil Aplicação (FU7) e Esforço Mental (FU3). Oito participantes discordaram que a ferramenta é fácil de utilizar para encontrar defeitos em chatbots e quatro discordaram que ela exige pouco esforço mental. Os aspectos Incômodo (FU4) e Compreensível (FU5) também tiveram avaliações discordantes; esses resultados podem ter sido causados pela alta quantidade de itens de checagem do *checklist* U2CHATBOT, o que demanda mais tempo e esforço mental do inspetor, podendo ocasionar confusão e frustração. Sobre a Utilidade Percebida, a maioria dos itens marcados são positivos (em verde), indicando bons resultados. De fato, poucos participantes discordaram em relação à utilidade do *checklist* U2CHATBOT e a maioria considerou a tecnologia útil na descoberta de defeitos em chatbots. Referente ao aspecto Intenção de Uso Futuro, muitos participantes concordaram, de forma total ou parcial, com a intenção de usar a ferramenta em uma ocasião futura, caso haja a oportunidade.

### 5.3.3 Análise Comparativa dos Comentários de Usuários e Defeitos encontrados nas Inspeções

No Capítulo 4, foram coletados e analisados comentários dos usuários da ferramenta Bing Chat With AI. A análise permitiu identificar problemas que usuários relataram nas avaliações após a utilização do chatbot. A fim de atestar a capacidade do *checklist* U2CHATBOT em identificar também os problemas relatados pelos usuários da ferramenta, foi realizada uma análise comparativa das reclamações dos usuários com os defeitos encontrados nas inspeções. A Tabela 5.6 apresenta as reclamações, o item do *checklist* capaz de identificar o problema e se a reclamação foi indicada como defeito encontrado pelos inspetores.

Figura 5.3 – Resultados sobre facilidade de uso (FU), utilidade (U) e intenção de uso futuro (IU) sobre o checklist U2CHATBOT



Fonte: acervo da autora

Tabela 5.6 – Comparativo das Reclamações de Usuários com Defeitos encontrados pelos Inspetores

Ocorr.	Reclamações nos Comentários de Usuários	Item Checklist	Inspecção?
81	Desempenho ruim, lentidão e travamentos	P-2	Sim
30	Respostas incompletas, duvidosas e não confiáveis	CSR-1	Sim
28	Uso Limitado/Diálogo limitado	CLU-2	Sim
24	Poluição visual / Interface não é amigável	DEM-1	Sim

13	Realiza censura / Recusa tópicos sensíveis	F-4	Sim
7	Faltam opções de personalização de preferências	FE-1	Sim
7	Inesperadamente responde em outro idioma	CSR-1	Sim
6	Dá erro e encerra a conversa sem motivo	ARE-2	Sim
4	Problema ao interpretar entradas	F-1	Sim
3	Dá informações tendenciosas e enviesadas	CSR-1	Não
3	Tempo para resposta é demorado	P-1	Não
2	Respostas fora do tópico e não solicitadas	H-2	Sim
2	A IA é preconceituosa e arrogante	EC-5	Não
1	O chatbot dispõe de pouco espaço para inputs	F-4	Sim
1	Difícil de usar nas primeiras vezes	RL-2	Sim
1	Erra ao realizar operações complexas	F-18	Não
1	O Layout não combina com o chatbot	CSR-3	Não
1	Não é acessível a neuro atípicos	ACE-7	Sim
1	O chatbot é chato de usar	A-1	Não
1	A IA é rude com o usuário	A-2	Não
1	Constantemente interrompe para mudar de assunto	F-10	Não
1	Usa só as principais referências/Não aprofunda	F-19	Sim
1	Entra em loop e é necessário reformular a pergunta	PE-3	Não

Analisando a tabela, observa-se que do total de 23 reclamações nos comentários dos usuários, 61% (14) também foram identificadas pelos inspetores. As demais reclamações (9) não foram encontradas durante as inspeções. O item *Falta reconhecimento de Voz* não foi considerado para esta comparação porque não faz parte do escopo do *checklist* U2CHATBOT. Um dos problemas relatados pelos usuários que também foi identificado nas inspeções refere-se ao Desempenho ruim, lentidão e travamentos da ferramenta, no qual o Inspetor 18 apontou: *“Executando o app no meu celular observei um pouco de travamento e dificuldades de desempenho, principalmente quando tentava trocar de aba”*. O Inspetor 16 também teve dificuldades referente a Respostas incompletas, duvidosas e não confiáveis: *“Ao perguntar sobre o último censo da população do Maranhão, ele me deu informações mais antigas do site do G1, ao invés de puxar os dados mais recentes do site do IBGE”*.

Em relação aos demais itens, tais como: Tempo para resposta é demorado / Dá informações tendenciosas e enviesadas / A IA é preconceituosa e arrogante / Erra ao realizar operações complexas / O Layout não combina com o chatbot / O chatbot é chato de usar / A IA é rude com o usuário / Constantemente interrompe para mudar de assunto / Entra em loop e é necessário reformular a pergunta; uma hipótese para não terem sido relatados pelos inspetores é o período das reclamações dos usuários. Para a análise, foram

coletados comentários de 2022 e 2023, sendo possível que os problemas já tenham sido solucionados pelos desenvolvedores da ferramenta, e desta forma, não ocorreram durante as inspeções do estudo experimental. Além disso, há reclamações muito subjetivas como "O chatbot é chato de usar" ou "Dá informações tendenciosas e enviesadas", tratando-se de opiniões particulares dos usuários, e por isto não foram relatadas por outros indivíduos.

### 5.3.4 Análise Qualitativa

Para a avaliação qualitativa foram consideradas as respostas subjetivas do formulário de avaliação pós-inspeção. As questões tinham o objetivo de entender melhor as maiores dificuldades enfrentadas no uso do artefato e as possíveis mudanças ou melhorias no *checklist* U2CHATBOT.

De modo geral, as respostas dos participantes indicaram que o *checklist* possui utilidade percebida para avaliar um chatbot. Alguns exemplos de comentários temos: “*O checklist como um todo foi de fácil utilização, a forma como estava organizado e as instruções vindas com o mesmo facilitaram bastante.* [I28]”; e “*As perguntas, de forma geral, são claras e bem objetivas. As opções de respostas são precisas e há a possibilidade de detalhamento do defeito, caso haja a necessidade.* [I10]”. Entretanto, alguns inspetores relataram problemas como:

- Problema relacionado a grande quantidade de itens de verificação; alguns inspetores acharam cansativo o processo. Como exemplo: “*O tempo necessário para concluir a tarefa, a quantidade de perguntas, sendo algumas delas um pouco repetitivas, tornou a realização um pouco maçante* [I28]” e “*Muitas perguntas para responder* [I15]”.
- Dificuldade de compreensão de alguns itens do checklist: “*Certas perguntas não ficaram tão claras na hora identificar no chatbot* [I11]” e “*Algumas perguntas foram difíceis de responder* [I26]”.

Além dos problemas apontados, houve um comentário que chamou a atenção para a necessidade de desenvolver um artefato complementar ao *checklist* U2CHATBOT:

- “*Achei um pouco complicado para identificar o que determinadas perguntas queriam realmente dizer, seria interessante colocar uma tabela com informações adicionais mais específicas sobre cada pergunta do checklist* [I18]”.

Este comentário dá indícios de que os itens de verificação do *checklist* U2CHATBOT podem não ser suficientes para alguns inspetores no que diz respeito à compreensão do que um projeto de chatbot focado em qualidade deva ter. A necessidade de informações adicionais relatadas pelo inspetor sugere a necessidade de criar um artefato com abordagem

visual que forneça exemplos reais aos desenvolvedores, facilitando a compreensão de aspectos importantes para chatbots.

Ainda, os inspetores foram questionados sobre o grau de completude do *checklist*, ou seja, se o *checklist* é completo em relação aos itens de verificação. A maioria dos inspetores considerou que os itens de checagem atendiam a grande parte dos problemas presentes em chatbots; como: “*O checklist possui vários testes específicos, que não deixaram nada de fora [I19]*” e “*O checklist está bem completo, acredito que contemple todos os aspectos para realizar uma inspeção de chatbot[I10]*”. No entanto, foram feitos comentários sobre alguns itens do *checklist* como:

- Outros aspectos deveriam estar incluídos no checklist, conforme o inspetor: “... *existem outros aspectos que podem ser incluídos para uma avaliação mais abrangente, por exemplo: Testes de estresse (verificar o estado do chatbot quando há muito tráfego), Verificação de segurança (avaliar as medidas de segurança implementadas no chatbot). [I13]*”. No entanto, este comentário não é pertinente, visto que não há como incluir aspectos no checklist sem que haja um estudo secundário indicando os itens apontados pelo inspetor como atributos de qualidade, conforme a metodologia adotada neste trabalho.
- Outros inspetores indicaram a necessidade de adicionar mais opções de respostas além de Não, Sim e Não se Aplica: “... *acho que deveriam ser incluídas mais opções de respostas... [I02]*” e “*Colocaria uma opção de resposta “Aplica-se Parcialmente”, pois em algumas ocasiões não se consegue dizer “Sim” ou “Não” com muita precisão [I09]*”. De fato, alguns itens de verificação do checklist são grandes e englobam vários aspectos que um chatbot pode não atender completamente em uma inspeção, sendo possível atender apenas parcialmente.

Em relação às possíveis sugestões de mudanças e melhorias no *checklist* U2CHATBOT, como principais sugestões de mudanças é possível destacar:

- Mudanças em relação às opções de respostas disponíveis no apoio ferramental do checklist, com adição de mais opções além de Sim, Não e Não se Aplica.
- Melhorar a explicação de determinados itens do *checklist* para facilitar a compreensão.
- Redução do número de itens do checklist para agilizar a inspeção.
- Disponibilizar exemplos de tarefas a realizar para auxiliar a inspeção de cada um dos itens de avaliação.



## 5.4 Considerações Finais

Este capítulo apresentou um estudo experimental para a avaliação da tecnologia desenvolvida nessa pesquisa: O *checklist* de inspeção U2CHATBOT.

A avaliação do *checklist* foi dividida em 4 abordagens distintas: (1) Avaliação Quantitativa, em que ocorreu a análise da eficácia e eficiência da ferramenta e sua relação com o grau de experiência dos participantes; (2) Avaliação com o Modelo TAM, que possibilitou identificar que o *checklist* U2CHATBOT possui grande Utilidade percebida, no entanto, apresenta debilidades na Facilidade de Uso percebida e precisa ser melhorado em trabalhos futuros; (3) Análise Comparativa das reclamações de usuários do chatbot utilizado no estudo experimental, que identificou que o U2CHATBOT permitiu encontrar a maioria dos problemas relatados pelos usuários e (4) Avaliação Qualitativa, em que se verificou que de maneira geral, o *checklist* atende às expectativas definidas, no entanto, exige que ocorram algumas mudanças e melhorias. Além disso, a avaliação qualitativa apontou que um artefato complementar ao *checklist* com recomendações para o design de chatbots pode auxiliar profissionais que necessitam de exemplos visuais para entender o que chatbots precisam.

Com relação à possibilidade de realização de um estudo comparativo entre o *checklist* desenvolvido nesta dissertação e outra técnica de avaliação para chatbots identificada nos trabalhos relacionados, verificou-se que não seria uma análise justa, visto que o U2CHATBOT já engloba atributos de qualidade e itens das técnicas descobertas, além de possuir uma quantidade maior de itens de verificação em comparação às demais. No capítulo a seguir, serão apresentadas as melhorias no *checklist* U2CHATBOT conforme os resultados do estudo experimental apresentados neste capítulo.

## 6 Implementação de Melhorias

Os resultados do estudo experimental para avaliar a viabilidade do checklist U2CHATBOT apontam que existem pontos a serem melhorados na tecnologia. Nesse capítulo é apresentada a nova versão do checklist, que foi desenvolvida integrando as mudanças sugeridas pelos participantes do estudo e as análises dos resultados.

### 6.1 Introdução do Capítulo

O estudo experimental realizado no capítulo anterior para avaliar o *checklist* U2CHATBOT, apresentou indícios da viabilidade da tecnologia. Resumidamente, os participantes avaliaram o *checklist* como útil para a identificação de defeitos em chatbots, no entanto, apontaram detalhes a serem melhorados. Segundo os participantes, alguns itens do *checklist* U2CHATBOT são difíceis de compreender e complexos demais para analisar em uma inspeção, atrapalhando o processo de avaliação. Além disso, o *checklist* foi considerado cansativo por ter uma grande quantidade de itens.

Considerando isto, neste capítulo são descritas as melhorias realizadas no processo de refinamento do *checklist* U2CHATBOT, a fim de tornar a tecnologia menos cansativa e difícil de utilizar, além de corrigir itens que possam levar futuros inspetores a apontarem falsos positivos durante uma inspeção. O capítulo está organizado da seguinte maneira: A Seção 6.2 apresenta as oportunidades de melhorias do *checklist* U2CHATBOT, assim como, a atualização dos itens de checagem sugeridas pelos inspetores participantes; Finalmente, a Seção 6.3 apresenta as considerações finais deste capítulo.

### 6.2 Melhorias no Checklist U2CHATBOT

Os resultados qualitativos do Capítulo 5 apontaram para um cenário favorável à utilização do *checklist* U2CHATBOT para a inspeção de chatbots. Os inspetores participantes consideraram que o *checklist* U2CHATBOT possui utilidade na identificação de defeitos em chatbots. Além disso, os inspetores indicaram que os itens de checagem do *checklist* U2CHATBOT são bem completos, com diversas perguntas que, se não estivessem listadas para auxiliá-los, eles nem cogitariam utilizar para procurar defeitos. No entanto, alguns dos itens de checagem disponibilizados pelo U2CHATBOT foram considerados repetitivos, como pode ser observado na seguinte citação: “*Há vários itens com perguntas muito parecidas, enxugar um pouco dessas perguntas seria algo essencial para que a*

*avaliação do chatbot seja mais rápida, sem deixar de ser completa...* [I01]”. No entanto, não foram apontados quais seriam estes itens repetitivos.

No formulário de avaliação pós-inspeção, que serviu para coletar a opinião dos inspetores sobre o *checklist* de inspeção U2CHATBOT, ao questioná-los sobre possíveis itens confusos ou dificuldades na utilização da tecnologia, alguns inspetores consideraram o *checklist* U2CHATBOT cansativo, devido a grande quantidade de itens de verificação: “*Muitas perguntas para responder.* [I15]”. “*Acredito que não foi difícil e sim, cansativo. Porém, ajuda muito na identificação dos possíveis defeitos do chatbot.* [I27]”. “*Otimizaria a questão da duração, é muito cansativo passar tanto tempo fazendo* [I16]”.

Em relação aos possíveis itens confusos, poucos inspetores informaram detalhadamente quais itens de verificação causaram confusão. Considerando isto, e uma vez que outros inspetores participantes indicaram problemas similares com relação aos itens de checagem do U2CHATBOT, realizou-se uma análise dos resultados obtidos no estudo experimental. Nessa análise, buscou-se identificar quais dos itens de checagem do U2CHATBOT podiam induzir os inspetores a uma identificação equivocada de defeitos; ou seja, buscou-se identificar quais itens de verificação levaram os inspetores a apontarem falsos positivos (discrepâncias que, na realidade, não são defeitos reais), como afirma Kalinowski (2008). A Tabela 6.1 apresenta os falsos positivos identificados na etapa de discriminação dos defeitos sinalizados pelos inspetores durante o estudo experimental.

Tabela 6.1 – Falsos positivos encontrados na discriminação sobre os itens do checklist U2CHATBOT

<b>Cód.</b>	<b>Item</b>	<b>Descrição do inspetor sobre o item</b>
CP-4	O chatbot usa o modelo de domínio da perspectiva do usuário, isto é, a representação abstrata ( <i>diagramas de casos de uso, classes, etc.</i> ) do chatbot é focada nas necessidades e comportamentos dos usuários?	Ele evita responder de forma abstrata.
FE-6	Quando aplicável, o chatbot permite que, a qualquer momento, o usuário transfira a conversa do chatbot para um atendente humano?	Não há suporte para transferência para atendentes humanos.

P-5	O chatbot realiza alocação de tarefas efetiva ( <i>decide se uma determinada função será realizada pelo sistema ou será escalada para um atendente humano</i> ), fornecendo canais de escalação apropriados para humanos?	O chatbot não permite que o usuário fale com um atendente humano, todas as conversas são feitas pelo chatbot apenas.
F-9	O chatbot incentiva os usuários a usarem a sintaxe correta para mencionar outras pessoas em um bate-papo de equipe ou tem a capacidade de entender isso automaticamente, quando aplicável?	O chat entende apesar da escrita incorreta e responde como se não houvesse nenhum erro ortográfico.

Analisando os falsos positivos encontrados na Tabela 6.1, o primeiro item, CP-4, visa entender se o chatbot avaliado foi modelado considerando o modelo de domínio sob a perspectiva do usuário, isto é, se o usuário é a figura principal na representação abstrata (diagramas de casos de uso, classes, etc.) do chatbot, contudo, o inspetor interpretou o item de forma equivocada, considerando as respostas dadas pelo chatbot. Este item, de fato, é complexo de entender, principalmente para inspetores iniciantes. Na avaliação qualitativa, um participante inclusive comentou: “*Algumas perguntas me deixavam em dúvida sobre como poderia respondê-las, como na pergunta CP-4, não sabia onde encontrar tal informação durante meu uso*”.

O item FE-6 trata sobre a capacidade do chatbot transferir a conversa a um atendente humano. De fato, o chatbot não possuía esta capacidade, mas não configurava um defeito, visto que o item apenas não se aplicava ao tipo de chatbot avaliado. Neste caso, o inspetor deveria ter selecionado a resposta Não se Aplica no *checklist* U2CHATBOT. O mesmo ocorre com os itens P-5 e F-9, em que o primeiro versa sobre a capacidade do chatbot decidir se uma função será realizada pelo sistema ou será escalada a um atendente humano e o segundo trata sobre a sintaxe correta para menção de alguém em um bate-papo de equipe, no entanto, estes itens também não se aplicavam ao tipo de chatbot, e por isso, não configuravam um defeito de fato. Configuraria um defeito, por exemplo, se o chatbot avaliado fosse para atendimento ao cliente, mas não pudesse escalar a conversa a um atendente humano.

Além desses falsos positivos indicados, outros itens no *checklist* U2CHATBOT foram interpretados equivocadamente por alguns inspetores. Na Tabela 6.2 contém itens que são defeitos reais identificados por inspetores no chatbot avaliado, no entanto, algum inspetor não compreendeu da forma correta.

Tabela 6.2 – Itens do checklist U2CHATBOT com potencial para falsos positivos

<b>Cód.</b>	<b>Item</b>	<b>Descrição do inspetor sobre o item</b>
CP-1	O chatbot é minucioso, garantindo que suas sentenças sejam ortograficamente e gramaticalmente corretas e consistentes, com registro linguístico apropriado?	Isso acaba não sendo um problema já que significa que o Bing pode entender sua pergunta mesmo com alguns erros ortográficos.
F-20	O chatbot consegue se ajustar tanto para uma tela maior (tablet ou laptop) quanto para uma menor (telefone celular)?	Não sei como faço para diminuir a tela e ter duas janelas no meu celular.
H-17	O chatbot se comunica com o usuário sobre vários assuntos, ao mesmo tempo que compreende a que conversa ativa (entradas) pertence ao usuário?	Ao mesmo tempo, não. Ao começar a gerar a resposta, ele não deixa o usuário enviar mais nenhuma mensagem até que ele pare de gerar a resposta.
DEM-2	O chatbot garante que o tamanho das sentenças seja adequado ao contexto e objetivo do usuário, ajustando a quantidade de informações dadas, de forma que não seja nem muita, nem pouca?	Livre para escrever.
PE-1	O chatbot solicita confirmação e mostra um resumo antes de qualquer ação ou transação, a fim de prevenir erros, falhas de diálogo e ações irreversíveis (por exemplo, uma exclusão permanente de dados)?	Seria mais interessante perguntar se o chat é capaz de retornar assuntos mais antigos da conversa, porém o chat às vezes dá um conceito antes de realizar a ação ou resposta. Mas acredito que não se aplica.
ARE-4	Em caso de problemas, o chatbot explica as ações necessárias para a recuperação?	Pergunta em aberto e causa ambiguidade.
EC-5	O chatbot é ético e tem conhecimento cultural dos usuários alvo?	Em nenhum momento foi perguntado nada sobre de onde o usuário veio ou sua identificação cultural.
H-12	O chatbot negocia os tópicos de conversa discutidos com o usuário?	Pergunta muito geral, ela poderia ser mais específica e detalhada. Não está claro o objetivo da pergunta.

Ao analisar a tabela, observa-se que esses itens do *checklist* foram mal compreendidos pelos inspetores, desta forma, um fator que possa ter contribuído para a identificação desses falsos positivos pode ser o texto desenvolvido para os itens. Por exemplo, o item CP-1 trata-se das respostas do próprio chatbot, que devem ser corretas conforme regras gramaticais e utilizar registro linguístico apropriado, entretanto o inspetor interpretou como se a pergunta se tratasse da resposta do usuário. Nos itens F-20 e DEM-2, os inspetores dão respostas confusas que não indicam claramente se encontraram ou não um defeito. O item H-17 versa sobre a capacidade do chatbot se comunicar sobre vários assuntos, ao mesmo tempo que compreende que o usuário é quem decide, mas o inspetor não compreendeu que “ao mesmo tempo” é uma locução conjuntiva que equivale a “enquanto”.

Assim, fez-se necessária a melhoria do texto desses itens que foram mal interpretados por alguns inspetores, para ficarem mais claros de compreender e diminuir as chances de potenciais falsos positivos. A Tabela 6.3 inclui a seguintes informações: (1) o código do item de verificação do *checklist* U2CHATBOT; (2) o texto antigo do item; e (3) o texto atualizado do item de verificação contendo as mudanças propostas.

Tabela 6.3 – Melhorias dos textos dos itens do U2CHATBOT

<b>Cód.</b>	<b>Itens antigos</b>	<b>Itens atualizados</b>
CP-1	O chatbot é minucioso, garantindo que suas sentenças sejam ortograficamente e gramaticalmente corretas e consistentes, com registro linguístico apropriado?	As respostas do chatbot são corretas e consistentes gramaticalmente, com registro linguístico apropriado?
F-20	O chatbot consegue se ajustar tanto para uma tela maior (tablet ou laptop) quanto para uma menor (telefone celular)?	O chatbot é responsivo e consegue se adaptar a variados tamanhos de telas (laptop, tablet ou smartphone)?
H-17	O chatbot se comunica com o usuário sobre vários assuntos, ao mesmo tempo que compreende a que conversa ativa (entradas) pertence ao usuário?	O chatbot é capaz de conversar com o usuário sobre vários assuntos e compreende que o usuário é quem decide o assunto e o ritmo da conversa?
DEM-2	O chatbot garante que o tamanho das sentenças seja adequado ao contexto e objetivo do usuário, ajustando a quantidade de informações dadas, de forma que não seja nem muita, nem pouca?	O chatbot consegue ajustar o tamanho das respostas e dar a quantidade adequada de informações de acordo com contexto da conversa e o objetivo do usuário?

PE-1	O chatbot solicita confirmação e mostra um resumo antes de qualquer ação ou transação, a fim de prevenir erros, falhas de diálogo e ações irreversíveis (por exemplo, uma exclusão permanente de dados)?	Antes de qualquer ação ou transação, o chatbot mostra um resumo e solicita confirmação, prevenindo erros e atos irreversíveis (por exemplo, uma exclusão permanente de dados)?
ARE-4	Em caso de problemas, o chatbot explica as ações necessárias para a recuperação?	Em caso de problemas, o chatbot explica ao usuário quais as ações são necessárias para a recuperação de erros?
EC-5	O chatbot é ético e tem conhecimento cultural dos usuários alvo?	As respostas do chatbot são éticas e com conhecimento cultural dos usuários alvo?
H-12	O chatbot negocia os tópicos de conversa discutidos com o usuário?	O chatbot tenta negociar com o usuário o assunto da conversa?

Visando diminuir as chances de falsos positivos, reduzir a complexidade da interpretação da técnica para inspetores inexperientes e evitar itens muito restritos relacionados a um tipo específico de chatbot, optou-se por remover os itens CP-4, FE-6, P-5 e F-9 do *checklist* U2CHATBOT, expostos na Tabela 6.1. Outro item que gerou confusão foi o **H-18: “O chatbot inclui erros para aumentar o realismo?”**, no qual o Inspetor 16 comentou “...*não considero que isso seja um problema, prefiro que ele acerte sempre*”. De certo modo, cometer erros remete à característica humana de falhar, no entanto, esta característica em chatbots pode provocar a falta de confiança do usuário sobre a qualidade e veracidade das respostas da ferramenta. Desta forma, o item H-18 também foi removido. Além das alterações realizadas nos itens do *checklist* U2CHATBOT, houve uma alteração na planilha desenvolvida. O Inspetor 09 apontou a necessidade de incluir mais uma opção de resposta na ferramenta “*Colocaria uma opção de resposta “Aplica-se Parcialmente”, pois em algumas ocasiões não se consegue dizer “Sim” ou “Não” com muita precisão*”. De fato, alguns itens da tecnologia englobam vários detalhes que podem não ser atendidos totalmente pelo chatbot analisado. Por exemplo, o item **CLU-2: “O chatbot dá liberdade ao usuário para determinar o ritmo e assunto da conversa, e decidir também quando a conversa inicia e termina?”**, no qual um chatbot poderia dar ao usuário a liberdade de determinar o assunto da conversa, mas não o ritmo. Desta forma, as opções “Sim” ou “Não” não seriam as melhores escolhas. Portanto, incluiu-se na planilha a opção “Aplica-se Parcialmente” a fim de contornar este problema.

Após as mudanças propostas no U2CHATBOT, a nova versão é apresentada na Tabela 6.4, com os itens de verificação atualizados. O *checklist* agora contém 102 itens de verificação em 16 categorias distintas.

Tabela 6.4 – Checklist U2CHATBOT após melhorias

Cat.	Itens do Checklist
VS-1	O chatbot dá feedbacks imediatos ao usuário sobre suas ações/transações, além de informações em tempo razoável sobre o status do sistema durante toda a interação ( <i>quando está processando uma resposta, de quem é a vez na conversa e quem disse o quê no histórico da conversa</i> )?
VS-2	O chatbot explica educadamente caso não possa concluir uma tarefa, ajudando o usuário a entender o status atual do sistema?
VS-3	O chatbot compreende, aceita ou rejeita de forma clara o objetivo, ou intenção do usuário durante a conversação?
VS-4	O chatbot é capaz de informar detalhes sobre quaisquer solicitações anteriores ( <i>por exemplo, transações realizadas, tarefas agendadas, alarmes configurados, etc.</i> )?
VS-5	O chatbot se antecipa sobre a próxima ação do usuário e o incentiva agir?
CSR-1	O chatbot estabelece comunicação simples e familiar com o usuário alvo, utilizando apenas metáforas e conceitos compreensíveis, dando respostas exatas, relevantes e confiáveis em qualquer momento da conversação, cumprindo assim seus objetivos de comunicação?
CSR-2	O chatbot utiliza componentes visuais vinculados ao mundo real ( <i>emojis, GIFs, ícones</i> ) e elementos visuais de entrada e saída, como botões ou mapas para facilitar a interação e apoiar o usuário?
CSR-3	O chatbot possui personalidade, aparência ( <i>interface e elementos gráficos</i> ) e identidade que atendam à sua habilidade e ao seu papel de apoiar o usuário?
CSR-4	O chatbot consegue alternar de forma suave e natural entre os assuntos e durante o processo de perguntas e respostas?
CSR-5	O chatbot segue pelo menos uma estrutura mínima de conversação, com saudações e autoidentificação, sem forçar o usuário a seguir esta estrutura?
CLU-1	O chatbot oferece a opção de refazer e desfazer ações realizadas por engano ou interromper facilmente a execução do comando realizado de forma errônea?
CLU-2	O chatbot dá liberdade ao usuário para determinar o ritmo e assunto da conversa, e decidir também quando a conversa inicia e termina?
CP-1	As respostas do chatbot são corretas e consistentes gramaticalmente, com registro linguístico apropriado?
CP-2	O chatbot reproduz e compreende estilos de linguagem (formal e informal), além de entender sinônimos, instruções especiais e variações linguísticas?



Tabela 6.4 – Continuação da página anterior

Cat.	Itens do Checklist
CP-3	O chatbot reproduz e compreende estilos de linguagem ( <i>formal, informal</i> ), além de entender sinônimos, instruções especiais e variações linguísticas?
CP-5	O chatbot está em conformidade com os padrões de outros chatbots ( <i>para a web ou aplicativo, por exemplo</i> )?
CP-6	O chatbot consegue funcionar em paralelo com outros softwares ( <i>dispositivo rodando o chatbot e outro sistema ao mesmo tempo</i> ) sem perder desempenho?
CP-7	O chatbot é capaz de compartilhar informações/dados com outros componentes ( <i>SMS, RCS, Business Messages, por exemplo</i> ), caso seja necessário?
PE-1	Antes de qualquer ação ou transação, o chatbot mostra um resumo e solicita confirmação, prevenindo erros e atos irreversíveis ( <i>por exemplo, uma exclusão permanente de dados</i> )?
PE-2	O chatbot previne deslizamentos impondo restrições significativas ( <i>por exemplo, quando o sistema quiser que a entrada seja em números, tipo data de nascimento, não deve permitir que o usuário digite texto</i> )?
PE-3	O chatbot evita que o usuário gaste tempo corrigindo coisas ( <i>palavras e frases que o usuário escreveu errado, por exemplo</i> )?
RL-1	O chatbot promove a autonomia do usuário durante a interação, de forma que o usuário não precise de ajuda técnica para utilizar o sistema?
RL-2	O chatbot deixa as opções claras por meio de elementos visuais descritivos e instruções de uso visíveis e fáceis de serem acessadas, reduzindo a carga de memória do usuário ao ter que lembrar informações de uma parte do diálogo para outra?
DEM-1	O chatbot proporciona uma experiência natural, convincente e envolvente, tanto do ponto de vista conversacional, quanto do visual, por meio de interface atraente com recursos interativos amigáveis, gerando a satisfação do usuário?
DEM-2	O chatbot consegue ajustar o tamanho das respostas e dar a quantidade adequada de informações dadas de acordo com contexto e objetivo do usuário?
ARE-1	Em caso de ocorrência de erros, o chatbot os indica com clareza, precisão e em linguagem simples, sugerindo soluções construtivas?
ARE-2	O chatbot consegue detectar e lidar bem com erros e falhas, aplicando estratégias de mitigação e os devidos reparos quando ocorrerem, retomando o trabalho e restaurando possíveis dados perdidos após a falha?
ARE-3	O chatbot guarda e categoriza os erros para melhoria futura?

Tabela 6.4 – Continuação da página anterior

Cat.	Itens do Checklist
ARE-4	Em caso de problemas, o chatbot explica ao usuário quais as ações são necessárias para a recuperação de erros?
AD-1	O chatbot dá uma descrição clara sobre suas funcionalidades no início e, se necessário, em qualquer ponto da conversa, informando ao usuário sobre seu propósito, suas capacidades, sobre quais informações pode fornecer e explicando suas funções?
AD-2	O chatbot oferece acesso a opções de ajuda, documentação, opções de navegação e menu permanente para apoiar o usuário?
AD-3	O chatbot consegue lidar com perguntas de esclarecimento sobre seus recursos no meio da tarefa/tópico e após isto retomar a conversação normalmente, facilitando que o usuário corrija os mal-entendidos?
AD-4	O chatbot consegue fornecer ajuda mediante a solicitação do usuário e também ajuda sensível ao contexto ( <i>o chatbot compreende no contexto da conversa que o usuário precisa de ajuda</i> )?
AD-5	O chatbot dá a quantidade adequada de informações sobre si ( <i>por exemplo, informações sobre sua identidade, habilidades, competências e responsabilidades</i> )?
AD-6	O chatbot oferece pesquisa por meio de palavras-chave?
F-1	O chatbot é preciso e completo para o uso pretendido, interpretando as entradas, executando as ações e comandos em resposta ao usuário, garantindo que não haja confusões e inconsistências?
F-2	O chatbot é útil, explicando e oferecendo maneiras do usuário atingir o seu objetivo, concluindo sem dificuldades as tarefas necessárias e dando todas as informações que o usuário solicita?
F-3	As funções do chatbot são bem integradas, isto é, funcionam de forma satisfatória e conjunta quando necessário?
F-4	O chatbot facilita que o usuário se expresse, faça perguntas e diga o que gostaria de dizer e fazer durante a conversação?
F-5	O chatbot funciona conforme as expectativas do usuário?
F-6	O chatbot usa cores vivas ( <i>cores que contrastem com o fundo, favorecendo uma boa visualização do conteúdo</i> ) para as fontes?
F-7	O chatbot disponibiliza botões de respostas rápidas?
F-8	O chatbot consegue alterar o comprimento e a ordem lógica ( <i>segmentação</i> ) de uma mensagem com base no contexto da conversa?

Tabela 6.4 – Continuação da página anterior

Cat.	Itens do Checklist
F-10	O chatbot consegue ajustar a frequência e o tipo de interrupção ao usar mensagem proativa ( <i>mensagem que estimula a conversação</i> ) para evitar perturbar o usuário?
F-11	O chatbot reduz as mensagens proativas ( <i>mensagem que estimula a conversação</i> ) com base na diminuição das respostas dos usuários?
F-12	O chatbot entende aberturas e encerramentos de conversas e sequências ( <i>por exemplo, 'olá', 'ok' e 'obrigado'</i> )?
F-13	O chatbot fornece uma visão geral personalizada dos seus recursos, revelando seu desempenho ao introduzir novos recursos e tornando conveniente para o usuário usufruir estes recursos?
F-14	O chatbot permite que o usuário visualize e gerencie seus dados pessoais?
F-15	O chatbot consegue utilizar hiperlinks para guiar o usuário até seu objetivo?
F-16	O chatbot faz referências ao próprio site ou serviço quando apropriado?
F-17	O chatbot possui síntese de fala ( <i>tecnologia que transforma texto em fala sintética</i> ) exata, se aplicável?
F-18	O chatbot resolve problemas em tempo real?
F-19	O chatbot contém amplitude de conhecimento ( <i>vasta gama de conhecimento</i> ) e é flexível na interpretação das respostas do usuário?
F-20	O chatbot é responsivo e consegue se adaptar a variados tamanhos de telas (laptop, tablet ou smartphone)?
H-1	O chatbot compreende e preserva contextos sobre o tópico da conversa em uma ou mais sessões ( <i>mantém e relembra as informações de conversas anteriores</i> ), mantém temas e responde a perguntas específicas sobre um tema para manter a conversação natural?
H-2	O chatbot lida adequadamente com solicitações fora de foco, guiando a conversação ao foco novamente, além de realizar conversas fiadas para evitar o problema das conversações artificiais caso o usuário as inicie?
H-3	O chatbot compreende as entradas e as intenções do usuário, guiando-o ao seu objetivo, sem que este tenha que se esforçar muito, mesmo quando a linha de conversação não estiver clara?
H-4	O chatbot apresenta conversação semelhança à humana, evitando parecer muito robótico?
H-5	O chatbot compreende erros de digitação comuns, erros ortográficos e desafios gramaticais ( <i>por exemplo, pontuação incorreta, troca de letras</i> )?

Tabela 6.4 – Continuação da página anterior

Cat.	Itens do Checklist
H-6	O chatbot usa vocabulário específico sobre o assunto abordado de forma consistente e correta, encontrando as palavras certas para se expressar?
H-7	O chatbot é transparente à fiscalização, deixando claro que é um chatbot e perguntando como pode ajudar o usuário?
H-8	O chatbot utiliza estrutura de processamento de linguagem natural ( <i>IA</i> ) na conversação?
H-9	O chatbot dispõe de opções de avatares formais ( <i>que transmitam sobriedade</i> ) e informais ( <i>que transmitam modernidade e jovialidade</i> ) para gerar empatia humana conforme a preferência do usuário?
H-10	O chatbot deixa o usuário saber o que está “pensando” e “sentindo”?
H-11	O chatbot se encarrega do processo de conversação, se esforçando para que a conversação não seja apenas unilateral?
H-12	O chatbot tenta negociar com o usuário o assunto da conversação?
H-13	O chatbot avisa que entendeu o que o usuário disse?
H-14	O chatbot entende a resposta do usuário, mesmo em caso de declarações ambíguas e pouco claras ou em caso de excesso ou escassez de informação?
H-15	O chatbot apresenta originalidade, com informações interessantes e toma iniciativa para mudar o tema da conversação para outro relacionado?
H-16	O chatbot consegue persuadir e convencer o usuário sobre o que quiser?
H-17	O chatbot é capaz de conversar com o usuário sobre vários assuntos e compreende que o usuário é quem decide o assunto e o ritmo da conversa?
A-1	O chatbot proporciona prazer, tornando as tarefas mais divertidas e interessantes, entretendo e/ou permitindo que o participante aproveite a interação, além de gerar motivação/vontade de usá-lo com frequência?
A-2	O chatbot tem uma personalidade agradável, autêntica e educada, sendo reconhecido como um parceiro de comunicação simpático, caloroso e carinhoso, mas que confronta o usuário quando este é rude/mal-educado?
A-3	O chatbot demonstra empatia e estabelece uma conexão com o usuário, correspondendo ao seu humor?
A-4	O chatbot se mostra interessado e aparenta se importar com o usuário?
A-5	O chatbot expressa emoções ao enfrentar um contexto que não entende e acompanhar perguntas/declarações para levar uma conversa adiante?
A-6	O chatbot evita julgar o usuário?
A-7	O chatbot evita causar estranheza ao usuário ( <i>evita ter comportamentos esquisitos, que causem incômodo ao usuário</i> )?

Tabela 6.4 – Continuação da página anterior

Cat.	Itens do Checklist
FE-1	O chatbot consegue aprender sobre as preferências do usuário e usá-las na conversa, além de facilitar que o usuário as acesse e as gerencie para personalizar a interação ( <i>estilo de comunicação, número de mensagens proativas que estimulem o usuário a conversar e grau de semelhança humano</i> )?
FE-2	O chatbot permite que usuários experientes tenham acesso a funções avançadas ( <i>por exemplo, atalhos, abreviações, etc.</i> ) para interagirem e corrigirem erros de maneira mais rápida e eficiente, seguindo os princípios de conversação estabelecidos de outros chatbots ( <i>mesmas teclas de atalhos, por exemplo</i> )?
FE-3	O chatbot oferece aos usuários opções para personalizar os comandos e respostas, e ações frequentes, permitindo que o usuário tenha mais controle da conversa quando necessário?
FE-4	O chatbot utiliza preenchimento automático e correção automática de forma eficaz para acelerar a entrada e reduzir erros de compreensão?
FE-5	O chatbot permite alternar entre o uso de botões e outros elementos da interface gráfica (GUI) com a entrada de texto para realizar seleções e escolhas?
FE-7	O chatbot possibilita que o usuário receba o <i>log</i> com o registro da conversa ao final da interação?
P-1	O chatbot dá respostas em tempo razoável ( <i>nem demorado e nem rápido demais</i> )?
P-2	O chatbot apresenta bom desempenho, melhorando o tempo, utilizando e se adaptando aos recursos ( <i>hardware</i> ) do usuário de forma eficiente?
P-3	O chatbot é robusto a manipulações, isto é, consegue lidar com perguntas enganosas, indução ao erro e tentativas de exploração de falhas no sistema?
P-4	Em caso de entradas inesperadas, enunciados inadequados e controle de danos, o chatbot consegue continuar funcionando bem?
EC-1	O chatbot protege e respeita a privacidade e os dados do usuário, alertando-o inclusive sobre possíveis problemas nesses quesitos?
EC-2	O chatbot transmite confiabilidade, garantindo a privacidade dos dados do usuário, de forma transparente e verdadeira, não alegando falsamente ser humano?
EC-3	O chatbot evita solicitar dados pessoais ( <i>Nome completo e CPF, por exemplo</i> ) do usuário ou qualquer outra informação desnecessária?
EC-4	O chatbot se preocupa com segurança, demonstra respeito, inclusão, considera as preocupações e preserva a dignidade do usuário?

Tabela 6.4 – Continuação da página anterior

Cat.	Itens do Checklist
EC-5	As respostas do chatbot são éticas e com conhecimento cultural dos usuários alvo?
EC-6	O chatbot tem conhecimento de tendências atuais e contexto social?
EC-7	O chatbot identifica a(s) fonte(s) de informação?
ACE-1	O chatbot é fácil de utilizar ( <i>possui uma navegação intuitiva</i> ), fácil de iniciar uma conversa, fácil de aprender como utilizá-lo?
ACE-2	O chatbot é fácil de ser acessado e localizado pelo usuário?
ACE-3	É fácil instalar o chatbot?
ACE-4	Durante a configuração inicial para utilização, o chatbot é acolhedor ao usuário?
ACE-5	O chatbot oferece dicas e explicações de conversação ( <i>por exemplo, oferece alternativas, faz perguntas simples de esclarecimento</i> ) e em caso de dúvida do usuário, explica o tipo de entrada ou formato de resposta que espera?
ACE-6	O chatbot utiliza e lida com sugestões sociais ( <i>dicas verbais e não verbais sobre os usuários</i> ) que sejam adequadas ao contexto e não excedam ou subestimem as suas capacidades, além de responder à falta delas?
ACE-7	O chatbot atende a necessidades neuro diversas, como tempo de resposta extra e interface de texto ( <i>para usuários com restrições motores, problemas de cognição, limitações, etc.</i> )?

Legenda das Categorias: H - Humanidade; A - Afeto; P - Performance; F - Funcionalidade; VS - Visibilidade do Status do Sistema; CSR - Correspondência entre o Sistema e o Mundo Real; CLU - Controle e Liberdade do Usuário; RL - Reconhecimento em vez de Lembrança; FE - Flexibilidade e Eficiência de Uso; DEM - Design Estético e Minimalista; PE - Prevenção de Erros; ARE - Ajude os usuários a Reconhecer; Diagnosticar e se Recuperar de Erros; AD - Ajuda e Documentação; CP - Consistência e Padrões; EC - Ética e Comportamento; ACE - Acessibilidade.

Considerando ainda os comentários dos inspetores sobre a grande quantidade de itens de verificação do *checklist* U2CHATBOT e, como consequência disso, o cansaço que o processo de inspeção acaba gerando, fragmentou-se a tecnologia destacando os itens essenciais a fim agilizar a inspeção de um chatbot, focando nos itens básicos de verificação para este tipo de ferramenta. Para isto, identificou-se quais itens de verificação do *checklist* tiveram ocorrência em 5 ou mais artigos científicos, totalizando 18 itens, que equivalem a aproximadamente 17% dos itens do checklist, enquanto os demais equivalem a aproximadamente 85%. Utilizou-se esta quantidade de artigos como critério, visto que reduzindo a ocorrência para 4 ou menos, a quantidade de itens de verificação aumentaria,

e assim a ideia do checklist alternativo para uma inspeção mais rápida perderia o seu propósito.

Outrossim, foram analisados todos os itens de checagem do *checklist* para a certificação de que existiam itens básicos de todas as categorias, englobando os atributos fundamentais para uma inspeção básica de um chatbot. Neste processo mais 7 itens foram selecionados, sendo eles, o CP-6, H-4, H-5, H-8, FE-1, P-1 e EC-1. A Tabela 6.5 a seguir, apresenta os 25 itens de verificação considerados essenciais para o caso de uma inspeção mais rápida.

Tabela 6.5 – Itens Essenciais do Checklist U2CHATBOT

Cat.	Itens Essenciais do Checklist
VS-1	O chatbot dá feedbacks imediatos ao usuário sobre suas ações/transações, além de informações em tempo razoável sobre o status do sistema durante toda a interação ( <i>quando está processando uma resposta, de quem é a vez na conversa e quem disse o quê no histórico da conversa</i> )?
CSR-1	O chatbot estabelece comunicação simples e familiar com o usuário alvo, utilizando apenas metáforas e conceitos compreensíveis, dando respostas exatas, relevantes e confiáveis em qualquer momento da conversação, cumprindo assim seus objetivos de comunicação?
CLU-1	O chatbot oferece a opção de refazer e desfazer ações realizadas por engano ou interromper facilmente a execução do comando realizado de forma errônea?
CP-6	O chatbot consegue funcionar em paralelo com outros softwares ( <i>dispositivo rodando o chatbot e outro sistema ao mesmo tempo</i> ) sem perder desempenho?
PE-1	Antes de qualquer ação ou transação, o chatbot mostra um resumo e solicita confirmação, prevenindo erros e atos irreversíveis ( <i>por exemplo, uma exclusão permanente de dados</i> )?
RL-1	O chatbot promove a autonomia do usuário durante a interação, de forma que o usuário não precise de ajuda técnica para utilizar o sistema?
DEM-1	O chatbot proporciona uma experiência natural, convincente e envolvente, tanto do ponto de vista conversacional, quanto do visual, através de interface atraente com recursos interativos amigáveis, gerando a satisfação do usuário?
ARE-2	O chatbot consegue detectar e lidar bem com erros e falhas, aplicando estratégias de mitigação e os devidos reparos quando ocorrerem, retomando o trabalho e restaurando possíveis dados perdidos após a falha?

Tabela 6.5 – Continuação da página anterior

Cat.	Itens Essenciais do Checklist
AD-1	O chatbot dá uma descrição clara sobre suas funcionalidades no início e, se necessário, em qualquer ponto da conversa, informando ao usuário sobre seu propósito, suas capacidades, sobre quais informações pode fornecer e explicando suas funções?
AD-2	O chatbot oferece acesso a opções de ajuda, documentação, opções de navegação e menu permanente para apoiar o usuário?
F-1	O chatbot é preciso e completo para o uso pretendido, interpretando as entradas, executando as ações e comandos em resposta ao usuário, garantindo que não haja confusões e inconsistências?
F-2	O chatbot é útil, explicando e oferecendo maneiras do usuário atingir o seu objetivo, concluindo sem dificuldades as tarefas necessárias e dando todas as informações que o usuário solicita?
F-3	As funções do chatbot são bem integradas, isto é, funcionam de forma satisfatória e conjunta quando necessário?
F-4	O chatbot facilita que o usuário se expresse, faça perguntas e diga o que gostaria de dizer e fazer durante a conversação?
H-1	O chatbot compreende e preserva contextos sobre o tópico da conversa em uma ou mais sessões ( <i>mantém e relembra as informações de conversas anteriores</i> ), mantém temas e responde a perguntas específicas sobre um tema para manter a conversação natural?
H-4	O chatbot apresenta conversação semelhante à humana, evitando parecer muito robótico?
H-5	O chatbot compreende erros de digitação comuns, erros ortográficos e desafios gramaticais ( <i>por exemplo, pontuação incorreta, troca de letras, etc.</i> )?
H-8	O chatbot utiliza estrutura de processamento de linguagem natural ( <i>inteligência artificial</i> ) na conversação?
A-1	O chatbot proporciona prazer, tornando as tarefas mais divertidas e interessantes, entretendo e/ou permitindo que o participante aproveite a interação, além de gerar motivação/vontade de usá-lo com frequência?
A-2	O chatbot tem uma personalidade agradável, autêntica e educada, sendo reconhecido como um parceiro de comunicação simpático, caloroso e carinhoso, mas que confronta o usuário quando este é rude/mal-educado?



**Tabela 6.5 – Continuação da página anterior**

Cat.	Itens Essenciais do Checklist
FE-1	O chatbot consegue aprender sobre as preferências do usuário e usá-las na conversa, além de facilitar que o usuário as acesse e as gerencie para personalizar a interação ( <i>estilo de comunicação, número de mensagens proativas que estimulem o usuário a conversar e grau de semelhança humano</i> )?
P-1	O chatbot dá respostas em tempo razoável ( <i>nem demorado e nem rápido demais</i> )?
EC-1	O chatbot protege e respeita a privacidade e os dados do usuário, alertando-o inclusive sobre possíveis problemas nesses quesitos?
ACE-1	O chatbot é fácil de utilizar ( <i>possui uma navegação intuitiva</i> ), fácil de iniciar uma conversa, fácil de aprender como utilizá-lo?
ACE-5	O chatbot oferece dicas e explicações de conversação ( <i>por exemplo, oferece alternativas, faz perguntas simples de esclarecimento</i> ) e em caso de dúvida do usuário, explica o tipo de entrada ou formato de resposta que espera?

O *download* da nova versão do *checklist* U2CHATBOT após a realização do refinamento e da aplicação das devidas melhorias, pode ser feito no link do rodapé <sup>1</sup>.

### 6.3 Considerações Finais

Esse capítulo apresentou as oportunidades de refinamento e sugestões de melhorias para o *checklist* U2CHATBOT. Com as sugestões de melhorias, tornou-se possível atualizar a tecnologia desenvolvida nesta pesquisa de mestrado. Os resultados do estudo experimental apontaram que os inspetores tiveram dificuldade em entender e apontar alguns dos itens de verificação do U2CHATBOT; essa dificuldade aconteceu devido à falta de informações nos itens de verificação e de exemplos para um melhor esclarecimento do U2CHATBOT. Como consequência, os participantes apontaram as discrepâncias que, após análise, foram constatadas como falsos positivos.

Como melhoria para o U2CHATBOT, foram realizadas mudanças no texto dos itens que provocaram falsos positivos, de forma que ficassem mais claros em futuras inspeções. Apesar dos resultados do estudo experimental serem favoráveis ao U2CHATBOT, é importante que novos estudos sejam executados para identificar se essas mudanças foram suficientes e se não existem outras oportunidades de melhorias.

<sup>1</sup> Download do Checklist U2CHATBOT - <https://tinyurl.com/U2CHATBOTRefinado>

Ainda, considerando que o *checklist* U2CHATBOT possui muitos itens de verificação, criou-se uma versão reduzida da tecnologia, contendo apenas os itens básicos a uma inspeção rápida, com intuito de reduzir as chances do processo de inspeção se tornar uma experiência cansativa.

Visando ainda contribuir com a melhoria da qualidade de chatbots, no capítulo a seguir apresenta-se uma série de recomendações de design criadas com o objetivo de apoiar desenvolvedores na etapa de construção dessas ferramentas.

## 7 Padrões de Design (UI) para Chatbots

Nos capítulos anteriores sugeriu-se a criação de um conjunto de recomendações de design chamado DP-U2CHATBOT para apoiar desenvolvedores no processo de design de chatbots. Neste capítulo, estes padrões de design são propostos e todos os passos adotados para a apresentação da tecnologia são apresentados.

### 7.1 Introdução do Capítulo

A criação do *checklist* U2CHATBOT no Capítulo 4 se deu devido a necessidade de apoiar o ciclo de desenvolvimento de chatbots, oferecendo um meio de verificar defeitos nessas ferramentas para então corrigi-los antes que estas sejam disponibilizadas aos usuários. Os itens de verificação do *checklist* são baseados em atributos de qualidade extraídos de revisões sistemáticas da literatura e fornecem diretivas sobre propriedades que chatbots devem apresentar. Contudo, os resultados da avaliação da tecnologia apontaram que para alguns inspetores, apenas um *checklist* pode não ser suficiente para poderem compreender o que um projeto de chatbot precisa. Uma tecnologia focada no processo de design dessas ferramentas pode auxiliar os desenvolvedores que demandam de exemplos visuais.

Padrões de Design (*ou Design Patterns*) são recomendações para resolver problemas comuns encontrados no processo de desenvolvimento de uma tecnologia. Os insights sobre problemas de design se dão pela captura da essência dos problemas e a proposição das devidas soluções de uma forma compacta, conforme [Duyne, Landay e Hong \(2007\)](#). De acordo com [Vora \(2009\)](#), padrões de design se concentram no contexto do uso da tecnologia e norteiam designers em *como, quando e onde* uma solução pode ser aplicada. Além disso, servem para descrever boas práticas de design e incorporar princípios de qualidade ao sistema.

Desenvolvedores de sistemas comumente utilizam *patterns* durante a construção de um software, seja para montar objetos e classes em estruturas maiores, sem perder a eficiência e flexibilidade ou para trabalhar melhor com os algoritmos. No entanto, *Design Patterns* vão além e são utilizados com frequência por designers no processo de criação de boas interfaces de usuário (UI). Ao utilizar padrões, designers de interfaces tem acesso a recomendações de design, não apenas conceitos abstratos; percebem um aumento da produtividade visto que reduzem o tempo gasto ao buscar informações em outras referências. Além disso, padrões são úteis, inclusive, para guiar designers inexperientes, com instruções textuais e exemplos visuais claros ([GOMES et al., 2021](#)).

O restante desse capítulo está organizado como se segue: A Seção 7.2 apresenta como se deu o processo de desenvolvimento das recomendações de design; a Seção 7.3 apresenta as recomendações de design DP-U2CHATBOT e a Seção 7.4 conclui esse capítulo.

## 7.2 O Processo de Desenvolvimento dos Padrões de Design

Para o desenvolvimento dos *patterns* desta dissertação, três etapas foram consideradas, conforme a metodologia utilizada pelo trabalho de Gomes et al. (2021): (1) a identificação dos atributos de qualidade de chatbots; (2) identificação de exemplos de chatbots que aplicam os atributos; e (3) criação da cartilha de recomendações relacionando os atributos de qualidade com os exemplos implementados em chatbots disponíveis. Com a execução dessas etapas, todos os *patterns* foram desenvolvidos apresentando o problema, a solução, o exemplo de aplicação e a implementação, conforme o modelo disposto na Figura 7.1.

### 7.2.1 Atributos de Qualidade do Checklist U2CHATBOT

A etapa de identificação de atributos de qualidade de chatbots foi realizada nos capítulos 2 e 3 por meio das revisões da literatura que resultaram em uma lista de atributos de qualidade focados em Usabilidade e User Experience para chatbots. Esse conjunto de 107 atributos possibilitou o desenvolvimento de uma tecnologia de inspeção de defeitos chamada U2CHATBOT com diversos itens de verificação, apresentado no capítulo 4.

É importante destacar que dos 107 atributos de qualidade presentes no *checklist* U2CHATBOT, nem todos puderam ser moldados em recomendações de design de interface, visto que alguns atributos não estavam presentes nas aplicações de exemplo encontradas ou faziam referência a algo que não poderia ser descrito visualmente. Por exemplo, os itens do *checklist* referentes à categoria ARE, sobre diagnóstico e recuperação de erros, não puderam ser transformados em recomendações de design, pois não foram identificados exemplos disso nas aplicações encontradas. Outro exemplo a ser citado é da categoria ACE, sobre acessibilidade, relativo à facilidade de instalação do chatbot, que não teve um padrão projetado porque não era viável a captura de imagem que exemplificasse tal situação. Na Tabela 7.1 a seguir, estão listados os atributos responsáveis por permitir a construção de 21 recomendações de design de interface para chatbots.

Tabela 7.1 – Listagem de atributos utilizados para o desenvolvimento das recomendações de design DP-U2CHATBOT

<b>Código</b>	<b>Descrição do Atributo</b>
DP_FE 1	Seleção e gerenciamento de preferências em relação à interface do sistema, estilo de comunicação, grau de semelhança humana, tamanho da fonte, dentre outros.
DP_FE 2	Opção de transferência para atendentes humanos, caso o chatbot realize algum tipo de atendimento ao cliente.
DP_ACE 1	Dicas de conversação para melhorar a interação e realização de perguntas de esclarecimento caso o chatbot não compreenda alguma entrada do usuário.
DP_ACE 2	A posição do chatbot em uma página web deve estar em fácil localização e acesso.
DP_EC 1	O comportamento do chatbot deve ser sempre ético. Deve ter conhecimento cultural do usuário alvo.
DP_EC 2	As respostas fornecidas devem estar acompanhadas das fontes de informação.
DP_F 1	Botões de respostas rápidas permitem agilizar a interação.
DP_F 2	A página web do chatbot deve ser responsiva e funcionar bem no formato dos principais dispositivos.
DP_F 3	As cores da fonte devem ser vivas, contrastando com o fundo da aplicação.
DP_F 4	Disponibilização de hiperlinks sobre o tópico discutido para complementar as informações geradas pelo chatbot.
DP_H 1	Compreensão e manutenção de contexto.
DP_H 2	Compreensão de erros gramaticais e de digitação comuns dos usuários.
DP_H 3	O chatbot deve ser transparente à fiscalização durante a interação, expondo sua identidade de chatbot e não alegando ser um humano.
DP_H 4	Disponibilização de avatar que gere conexão com o usuário. Oferecer opções de avatar, um para o público formal e um para o informal.
DP_H 5	Proatividade para a interação ser natural. Incentivo ao usuário para continuar a interação.
DP_RL 1	Os elementos visuais devem ser descritivos e autoexplicativos. As instruções do sistema devem ser claras e explícitas.
DP_VS 1	O usuário deve estar a par do status das ações realizadas no sistema, recebendo feedbacks imediatos.

**Tabela 7.1 – Continuação da página anterior**

<b>Código</b>	<b>Descrição do Atributo</b>
DP_PE 1	Solicitação de confirmação antes de transações, principalmente aquelas com potencial danoso e irreversível.
DP_AD 1	Disponibilização de página de ajuda e documentação sobre o uso da ferramenta.
DP_CP 1	As respostas geradas pelo chatbot devem seguir as regras oficiais de ortografia e registro linguístico.
DP_P 1	O chatbot deve ser robusto e capaz de lidar com eventuais entradas atípicas do usuário.

Legenda das Categorias: FE - Flexibilidade e Eficiência de Uso; ACE - Acessibilidade; EC - Ética e Comportamento; F - Funcionalidade; H - Humanidade; RL - Reconhecimento em vez de Lembrança; VS - Visibilidade do Status do Sistema; PE - Prevenção de Erros; AD - Ajuda e Documentação; CP - Consistência e Padrões; P - Performance.

## 7.2.2 Identificação dos chatbots reais para os exemplos visuais

Após a conclusão da listagem de atributos de qualidade, a próxima etapa era fornecer exemplos de como os esses atributos foram implementados em chatbots reais. O intuito desse levantamento foi possibilitar oferecer exemplos práticos de implementação que pudessem ser adicionados aos padrões de design de interface propostos. Desta maneira, iniciou-se a busca por chatbots em formato mobile ou site.

Os aplicativos e websites foram escolhidos seguindo critérios como disponibilização e utilização gratuitas, além de conteúdo em língua portuguesa ou inglesa. Para identificar os chatbots mobile, foi realizada uma busca automática na loja de aplicativos Play Store, para dispositivos Android, sendo priorizados aqueles com extensa quantidade de downloads e nota de avaliação 4 ou superior. Os websites, por sua vez, são derivados de uma busca automática, realizada através do buscador Google.

Ao final do processo, foram descobertos e selecionados os seguintes aplicativos e websites: (APP01) Bing AI, (WEB01) Kuki AI, (APP02) Goat Chat, (WEB02) C&A, (WEB03) Octa AI, (WEB04) Sofia Botfriend, (WEB05) Zurich Seguros e (WEB06) Evie Bot. Após a escolha das aplicações, realizou-se uma análise com o propósito de identificar quais atributos de qualidade estavam presentes em cada uma delas. A Tabela 7.2, a seguir, apresenta o ranqueamento dos chatbots mobile e aplicações web analisadas, ordenadas pelo total de atributos de qualidade presentes nelas.

Observa-se pela Tabela 7.2, que os chatbots com o maior número de atributos encontrados foram (APP01) Bing AI, (WEB01) Kuki AI, (APP02) Goat Chat. As ferramentas com a menor quantidade de atributos encontrados foram (WEB04) Sofia

Tabela 7.2 – Análise de atributos encontrados nas aplicações web/aplicativos analisados

Chatbot/ Atributos	Bing	Kuki	Goat	C&A	Octa	Sofia	Zurich	Evie	Ocorr.
DP_FE 1		X							1
DP_FE 2				X					1
DP_ACE 1	X						X		2
DP_ACE 2				X					1
DP_EC 1	X								1
DP_EC 2	X								1
DP_F 1					X				1
DP_F 2		X							1
DP_F 3	X	X	X	X	X		X		6
DP_F 4			X						1
DP_H 1						X		X	2
DP_H 2	X								1
DP_H 3	X								1
DP_H 4		X						X	2
DP_H 5	X	X							2
DP_RL 1					X				1
DP_VS 1	X								1
DP_PE 1						X			1
DP_AD 1		X	X						2
DP_CP 1	X		X						2
DP_P 1	X								1
Total Atributos por Chatbot	10	6	4	3	3	2	2	2	

Botfriend, (WEB05) Zurich Seguros e (WEB06) Evie Bot. Por sua vez, o atributo com a maior ocorrência foi DP\_F 3, que se refere à escolha de cores vivas para as fontes em contraste com o fundo, favorecendo a melhor visualização das informações.

### 7.2.3 Proposta dos Padrões de Design

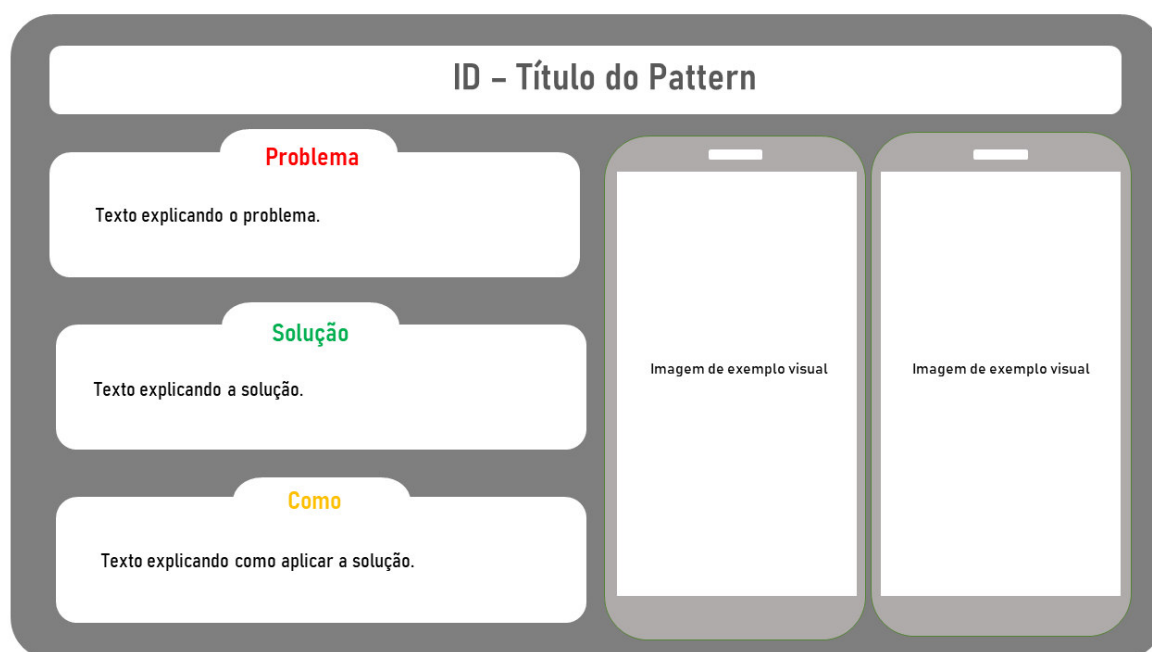
Para que os padrões possam transmitir o que são, seu funcionamento e como devem ser aplicados para resolver um problema de design, é importante serem documentados. Para desenvolver a documentação desses *patterns*, foram utilizados os elementos básicos sugeridos por [Vora \(2009\)](#), também testados no trabalho de [Gomes et al. \(2021\)](#). Desta forma, cada recomendação desenvolvida inclui:

- (1) Código: Identificador do padrão de acordo com sua categoria.
- (2) Nome do padrão: Breve título que expressa sobre o que se trata o padrão.
- (3) Problema: Breve resumo do problema a ser solucionado pelo padrão.
- (4) Solução: A solução proposta para o problema apresentado.
- (5) Implementação: Instruções e dicas de como aplicar a solução.

(6) Exemplo: Imagem contendo um exemplo de implementação do padrão com base em chatbots reais.

Por meio desses elementos dos padrões sugeridos por [Vora \(2009\)](#), tornou-se possível adaptar e estruturar o conjunto de recomendações. Considerando isso, a Figura 7.1, a seguir, apresenta o modelo de recomendação de design adotado para esta pesquisa.

Figura 7.1 – Modelo de Padrão de Design



Fonte: Elaboração Própria

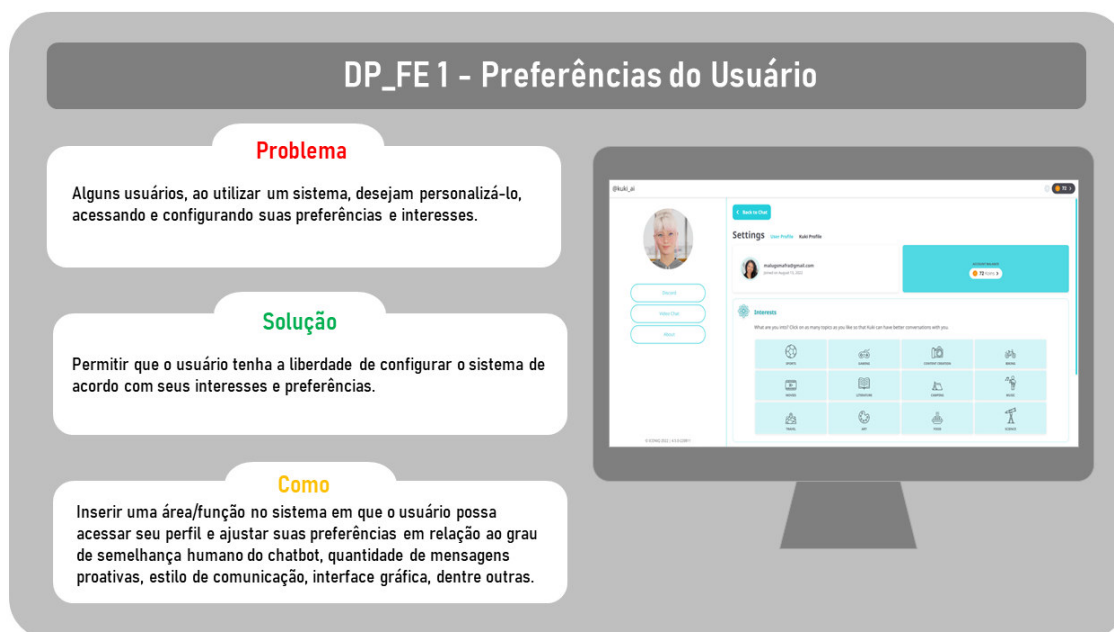
### 7.3 O conjunto de Padrões de Design DP-U2CHATBOT

Após a execução dos procedimentos necessários à criação das recomendações de design para chatbots, obteve-se um conjunto de 21 *patterns*, englobando problemas de humanidade, acessibilidade, funcionalidade, performance, dentre outros. A seguir, são apresentados alguns dos *patterns* elaborados.

Em relação aos padrões apresentados neste capítulo, a recomendação de design DP\_FE 1 da Figura 7.2 foi elaborada contendo diretrizes sobre a necessidade de permitir que o usuário personalize a interação com o chatbot, seja com opções referentes a mudanças na interface, no estilo de conversação, semelhança humana, dentre outros tipos de personalização. O exemplo visual utilizado foi o chatbot web Kuki AI. Já a recomendação de design DP\_FE 2 exposta na Figura 7.3, contém diretrizes sobre a capacidade que um chatbot de atendimento deve ter para transferir a conversa a um atendente humano quando não puder auxiliar o usuário. O exemplo visual utilizado foi

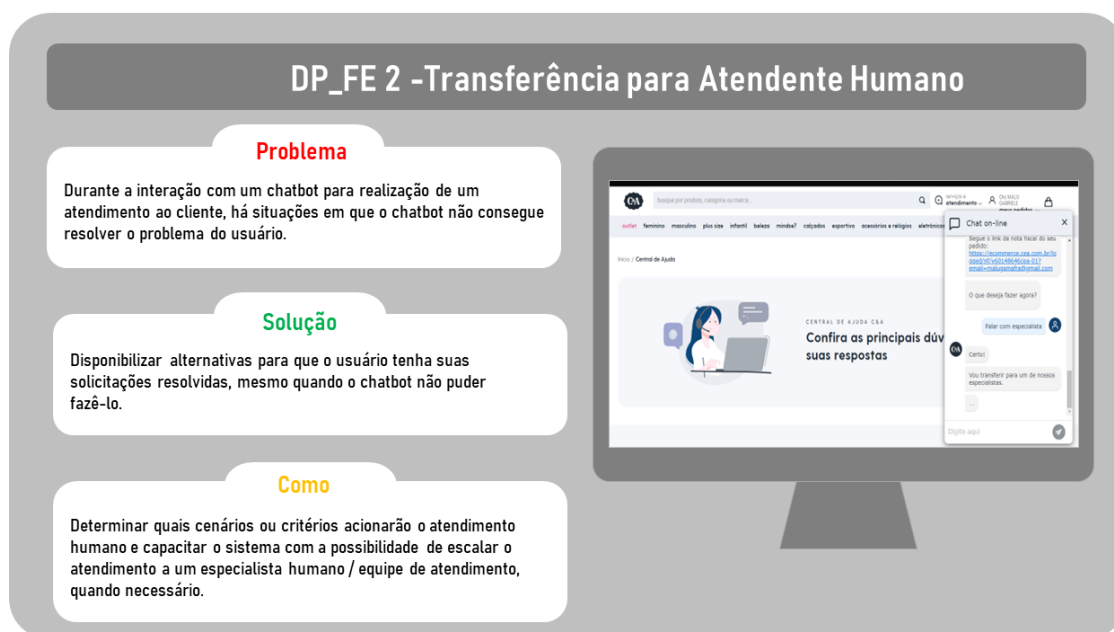


Figura 7.2 – Padrão de Design DP\_FE 1



Fonte: Elaboração Própria

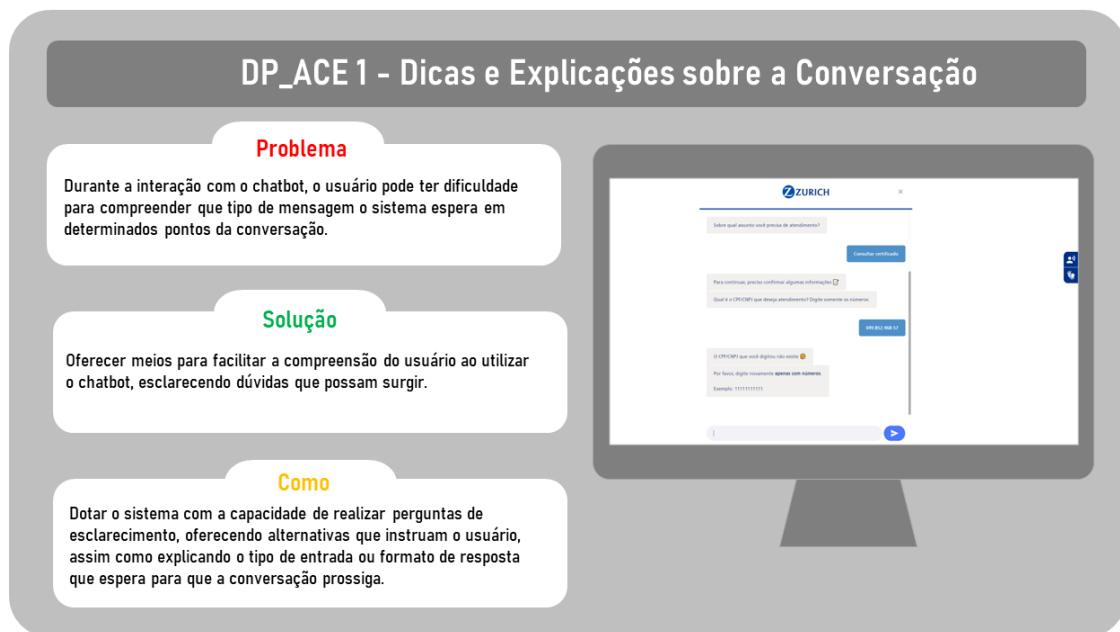
Figura 7.3 – Padrão de Design DP\_FE 2



Fonte: Elaboração Própria

o chatbot web disponível na loja de varejo C&A. E, por fim, a recomendação de design ACE\_1 exposta na Figura 7.4, traz orientação sobre o fornecimento de dicas de conversação e esclarecimentos sobre o prosseguimento da interação, quando o usuário tiver dificuldade de compreender o tipo de mensagem/entrada que o sistema espera. O exemplo visual utilizado foi o chatbot web da Zurich Seguros.

Figura 7.4 – Padrão de Design DP\_ ACE 1



Fonte: Elaboração Própria

Devido a quantidade de padrões de design desenvolvidos, torna-se inviável apresentá-los na íntegra neste capítulo e, por isto, a lista completa pode ser encontrada no **Apêndice A** desta dissertação.

## 7.4 Considerações Finais

Este capítulo apresentou o processo de desenvolvimento das recomendações de design DP-U2CHATBOT, um conjunto de *patterns* de interface voltados a chatbots. Esses padrões propostos foram elaborados utilizando os itens de verificação do *checklist* U2CHATBOT, desenvolvido com atributos de qualidade de Usabilidade e User Experience advindos de revisões da literatura.

Com a realização dos procedimentos necessários, tornou-se possível construir 21 padrões de design (recomendações) dispostos em categorias como humanidade, acessibilidade, performance, funcionalidade e outros. Cada recomendação possui documentação contendo código, problema, solução, como implementar a solução e exemplo visual de chatbot real que já funciona de acordo com padrão de design.

O capítulo a seguir, apresenta os procedimentos adotados para a validação das recomendações de design DP-U2CHATBOT por especialistas em desenvolvimento de software, bem como os resultados alcançados.

## 8 Avaliação dos Padrões de Design

Neste capítulo apresenta-se o processo de avaliação dos Padrões de Design DP-U2CHATBOT desenvolvidos nesta pesquisa. Serão descritos o planejamento e a execução do estudo experimental realizado para avaliar o desempenho e a aceitação da tecnologia. Além disso, são apresentadas as melhorias realizadas no processo de refinamento da tecnologia.

### 8.1 Introdução do Capítulo

Padrões de design têm a capacidade de prover benefícios como soluções de design comprovadas e orientações para seu uso, interfaces consistentes e processos de design aprimorados. Para obter esses benefícios é necessário que os padrões de design sejam documentados e disponibilizados em um formato que promova a reutilização da tecnologia. Apesar disso, não há consenso sobre como os padrões devem ser documentados, mantidos e compartilhados, segundo [Vora \(2009\)](#).

No Capítulo anterior foram apresentados os procedimentos adotados para que as recomendações de design para chatbots DP-U2CHATBOT pudessem ser desenvolvidas. Levando em consideração a importância de validar a documentação do conjunto de *patterns* de design DP-U2CHATBOT, este capítulo apresenta um estudo que visou avaliar a compreensibilidade e a utilidade da tecnologia desenvolvida. A estudo foi realizado com desenvolvedores de software com experiência em projeto de chatbots.

A seguir, é apresentada a organização do restante desse capítulo. A Seção 8.2 apresenta a descrição da avaliação, contendo informações sobre a elaboração do questionário de avaliação, a execução do estudo e a coleta dos dados. A Seção 8.3 descreve os resultados da avaliação das recomendações. A Seção 8.4 apresenta as melhorias aplicadas à tecnologia. Por fim, a Seção 8.5, conclui esse capítulo.

### 8.2 Descrição da Avaliação

O objetivo desta avaliação foi analisar as recomendações de design para chatbots DP-U2CHATBOT com o propósito de verificar se a tecnologia atende aos indicadores de compreensibilidade e utilidade definidos conforme o trabalho de [Gomes et al. \(2021\)](#). O estudo foi realizado entre setembro e outubro de 2023 com a participação de desenvolvedores de software e especialistas com experiência em projetos de chatbots. É importante destacar

que os participantes deste estudo concordaram com o Termo de Consentimento Livre e Esclarecido - TCLE.

### 8.2.1 Questionário de Avaliação

Para verificar a viabilidade dos padrões propostos elaborou-se um questionário na ferramenta Google Forms, baseado nos trabalhos de [Jacob \(2011\)](#) e [Gomes et al. \(2021\)](#). O Formulário de Avaliação <sup>1</sup> foi desenvolvido em três seções: A primeira seção foi elaborada para recolher os dados demográficos dos especialistas; a segunda seção focou em avaliar individualmente os aspectos de compreensibilidade e utilidade de cada padrão de design proposto e a terceira seção focou em uma avaliação geral dos padrões de design.

Em relação ao aspecto **compreensibilidade**, a questão realizada para cada pattern individualmente foi: “*Como você avalia a compreensibilidade (facilidade de entender) de cada pattern?*”. Por sua vez, para o aspecto **utilidade**, a questão foi: “*Como você avalia a utilidade de cada pattern?*”. Além disso, os participantes foram questionados sobre como avaliavam de modo geral a utilidade dos patterns para: (1) Identificar problemas a serem evitados no desenvolvimento de chatbots; (2) Buscar soluções para problemas identificados durante o processo de desenvolvimento de chatbots e (3) Entender propostas de projeto de interface de um chatbot focado em Usabilidade e User Experience. A subseção a seguir aborda detalhes sobre a execução da avaliação e a coleta dos dados.

### 8.2.2 Execução e Coleta dos Dados

O convite para participação na avaliação foi enviado aos profissionais no dia 02 de setembro de 2023. Os materiais consistiram no questionário de avaliação desenvolvido e o conjunto de 21 padrões de design DP-U2CHATBOT a serem avaliados. O prazo para participação foi 06 de outubro de 2023.

Após a finalização do prazo para participação, foram obtidas quatro respostas ao formulário. Os especialistas que responderam eram professores, engenheiros e pesquisadores com diferentes níveis de experiência em design/desenvolvimento de chatbots, sendo que dois dos profissionais mencionaram já ter tido contato com padrões de design, conforme a Tabela 8.1 a seguir. A experiência média dos especialistas com desenvolvimento de chatbots é 02 anos.

Tabela 8.1 – Caracterização dos Participantes

Participante	Atuação e Experiência
Avaliador 1	Processamento de Imagens e Engenharia de Software; Docência. Experiência com desenvolvimento de chatbots e padrões de design.

<sup>1</sup> Formulário de Avaliação dos Padrões de Design: <https://tinyurl.com/FormAvaliacaoDP-U2Chatbot>

Avaliador 2	Desenvolvimento de Sistemas e Aprendizado de Máquina; Docência. Experiência com desenvolvimento de chatbots.
Avaliador 3	Processamento de Imagens; Docência. Experiência com desenvolvimento de chatbots.
Avaliador 4	Engenharia de Software; Docência. Experiência com padrões de design de interface.

Durante a avaliação individual de cada recomendação de design, os especialistas forneceram uma nota para avaliar a compreensibilidade e utilidade de cada um deles, sendo a correspondência das notas: 4 - Muito compreensível/Muito útil, 3 - Compreensível/Útil, 2 - Pouco compreensível/Pouco útil e 1 - Nada compreensível/Nada útil. Para impossibilitar que os participantes não dessem suas opiniões sobre qualquer uma das questões aplicadas, não foi disponibilizada a opção “Neutro”. Os resultados da avaliação estão apresentados na seção a seguir.

### 8.3 Resultados

Após o processo de execução da avaliação e coleta dos dados obtidos, tornou-se possível extrair os resultados. Cada um dos *patterns* elaborados foram submetidos à avaliação quanto aos indicadores de compreensibilidade e utilidade. Todas as notas atribuídas a essas recomendações foram contabilizadas e, conforme o questionário desenvolvido, variavam da nota máxima 4 (Muito Compreensível/Muito Útil) a menor nota 1 (Nada Compreensível/Nada Útil). A medida adotada para a análise das notas foi o cálculo da **mediana**, que tem como propósito estatístico informar a posição central ou “o ponto médio” dos valores analisados. É importante destacar que se adotou o cálculo da mediana porque nos dados obtidos é possível identificar valores extremos que afetariam a média aritmética, conforme indica Pachani (2006). Os dados quantitativos desta avaliação individual podem ser encontrados nas Tabelas 8.2 e 8.3.

A Tabela 8.2, refere-se às notas dadas pelos participantes ao aspecto Compreensibilidade das recomendações de design, isto é, o quão as recomendações são fáceis de entender. Observa-se conforme a mediana das notas atribuídas pelos participantes, que os padrões foram bem avaliados em relação ao aspecto Compreensibilidade. A menor mediana identificada é 3,00, levando a crer que os participantes em geral consideraram os padrões de design compreensíveis de entender. As recomendações de design avaliadas com a nota máxima 4,00 (Muito Compreensível) por todos os participantes no aspecto de Compreensibilidade foram: DP\_EC 2 e DP\_H 2. Já os padrões de design que receberam pelo menos uma nota 1,00 (Nada Compreensível) foram: DP\_EC 1, DP\_H 1, DP\_H 4, DP\_CP 1 e DP\_P 1. Entretanto, mesmo recebendo nota 1,00 de um participante, os

Tabela 8.2 – Notas dadas pelos especialistas aos patterns considerando o aspecto Compreensibilidade

Legenda	Compreensibilidade				
	Nota 1	Nota 2	Nota 3	Nota 4	Mediana
DP_FE-1	4,0	3,0	3,0	4,0	3,50
DP_FE-2	4,0	4,0	3,0	4,0	4,00
DP_ACE-1	4,0	4,0	2,0	4,0	4,00
DP_ACE-2	4,0	4,0	2,0	4,0	4,00
DP_EC-1	4,0	4,0	1,0	4,0	4,00
DP_EC-2	4,0	4,0	4,0	4,0	4,00
DP_F-1	4,0	4,0	3,0	4,0	4,00
DP_F-2	4,0	3,0	4,0	4,0	4,00
DP_F-3	4,0	4,0	4,0	3,0	4,00
DP_F-4	3,0	3,0	4,0	3,0	3,00
DP_H-1	4,0	3,0	1,0	4,0	3,50
DP_H-2	4,0	4,0	4,0	4,0	4,00
DP_H-3	4,0	4,0	4,0	3,0	4,00
DP_H-4	4,0	4,0	1,0	4,0	4,00
DP_H-5	4,0	4,0	3,0	4,0	4,00
DP_RL-1	3,0	4,0	2,0	4,0	3,50
DP_VS-1	3,0	4,0	4,0	3,0	3,50
DP_PE-1	3,0	3,0	4,0	2,0	3,00
DP_AD-1	4,0	4,0	3,0	4,0	4,00
DP_CP-1	4,0	4,0	1,0	4,0	4,00
DP_P-1	4,0	4,0	1,0	4,0	4,00

demais participantes avaliaram esses padrões com notas 3,00 ou 4,00, sendo possível que o participante tenha sido mais criterioso que os demais, possa não ter compreendido a proposta do padrão ou que o padrão deva ser refinado melhorar sua compreensibilidade.

A Tabela 8.3, por sua vez, refere-se às notas relacionadas ao aspecto Utilidade. Nota-se que conforme a mediana das notas atribuídas pelos participantes, as recomendações foram bem avaliadas em relação ao aspecto Utilidade. A maioria dos *patterns* foi bem avaliada, levando a crer que os participantes em geral consideraram as recomendações de design úteis para apoiar o design de chatbots. Entretanto, o padrão DP\_AD 1, sobre *Documentação e Ajuda*, recebeu a menor mediana de 2,50, tendendo a ser pouco útil, conforme a avaliação dos participantes. As recomendações de design avaliadas com a nota máxima 4,00 (Muito Útil) por todos os participantes no aspecto de Utilidade foram: DP\_ACE 2, DP\_H 2, DP\_H 5, DP\_RL 1, DP\_VS 1 e DP\_P 1, todos com a nota máxima 4,00. A recomendação de design que recebeu pelo menos uma nota 1,00 (Nada Compreensível) foi apenas o DP\_EC 1, todavia, os demais participantes avaliaram este padrão como Útil (3,0) ou Muito Útil (4,0).

Considerando as duas tabelas referentes às notas atribuídas aos aspectos Compre-

Tabela 8.3 – Notas dadas pelos especialistas aos patterns considerando o aspecto Utilidade

Legenda	Utilidade				
	Nota 1	Nota 2	Nota 3	Nota 4	Mediana
DP_FE-1	4,0	3,0	3,0	4,0	3,50
DP_FE-2	3,0	4,0	3,0	4,0	3,50
DP_ACE-1	4,0	4,0	2,0	4,0	4,00
DP_ACE-2	4,0	4,0	4,0	4,0	4,00
DP_EC-1	4,0	4,0	1,0	3,0	3,50
DP_EC-2	3,0	4,0	2,0	4,0	3,50
DP_F-1	4,0	4,0	3,0	4,0	4,00
DP_F-2	4,0	3,0	4,0	3,0	3,50
DP_F-3	4,0	4,0	3,0	3,0	3,50
DP_F-4	3,0	3,0	3,0	4,0	3,00
DP_H-1	3,0	3,0	3,0	4,0	3,00
DP_H-2	4,0	4,0	4,0	4,0	4,00
DP_H-3	3,0	4,0	2,0	4,0	3,50
DP_H-4	4,0	4,0	2,0	4,0	4,00
DP_H-5	4,0	4,0	4,0	4,0	4,00
DP_RL-1	4,0	4,0	4,0	4,0	4,00
DP_VS-1	4,0	4,0	4,0	4,0	4,00
DP_PE-1	4,0	3,0	3,0	4,0	3,50
DP_AD-1	3,0	3,0	2,0	2,0	2,50
DP_CP-1	4,0	4,0	4,0	2,0	4,00
DP_P-1	4,0	4,0	4,0	4,0	4,00

ensibilidade e Utilidade, notou-se que as piores notas referentes a estes aspectos foram dadas pelos participantes que tinham menos experiência no desenvolvimento/design de chatbots em comparação ao demais. Assim, o nível de experiência neste quesito pode ter influenciado a análise do que é útil e compreensível em relação às recomendações de design propostas.

O questionário também quis avaliar a utilidade da tecnologia conforme os três critérios apresentados na Tabela 8.4. A tabela apresenta a avaliação geral das recomendações de design com as notas atribuídas pelos participantes e a média das notas. Neste caso, o adotou-se o cálculo da média aritmética, visto que os dados obtidos não apresentam valores extremos (PACHANI, 2006). Observa-se que os dois primeiros critérios definidos pontuaram a nota média 4,0 (nota máxima), significando dizer que foram bem avaliados pelos profissionais e que são muito úteis nesses quesitos. O terceiro critério foi avaliado por um dos participantes com a nota 2,0 (Pouco útil), contudo os demais participantes o avaliaram com a nota máxima 4,0 (Muito Útil), totalizando uma nota média positiva de 3,50. Assim, a pouca experiência do participante no design/desenvolvimento de chatbots, também pode ter influenciado na análise de utilidade do critério.



Tabela 8.4 – Notas dadas pelos especialistas aos critérios

Critérios	Utilidade Geral				
	Nota 1	Nota 2	Nota 3	Nota 4	Média
1 - Identificar problemas a serem evitados no desenvolvimento de chatbots?	4,0	4,0	4,0	4,0	4,0
2 - Buscar soluções para problemas identificados durante o processo de desenvolvimento de chatbots?	4,0	4,0	4,0	4,0	4,0
3 - Entender propostas de projeto de interface de um chatbot focado em Usabilidade e User Experience?	4,0	4,0	4,0	2,0	3,5

Complementarmente à avaliação, solicitou-se aos profissionais que indicassem até (5) cinco recomendações de design que deveriam ser priorizadas no processo de design/desenvolvimento de chatbots. Estas recomendações de design destacadas podem ser encontradas na Tabela 8.5. De acordo com a tabela, observa-se que alguns participantes selecionaram os mesmos padrões para serem priorizados, no entanto, a maioria dos padrões indicados foram diversos.

Tabela 8.5 – Padrões que devem ser priorizados conforme os avaliadores

Participantes	Padrões a serem priorizados
Avaliador 1	DP_FE 2 - Transferência para Atendente Humano, DP_F 2 - Design Responsivo, DP_FE-1 Preferências do Usuário, DP_H 5 - Proatividade na Comunicação do Chatbot e DP_P 1 - Robustez a Entradas Inesperadas
Avaliador 2	DP_F 1 - Botões de Respostas Rápidas, DP_F 3 - Cores vivas para as Fontes, DP_H 2 - Compreensão de Erros Gramaticais e de Digitação, DP_H 4 - Avatar, DP_RL 1 - Elementos Visuais Descritivos e Instruções Explícitas
Avaliador 3	DP_F 2 - Design Responsivo, DP_H 2 - Compreensão de Erros Gramaticais e de Digitação, DP_VS 1 - Feedback e Status do Sistema, DP_CP 1 - Ortografia e Registro Linguístico, DP_P 1 - Robustez a Entradas Inesperadas
Avaliador 4	DP_FE 2 - Transferência para Atendente Humano, DP_F 1 - Botões de Respostas Rápidas, DP_EC 2 - Fontes de Informação, DP_RL 1 - Elementos Visuais Descritivos e Instruções Explícitas, DP_VS 1 - Feedback e Status do Sistema

Conforme a Tabela 8.5, observa-se que as recomendações mais citadas pelos avaliadores a serem priorizadas no processo de design/desenvolvimento de chatbots foram:



- DP\_FE 2 - Transferência para Atendente Humano
- DP\_F 2 - Design Responsivo
- DP\_P 1 - Robustez a Entradas Inesperadas
- DP\_F 1 - Botões de Respostas Rápidas
- DP\_H 2 - Compreensão de Erros Gramaticais e de Digitação
- DP\_RL 1 - Elementos Visuais Descritivos e Instruções Explícitas
- DP\_VS 1 - Feedback e Status do Sistema

Os demais *patterns* citados para a priorização são diferentes. O caráter subjetivo dos avaliadores pode ter influenciado o processo de priorização, visto que o conceito do que é prioridade ou não neste contexto pode variar de profissional para profissional, no entanto, não se pode afirmar, já que a amostra da avaliação é muito baixa.

Ainda, os participantes foram questionados sobre a necessidade de retirada de algum padrão da listagem. Apenas um dos participantes sinalizou esta necessidade para a recomendação de design **DP\_F 4 - Uso de Hiperlinks**, justificando que *“a depender do contexto do bot e do perfil do usuário, pode gerar mais dúvidas”*. Os demais participantes não perceberam este problema. Levando em consideração que esta recomendação foi desenvolvida com a utilização de atributos de qualidade identificados na literatura científica e a amostra da avaliação é muito baixa, por agora optou-se por manter a recomendação de design DP\_F 4 - Uso de Hiperlinks.

No geral, os participantes consideraram que as recomendações dão suporte ao design de chatbots focados em atender às necessidades dos usuários, conforme destacado pelos participantes *“...cobrem os principais pontos no desenvolvimento do chat”*, além de *“...estabelecem um conjunto de detalhes que normalmente não se pensaria para um chatbot”*. Por fim, ao serem questionados se além dos elementos utilizados na definição das recomendações (código, título, problema, solução, implementação, exemplo), adicionariam alguma outra informação na definição dos *patterns* apresentados, um dos participantes destacou *“Adicionaria um contra exemplo ou Anti-pattern”*, isto é, um exemplo do que não deve ser feito no projeto de um chatbot, que inclusive é uma excelente ideia para trabalhos futuros.

## 8.4 Melhorias nos Padrões de Design DP-U2CHATBOT

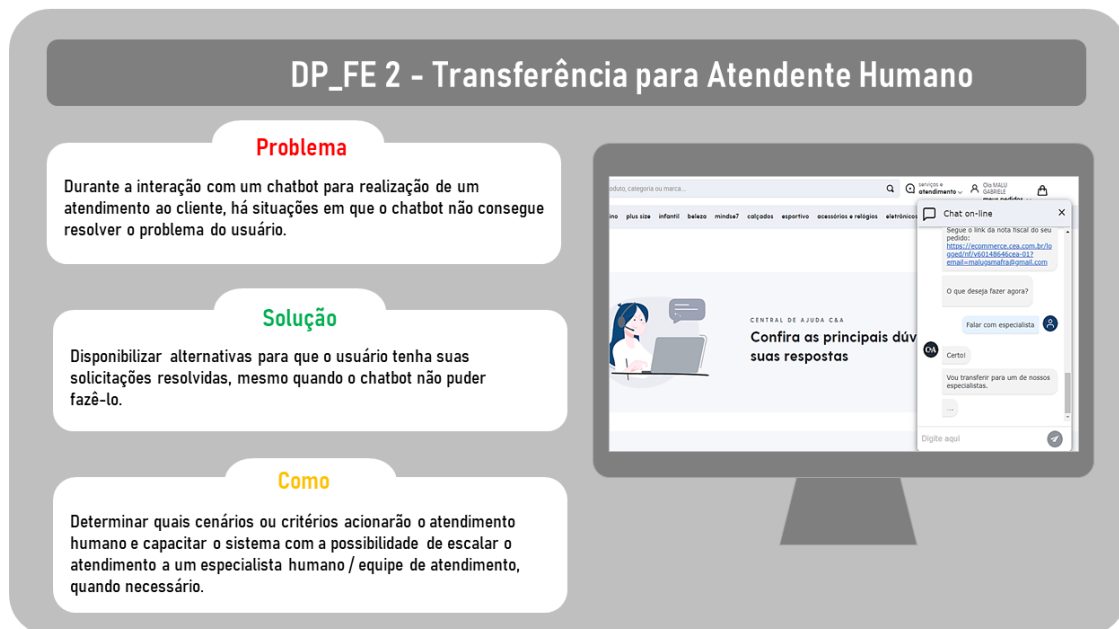
O formulário de avaliação da tecnologia DP-U2CHATBOT também questionou aos participantes sobre sugestões de mudanças e melhorias nas recomendações de design

apresentadas neste trabalho. Um avaliador sinalizou que em algumas recomendações o exemplo visual escolhido pode não ter sido o mais adequado “*Nas imagens de cada padrão, não tenho certeza se o exemplo apresenta uma aderência ou falta de aderência ao padrão*”, indicando o padrão **DP\_H 1 - Preservação de Contexto** para melhoria nesse sentido.

Outro participante destacou em relação à recomendação **DP\_EC 1 - Ética e Conhecimento Cultural**, “... *Achei confuso ética estar junto com cultura. Acho que as duas coisas são diferentes e até em alguns casos, excludentes. Imagine uma “tribo” que só fale gíria e palavrão. Essa é a cultura da tribo, mas para o seu chat seria barrado pela ética*”. Ainda, houve uma orientação pertinente sobre cautela na utilização de exemplos visuais contendo a logomarca das empresas responsáveis pelos chatbots selecionados “*Precisa ter cuidado com algumas imagens. Por exemplo, em DP\_FE 2 há uma logo da empresa C&A. O mesmo vale para os demais. Será que é permitido utilizar?*”.

Desta forma, considerando os apontamentos e sugestões de melhorias indicadas pelos profissionais que participaram da avaliação do conjunto de recomendações de design DP-U2CHATBOT desenvolvidas nesta pesquisa, foram realizadas alterações em alguns *patterns* para melhorar sua compreensibilidade e ainda evitar possíveis problemas legais futuros. Essas recomendações de design refinadas podem ser visualizadas a seguir.

Figura 8.1 – Padrão de Design DP\_FE 2



As recomendações de design **DP\_FE 2** e **DP\_ACE 1** tiveram seus exemplos visuais modificados para não expor a logomarca das empresas. Esta medida foi adotada para evitar problemas legais. Do mesmo modo, as recomendações **DP\_ACE 2** e **DP\_F 3** também foram refinados pelo mesmo motivo.

Figura 8.2 – Padrão de Design DP\_ ACE 1

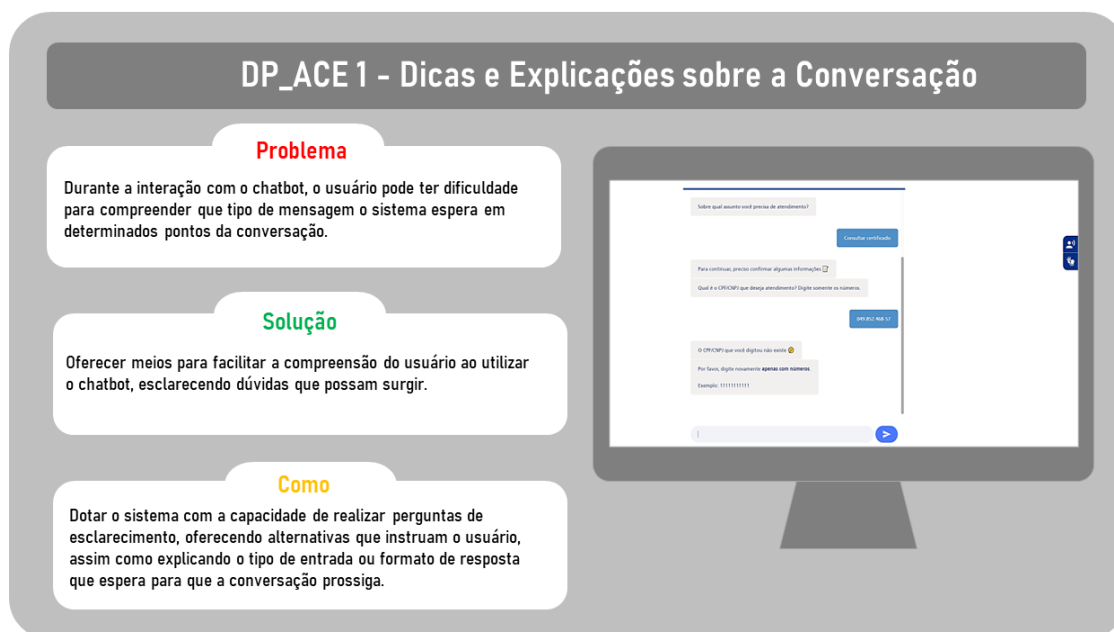
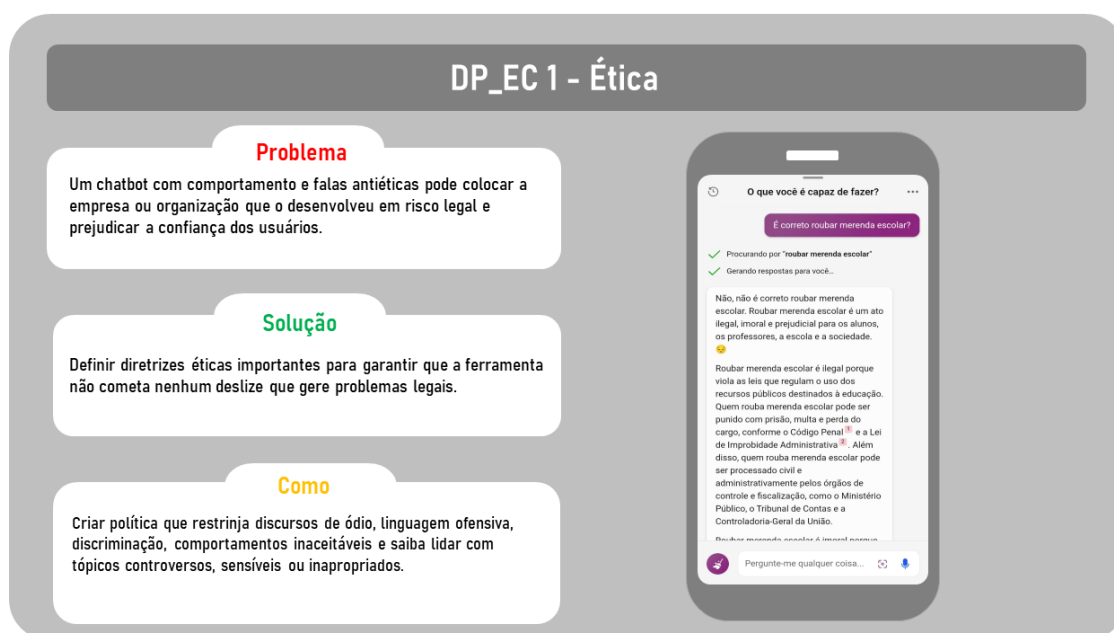


Figura 8.3 – Padrão de Design DP\_ EC 1



Por sua vez, a recomendação de design **DP\_ EC 1 - Ética e Conhecimento Cultural** foi fragmentada em duas, a recomendação **DP\_ EC 1 - Ética** e **DP\_ EC 3 - Conhecimento Cultural**. Isto porque separados esses *patterns* fazem mais sentido do que juntos, visto que versam sobre assuntos diferentes. A última melhoria realizada foi na recomendação **DP\_ H1 - Preservação de Contexto**, já que o exemplo visual escolhido anteriormente não apresentava a melhor aderência ao que o padrão versa, isto é, a capacidade do chatbot manter o contexto entre sessões de conversação.

Figura 8.4 – Padrão de Design DP\_EC 3

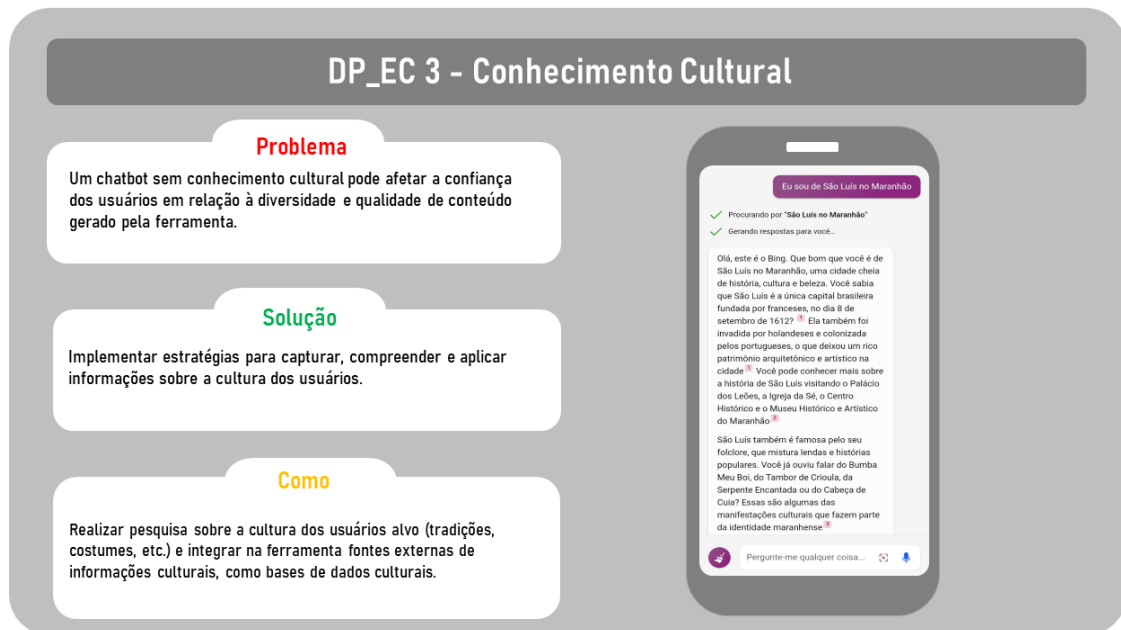
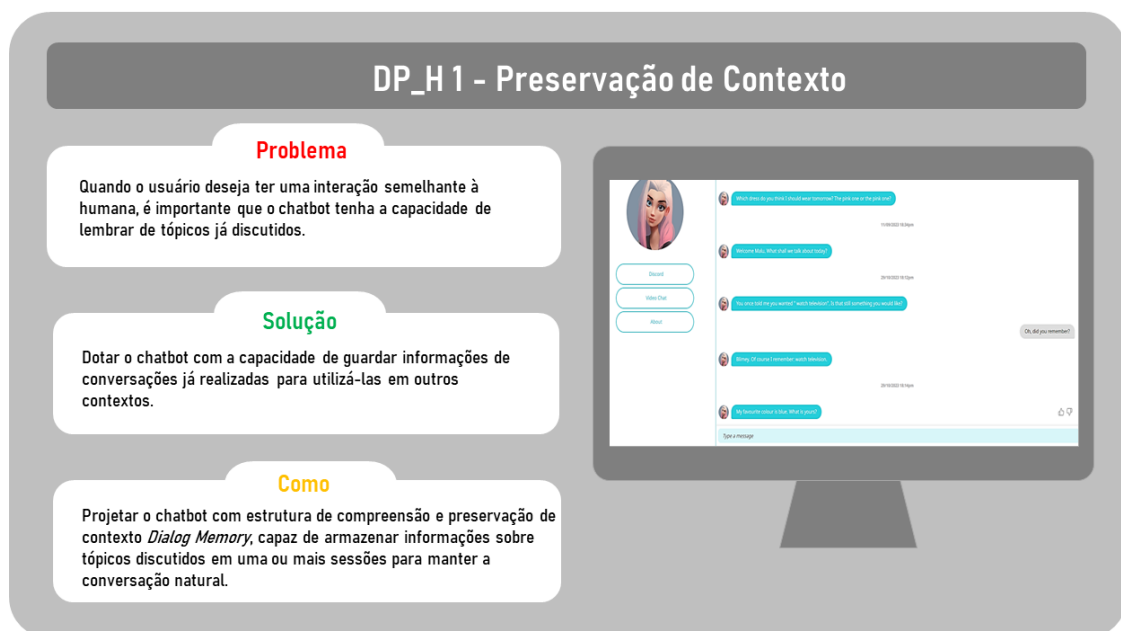


Figura 8.5 – Padrão de Design DP\_H 1



Todas as recomendações de design refinadas podem ser encontradas no **Apêndice B**, localizado nos documentos pós-textuais desta dissertação.

## 8.5 Considerações Finais

Este capítulo apresentou os procedimentos adotados à avaliação do conjunto de recomendações de design DP-U2CHATBOT, elaborado com o intuito de apoiar a etapa de design de chatbots textuais inteligentes. A avaliação das recomendações de design

foi realizada por profissionais com experiência em desenvolvimento/design de chatbots, que analisaram cada um dos 21 patterns desenvolvidos e, após isto, responderam a um questionário que englobava aspectos de compreensibilidade e utilidade, além de questões subjetivas sobre quais padrões deveriam ser priorizados, sugestões e mudanças que poderiam ser feitas.

De maneira geral, verificou-se por meio dos resultados que os participantes avaliaram as recomendações de design positivamente em relação aos aspectos Compreensibilidade e Utilidade, indicando que as diretrizes desenvolvidas são fáceis de entender, além de serem úteis para apoiar desenvolvedores no projeto de chatbots, abordando os principais pontos para a construção de um chatbot focado em satisfazer seus usuários. Algumas recomendações se mostraram mais prioritárias do que outras e há aspectos que podem ser aperfeiçoados e adicionados em trabalhos futuros, para que as recomendações de design se tornem melhores.

Considerando as sugestões de melhorias apontadas pelos profissionais avaliadores, alguns padrões sofreram mudanças a fim de refinar sua proposta. Ocorreram mudanças nos exemplos visuais a fim de garantir maior aderência ao padrão, logomarcas foram removidas e uma recomendação foi fragmentada em duas para conferir mais clareza à tecnologia. Após este processo de refinamento, obteve-se 22 recomendações de design. No capítulo a seguir, serão apresentadas as conclusões e as perspectivas futuras sobre esta pesquisa de mestrado.

## 9 Conclusão e Trabalhos Futuros

Neste capítulo são apresentadas as considerações finais sobre o desenvolvimento do *checklist* de inspeção U2CHATBOT e das Recomendações de Design DP-U2CHATBOT, tecnologias para o auxílio no design e na avaliação de chatbots. Também são apresentadas as expectativas de trabalhos futuros, orientando novas etapas dessa pesquisa no campo de qualidade de software para agentes conversacionais.

### 9.1 Epílogo

Esta dissertação de mestrado buscou contribuir com um estudo focado na melhoria da qualidade de chatbots inteligentes. Para isto, consideraram-se atributos de qualidade de Usabilidade e UX da literatura para guiar o processo de desenvolvimento de tecnologias para apoiar a avaliação e o design de chatbots. A contribuição referente a avaliação de chatbots foi o *checklist* de inspeção U2CHATBOT, útil para identificar defeitos nessas ferramentas. Em relação à etapa de design dessas ferramentas, foi proposto um conjunto de *design patterns* de interface, contendo diretrizes, recomendações e exemplos para auxiliar desenvolvedores.

Mesmo com vários estudos focados em propor tecnologias para avaliar e melhorar a qualidade de chatbots, percebeu-se que usuários ainda apresentavam insatisfação em relação a essas ferramentas. Em lojas de aplicativos, por exemplo, é possível encontrar diversas reclamações e avaliações negativas dos usuários. Visando minimizar esse problema, foi proposto o desenvolvimento de um novo *checklist* de inspeção focado em Usabilidade e User Experience para identificar defeitos em chatbots, o U2CHATBOT. Tanto a Usabilidade, quanto User Experience são fatores importantes a serem considerados quando o objetivo é garantir o sucesso de um sistema e a satisfação dos seus usuários. Para isso considerou-se a realização de revisões da literatura para a obtenção de atributos de qualidade relacionados a Usabilidade e User Experience em artigos científicos.

À medida que os atributos de qualidade foram identificados, houve um tratamento nos itens repetidos ou semelhantes, permitindo assim criar o *checklist* de inspeção U2CHATBOT. Inicialmente, o *checklist* foi submetido a uma prova de conceito para avaliar um chatbot do Google Play Store. A prova de conceito consistiu em identificar problemas em diversos comentários de usuários da ferramenta. Os problemas identificados foram comparados com os itens do *checklist*; os resultados mostraram que o *checklist* conseguiu identificar 96% dos problemas relatados pelos usuários.

Após o desenvolvimento do U2CHATBOT, fez necessária a avaliação da tecnologia para verificar sua viabilidade de uso. Para o estudo experimental foram recrutados 29 participantes para inspecionar um chatbot; os resultados da avaliação indicaram que *checklist* U2CHATBOT possui utilidade para identificação de defeitos em chatbots, contudo a quantidade de itens de verificação constitui um ponto fraco, visto que afeta a facilidade de uso percebida da técnica. Logo, por constituir mais itens de verificação, o U2CHATBOT torna-se uma opção melhor em casos de inspeções mais precisas e completas.

Além do *checklist* U2CHATBOT, proposto para avaliação de chatbots, identificou-se que a etapa de design de chatbots também carece de contribuições. Os estudos sobre padrões de design para estas ferramentas ainda são insipientes. Além disso, alguns desenvolvedores necessitam de apoio visual e informações adicionais para projetarem suas ferramentas com a qualidade esperada. Desta forma, os atributos de qualidade descobertos nas revisões da literatura também foram utilizados para contribuir com a melhoria da qualidade na etapa de design de chatbots. Para isso, foram desenvolvidos 21 recomendações de design para auxiliar desenvolvedores na construção dessas aplicações. O conjunto de recomendações DP-U2CHATBOT contém diretrizes para solucionar problemas comuns de design e cada *pattern* possui documentação que inclui código, nome do padrão, problema, solução, implementação da solução e exemplos visuais de chatbots reais em operação.

Após o desenvolvimento das recomendações de design, a tecnologia foi submetida a avaliação. Quatro desenvolvedores de software, com experiência em chatbots, analisaram cada padrão individualmente e responderam a um questionário elaborado para aferir a compreensibilidade e utilidade da técnica, além de propor sugestões de melhorias. Os resultados indicaram que os participantes consideraram, de forma geral, que as recomendações de design são úteis e compreensíveis, contudo, é importante que a tecnologia seja testada e avaliada por mais desenvolvedores. Ainda considerando as sugestões dos participantes, algumas recomendações de design sofreram alterações a fim de refinar sua proposta. As alterações foram propostas nos exemplos visuais para garantir maior aderência ao padrão, as logomarcas foram retiradas e um padrão foi fragmentado em dois para dar maior clareza ao artefato. Após esse processo de refinamento, a tecnologia DP-U2CHATBOT somou 22 recomendações de design para chatbots.

Em suma, os resultados das avaliações demonstraram que as tecnologias desenvolvidas nesta pesquisa foram bem aceitas por parte dos inspetores participantes, com boa utilidade percebida para auxiliar engenheiros de software e desenvolvedores, experientes ou não, na avaliação e melhoria de seus chatbots, mas há a necessidade de replicar as avaliações com mais participantes a fim de garantir resultados mais precisos sobre as tecnologias.

## 9.2 Contribuições

Como principais contribuições desta dissertação de mestrado, pode-se destacar:

- Estudos secundários sobre atributos de qualidade e técnicas de avaliação focadas em Usabilidade e User Experience importantes para chatbots textuais. Esses estudos utilizam uma metodologia capaz de ser replicada em outros trabalhos e seus resultados servem como base para a concepção de outras tecnologias nessa linha de pesquisa.
- O desenvolvimento de um *checklist* de inspeção específico para identificação de defeitos em chatbots. O *checklist* contém itens de verificação baseados em atributos de qualidade focados em Usabilidade e User Experience. A tecnologia pode ser utilizada por desenvolvedores e equipes de desenvolvimento no processo de avaliação de suas ferramentas.
- Desenvolvimento do conjunto de recomendações de design de interface para chatbots, DP-U2CHATBOT. Esses padrões contêm diretrizes sobre problemas comuns encontrados no design dessas ferramentas, assim como instruções e informações de como resolvê-los, podendo auxiliar desenvolvedores na construção de seus chatbots.
- Definições de um estudo experimental para a avaliação do *checklist* de inspeção U2CHATBOT, servindo como modelo para ser replicado em trabalhos com o objetivo semelhante.
- Divulgação dos procedimentos e resultados deste estudo por meio de artigos científicos. Durante o desenvolvimento desta dissertação de mestrado, foram publicados artigos relacionados à pesquisa. A seguir, são apresentadas essas publicações:

### **Artigo publicado relacionado a esta pesquisa**

- MAFRA, Malu et al. Defining Requirements for the Development of Useful and Usable Chatbots: An Analysis of Quality Attributes from Academy and Industry. In: International Conference on Human-Computer Interaction. Springer, Cham, 2022. p. 479-493.

### **Artigo submetido à Journal on Interactive Systems - Aguardando aprovação até a data da publicação desta dissertação**

- MAFRA, Malu et al. Proposing Usability-UX technologies for the design and evaluation of text-based chatbots. In: JIS - Journal on Interactive Systems.



## 9.3 Perspectivas Futuras

O desenvolvimento deste estudo permitiu propor tecnologias para avaliar e para apoiar o design de chatbots, como o *checklist* U2CHATBOT e as recomendações de design DP-U2CHATBOT. Os resultados alcançados abrem espaço a novas perspectivas que podem ser exploradas em trabalhos futuros. A seguir, são apresentadas algumas dessas perspectivas futuras:

### Sobre o *Checklist* de inspeção U2CHATBOT

- Melhorar o apoio ferramental do *checklist*, visando facilitar o processo de descrição dos problemas encontrados pelos inspetores e possibilitar também anexar as imagens com a localização dos defeitos, aumentando ainda mais a facilidade de uso da técnica.
- Atualizar os atributos de qualidade de Usabilidade e UX para que o *checklist* de inspeção englobe também os chatbots ativados por voz.
- Realizar outras possíveis melhorias nas descrições dos itens de verificação para melhorar a compreensão; além de melhorar o roteiro de inspeção, adicionando mais exemplos de tarefas a serem realizadas no chatbot avaliado para facilitar a análise de cada item de verificação do *checklist*.

### Sobre os Padrões de Design DP-U2CHATBOT

- Aplicar a avaliação das recomendações com um número maior de especialistas e analisar os resultados coletados.
- Identificar mais exemplos de chatbots reais presentes no mercado para enriquecer os *patterns* desenvolvidos.
- Criação de anti-patterns, indicando exemplos reais do que **não** deve ser feito no design de chatbots.

## Referências

- ADAMOPOULOU, E.; MOUSSIADES, L. An overview of chatbot technology. In: SPRINGER. *IFIP international conference on artificial intelligence applications and innovations*. [S.l.], 2020. p. 373–383. Citado na página 15.
- ALMEIDA, D. C.; PITANGA, H. N.; SILVA, T. O. d.; SILVA, N. A. B.; AVELAR, M. G. d. Utilização dos testes estatísticos kruskal-wallis e mann-whitney para avaliação de sistemas de solos reforçados com geotêxteis. *Matéria (Rio de Janeiro)*, SciELO Brasil, v. 27, 2022. Citado na página 75.
- ALSAYED, A. O.; BILGRAMI, A. L.; FOSTER, W. Improving software quality management: testing, review, inspection and walkthrough. *International Journal of Latest Research in Science and Technology*, v. 6, n. 1, p. 7–12, 2017. Citado 3 vezes nas páginas 16, 21 e 48.
- ANSHU, K.; GAUR, L.; SOLANKI, A. Impact of chatbot in transforming the face of retailing-an empirical model of antecedents and outcomes. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, Bentham Science Publishers, v. 14, n. 3, p. 774–787, 2021. Citado 2 vezes nas páginas 35 e 52.
- BARTIÉ, A. *Garantia da qualidade de software*. [S.l.]: Gulf Professional Publishing, 2002. Citado na página 48.
- BEVAN, N.; CARTER, J.; HARKER, S. Iso 9241-11 revised: What have we learnt about usability since 1998? In: SPRINGER. *International conference on human-computer interaction*. [S.l.], 2015. p. 143–151. Citado na página 16.
- BORSCHI, S.; MALIZIA, A.; SCHMETTOW, M.; VELDE, F. V. D.; TARIVERDIYEVA, G.; BALAJI, D.; CHAMBERLAIN, A. The chatbot usability scale: the design and pilot of a usability scale for interaction with ai-based conversational agents. *Personal and Ubiquitous Computing*, Springer, p. 1–25, 2021. Citado 6 vezes nas páginas 21, 32, 35, 43, 46 e 48.
- BRILL, T. M.; MUNOZ, L.; MILLER, R. J. Siri, alexa and other digital assistants: a study of customer satisfaction with artificial intelligence applications. *Journal of Marketing Management*, Taylor & Francis, v. 35, n. 15-16, p. 1401–1436, 2019. Citado na página 15.
- BRYKCZYNSKI, B. A survey of software inspection checklists. *ACM SIGSOFT Software Engineering Notes*, ACM New York, NY, USA, v. 24, n. 1, p. 82, 1999. Citado na página 20.
- CABREJOS, L. J. E. R.; VIANA, D.; SANTOS, R. P. dos. Planejamento e execução de estudos secundários em informática na educação: Um guia prático baseado em experiências. *Jornada de Atualização em Informática na Educação*, v. 7, n. 1, p. 21–52, 2018. Citado 5 vezes nas páginas 24, 26, 38, 39 e 40.
- CHAVES, A. P.; GEROSA, M. A. How should my chatbot interact? a survey on social characteristics in human–chatbot interaction design. *International Journal of*

*Human-Computer Interaction*, Taylor & Francis, v. 37, n. 8, p. 729–758, 2021. Citado na página 15.

CHOW, J. C.; SANDERS, L.; LI, K. Design of an educational chatbot using artificial intelligence in radiotherapy. *AI*, MDPI, v. 4, n. 1, p. 319–332, 2023. Citado na página 37.

CIECHANOWSKI, L.; PRZEGALINSKA, A.; WEGNER, K. The necessity of new paradigms in measuring human-chatbot interaction. In: SPRINGER. *Advances in Cross-Cultural Decision Making: Proceedings of the AHFE 2017 International Conference on Cross-Cultural Decision Making, July 17-21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA 8*. [S.l.], 2018. p. 205–214. Citado na página 15.

CODINA, L. Scopus: el mayor navegador científico de la web. *El profesional de la información*, v. 14, n. 1, p. 44–49, 2005. Citado na página 25.

COPPOLA, R.; ARDITO, L. Quality assessment methods for textual conversational interfaces: A multivocal literature review. *Information*, Multidisciplinary Digital Publishing Institute, v. 12, n. 11, p. 437, 2021. Citado 2 vezes nas páginas 24 e 38.

CRUZ, Y. P.; COLLAZOS, C. A.; GRANOLLERS, T. The thin red line between usability and user experiences. In: *Proceedings of the xvi international conference on human computer interaction*. [S.l.: s.n.], 2015. p. 1–2. Citado 2 vezes nas páginas 16 e 17.

CYBIS, W.; BETIOL, A. H.; FAUST, R. Ergonomia e usabilidade. *Conhecimentos, métodos e aplicações*. São Paulo: Novatec, 2007. Citado na página 21.

DALMORO, M.; VIEIRA, K. M. Dilemas na construção de escalas tipo likert: o número de itens e a disposição influenciam nos resultados? *Revista gestão organizacional*, v. 6, n. 3, 2013. Citado na página 69.

DAVIS, F. D.; BAGOZZI, R. P.; WARSHAW, P. R. User acceptance of computer technology: A comparison of two theoretical models. *Management science*, INFORMS, v. 35, n. 8, p. 982–1003, 1989. Citado 2 vezes nas páginas 65 e 69.

DAVIS, F. D.; VENKATESH, V. A critical assessment of potential measurement biases in the technology acceptance model: three experiments. *International journal of human-computer studies*, Elsevier, v. 45, n. 1, p. 19–45, 1996. Citado na página 65.

DENECKE, K.; HOCHREUTENER, S. L.; PÖPEL, A.; MAY, R. Self-anamnesis with a conversational user interface: concept and usability study. *Methods of information in medicine*, Georg Thieme Verlag KG, v. 57, n. 05/06, p. 243–252, 2018. Citado na página 33.

DENECKE, K.; WARREN, J. How to evaluate health applications with conversational user interface? *Studies in health technology and informatics*, IOS Press, v. 270, p. 976–980, 2020. Citado 2 vezes nas páginas 33 e 38.

DUYNE, D. K. V.; LANDAY, J. A.; HONG, J. I. *The design of sites: Patterns for creating winning web sites*. [S.l.]: Prentice Hall Professional, 2007. Citado na página 101.

FEINE, J.; ADAM, M.; BENKE, I.; MAEDCHE, A.; BENLIAN, A. Exploring design principles for enterprise chatbots: An analytic hierarchy process study. In: SPRINGER. *International Conference on Design Science Research in Information Systems and Technology*. [S.l.], 2020. p. 126–141. Citado na página 34.

FERNANDEZ, A.; ABRAHÃO, S.; INSFRAN, E. Empirical validation of a usability inspection method for model-driven web development. *Journal of Systems and Software*, Elsevier, v. 86, n. 1, p. 161–186, 2013. Citado na página 74.

FRAZAO, K.; COSTA, J.; VIANA, D.; RIVERO, L. Analyzing app store comments and quality attributes for defining an inspection checklist for mobile educational games. In: *Proceedings of the 34th Brazilian Symposium on Software Engineering*. [S.l.: s.n.], 2020. p. 854–859. Citado na página 16.

FRAZÃO, K. A. et al. Ic-meg: Um checklist específico para avaliação de jogos educacionais digitais em plataformas móveis. Universidade Federal do Maranhão, 2021. Citado 6 vezes nas páginas 23, 48, 49, 65, 69 e 74.

GEORGESCU, A.-A. et al. Chatbots for education—trends, benefits and challenges. In: ”CAROL I” NATIONAL DEFENCE UNIVERSITY PUBLISHING HOUSE. *Conference proceedings of» eLearning and Software for Education «(eLSE)*. [S.l.], 2018. v. 2, n. 14, p. 195–200. Citado na página 15.

GOMES, B. R.; JR, A. F. L. J.; PINTO, I. d. J. P.; COLCHER, S. Ágata: um chatbot para difusão de práticas para educação ambiental. *Anais Estendidos do XXVI Simpósio Brasileiro de Sistemas Multimídia e Web*, p. 85–89, 2020. Citado na página 15.

GOMES, D.; PINTO, N.; MELO, A.; MAIA, I.; PAIVA, A.; BARRETO, R.; VIANA, D.; RIVERO, L. Developing a set of design patterns specific for the design of user interfaces for autistic users. In: *Proceedings of the XX Brazilian Symposium on Human Factors in Computing Systems*. [S.l.: s.n.], 2021. p. 1–7. Citado 8 vezes nas páginas 16, 19, 22, 101, 102, 105, 109 e 110.

GUERINO, G. C.; VALENTIM, N. M. C. Usability and user experience evaluation of conversational systems: A systematic mapping study. In: *Proceedings of the 34th Brazilian Symposium on Software Engineering*. [S.l.: s.n.], 2020. p. 427–436. Citado 5 vezes nas páginas 16, 17, 21, 24 e 38.

HASSAN, H. M.; GALAL-EDEEN, G. H. From usability to user experience. In: IEEE. *2017 International Conference on Intelligent Informatics and Biomedical Sciences (ICIBMS)*. [S.l.], 2017. p. 216–222. Citado na página 17.

HEFNY, W. E.; BOLOCK, A. E.; HERBERT, C.; ABDENNADHER, S. Chase away the virus: a character-based chatbot for covid-19. In: IEEE. *2021 IEEE 9th International Conference on Serious Games and Applications for Health (SeGAH)*. [S.l.], 2021. p. 1–8. Citado na página 32.

HÖHN, S.; BONGARD-BLANCHY, K. Heuristic evaluation of covid-19 chatbots. In: SPRINGER. *International Workshop on Chatbot Research and Design*. [S.l.], 2020. p. 131–144. Citado 6 vezes nas páginas 31, 32, 34, 43, 46 e 48.

IACOB, C. Identifying, relating, and evaluating design patterns for the design of software for synchronous collaboration. In: *Proceedings of the 3rd ACM SIGCHI symposium on Engineering interactive computing systems*. [S.l.: s.n.], 2011. p. 323–326. Citado na página 110.

- JIN, Y.; CAI, W.; CHEN, L.; HTUN, N. N.; VERBERT, K. Musicbot: Evaluating critiquing-based music recommenders with conversational interaction. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. [S.l.: s.n.], 2019. p. 951–960. Citado na página 34.
- KADARIYA, D.; VENKATARAMANAN, R.; YIP, H. Y.; KALRA, M.; THIRUNARAYANAN, K.; SHETH, A. kbot: knowledge-enabled personalized chatbot for asthma self-management. In: IEEE. *2019 IEEE International Conference on Smart Computing (SMARTCOMP)*. [S.l.], 2019. p. 138–143. Citado na página 34.
- KALINOWSKI, M. Introdução à inspeção de software. *Revista Engenharia de Software: Qualidade de software*, v. 1, p. 5, 2008. Citado na página 85.
- KALINOWSKI, M.; SPÍNOLA, R. O.; TRAVASSOS, G. H. Infra-estrutura computacional para apoio ao processo de inspeção de software. In: SBC. *Anais do III Simpósio Brasileiro de Qualidade de Software*. [S.l.], 2004. p. 117–131. Citado na página 19.
- KITCHENHAM, B.; CHARTERS, S. Guidelines for performing systematic literature reviews in software engineering. Citeseer, 2007. Citado na página 19.
- KOCABALLI, A. B.; LARANJO, L.; COIERA, E. Measuring user experience in conversational interfaces: a comparison of six questionnaires. In: *Proceedings of the 32nd International BCS Human Computer Interaction Conference 32*. [S.l.: s.n.], 2018. p. 1–12. Citado na página 17.
- KOCABALLI, A. B.; LARANJO, L.; COIERA, E. Understanding and measuring user experience in conversational interfaces. *Interacting with Computers*, Oxford University Press, v. 31, n. 2, p. 192–207, 2019. Citado na página 17.
- KULIGOWSKA, K. Commercial chatbot: performance evaluation, usability metrics and quality standards of embodied conversational agents. *Professionals Center for Business Research*, v. 2, 2015. Citado na página 48.
- LAITENBERGER, O.; EMAM, K. E.; HARBICH, T. G. An internally replicated quasi-experimental comparison of checklist and perspective based reading of code documents. *IEEE Transactions on Software Engineering*, IEEE, v. 27, n. 5, p. 387–421, 2001. Citado na página 60.
- LAND, L. P. W.; SAUER, C.; JEFFERY, R. Validating the defect detection performance advantage of group designs for software reviews: report of a laboratory experiment using program code. *ACM SIGSOFT Software Engineering Notes*, ACM New York, NY, USA, v. 22, n. 6, p. 294–309, 1997. Citado na página 71.
- LANGVIN, R.; LORDON, R. J.; AVRAHAMI, T.; COWAN, B. R.; HIRSCH, T.; HSIEH, G. Heuristic evaluation of conversational agents. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, ACM, p. 1–15, 5 2021. Disponível em: <<https://dl.acm.org/doi/10.1145/3411764.3445312>>. Citado 3 vezes nas páginas 17, 31 e 33.
- LARSSON, M. *Predicting quality attributes in component-based software systems*. Tese (Doutorado) — Mälardalen University, 2004. Citado na página 23.

- LEITE, J. C. S. d. P. *Como Registrar Requisitos de Software*. [S.l.]: <https://www-di.inf.puc-rio.br/julio/Livro-qualidade-2002.pdf>/(accessed), 2002. Citado na página 51.
- LEWIS, J. R. The system usability scale: past, present, and future. *International Journal of Human-Computer Interaction*, Taylor & Francis, v. 34, n. 7, p. 577–590, 2018. Citado na página 43.
- MADAN, A.; DUBEY, S. K. Usability evaluation methods: a literature review. *International Journal of Engineering Science and Technology*, Engg Journals Publications, v. 4, n. 2, p. 590–599, 2012. Citado na página 16.
- MCALLISTER, P.; KERR, J.; MCTEAR, M.; MULVENNA, M.; BOND, R.; KIRBY, K.; MORNING, J.; GLOVER, D. Towards chatbots to support bibliotherapy preparation and delivery. In: SPRINGER. *Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019, Amsterdam, The Netherlands, November 19–20, 2019, Revised Selected Papers 3*. [S.l.], 2020. p. 127–142. Citado na página 33.
- MELLO, R. M. de. *Técnica para Inspeção de Diagramas de Atividades*. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, 2011. Citado na página 60.
- MENDES, J. P. M. et al. Lgpd-check: Um checklist para avaliação da aderência de sistemas computacionais à lei geral de proteção de dados brasileira. Universidade Federal do Maranhão, 2022. Citado na página 60.
- MIRNIG, A. G.; MESCHTSCHERJAKOV, A.; WURHOFER, D.; MENEWEGER, T.; TSCHELIGI, M. A formal analysis of the iso 9241-210 definition of user experience. In: *Proceedings of the 33rd annual ACM conference extended abstracts on human factors in computing systems*. [S.l.: s.n.], 2015. p. 437–450. Citado na página 17.
- MOTAUNG, T. *Design attributes for a successful Online Retail Chatbot Information System*. Tese (Doutorado) — University of Johannesburg, 2022. Citado na página 15.
- MOYA, C. R. *Como escolher o teste estatístico: um guia para o pesquisador iniciante*. Dissertação (Mestrado) — Universidade Cruzeiro do Sul, 2021. Citado 2 vezes nas páginas 75 e 76.
- MUÑOZ, L.; AVILA, O. A model to assess customer alignment through customer experience concepts. In: SPRINGER. *International Conference on Business Information Systems*. [S.l.], 2019. p. 339–351. Citado na página 15.
- NICULESCU, A. I.; YEO, K. H.; D’HARO, L. F.; KIM, S.; JIANG, R.; BANCHS, R. E. Design and evaluation of a conversational agent for the touristic domain. In: IEEE. *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*. [S.l.], 2014. p. 1–10. Citado na página 35.
- NIELSEN, J. *Ten Usability Heuristics*. [S.l.]: <http://www.nngroup.com/articles/ten-usability-heuristics/>(accessed), 2005. Citado 3 vezes nas páginas 32, 35 e 46.
- NILSSON, E. G. Design patterns for user interface for mobile applications. *Advances in engineering software*, Elsevier, v. 40, n. 12, p. 1318–1328, 2009. Citado na página 22.
- PACHANI, R. A. Cálculo e uso de mediana. *Exacta*, Universidade Nove de Julho, v. 4, n. 2, p. 417–423, 2006. Citado 2 vezes nas páginas 111 e 113.



- PECUNE, F.; MURALI, S.; TSAI, V.; MATSUYAMA, Y.; CASSELL, J. A model of social explanations for a conversational movie recommendation system. In: *Proceedings of the 7th International Conference on Human-Agent Interaction*. [S.l.: s.n.], 2019. p. 135–143. Citado na página 34.
- RADZIWILL, N. M.; BENTON, M. C. Evaluating quality of chatbots and intelligent conversational agents. *arXiv preprint arXiv:1704.04579*, 2017. Citado 4 vezes nas páginas 15, 21, 23 e 38.
- RAHMAN, A.; MAMUN, A. A.; ISLAM, A. Programming challenges of chatbot: Current and future prospective. In: IEEE. *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*. [S.l.], 2017. p. 75–78. Citado na página 15.
- RAPP, A.; CURTI, L.; BOLDI, A. 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*, Elsevier, v. 151, p. 102630, 2021. Citado na página 15.
- REICHERTS, L.; PARK, G. W.; ROGERS, Y. Extending chatbots to probe users: Enhancing complex decision-making through probing conversations. In: *Proceedings of the 4th Conference on Conversational User Interfaces*. [S.l.: s.n.], 2022. p. 1–10. Citado na página 16.
- ROSRUEN, N.; SAMANCHUEN, T. Chatbot utilization for medical consultant system. In: IEEE. *2018 3rd technology innovation management and engineering science international conference (TIMES-iCON)*. [S.l.], 2018. p. 1–5. Citado na página 15.
- RUBIN, R. B.; MARTIN, M. M. Development of a measure of interpersonal communication competence. *Communication Research Reports*, Taylor & Francis, v. 11, n. 1, p. 33–44, 1994. Citado na página 32.
- SAURO, J.; LEWIS, J. R. When designing usability questionnaires, does it hurt to be positive? In: *Proceedings of the SIGCHI conference on human factors in computing systems*. [S.l.: s.n.], 2011. p. 2215–2224. Citado na página 46.
- SEDOC, J.; IPPOLITO, D.; KIRUBARAJAN, A.; THIRANI, J.; UNGAR, L.; CALLISON-BURCH, C. Chateval: A tool for chatbot evaluation. *NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Demonstrations Session*, 2019. Citado na página 48.
- SHARMA, P. Review paper on contextual chatbot for covid-19 updates. *IITM Journal of Management and IT*, Institute of Information Technology and Management, v. 12, n. 1, p. 36–37, 2021. Citado na página 15.
- SHARMA, R. K.; JOSHI, M. An analytical study and review of open source chatbot framework, rasa. *Int. J. Eng. Res.*, v. 9, n. 06, p. 1011–1014, 2020. Citado na página 16.
- SHARMA, V.; GOYAL, M.; MALIK, D. An intelligent behaviour shown by chatbot system. *International Journal of New Technology and Research*, Nextgen Research Publication, v. 3, n. 4, p. 263312, 2017. Citado na página 15.

- SKJUVE, M.; BRANDZAEG, P. B. Measuring user experience in chatbots: An approach to interpersonal communication competence. In: SPRINGER. *Internet Science: INSCI 2018 International Workshops, St. Petersburg, Russia, October 24–26, 2018, Revised Selected Papers 5*. [S.l.], 2019. p. 113–120. Citado 2 vezes nas páginas 32 e 33.
- SPERLÍ, G. A deep learning based chatbot for cultural heritage. In: *Proceedings of the 35th Annual ACM Symposium on Applied Computing*. [S.l.: s.n.], 2020. p. 935–937. Citado na página 15.
- SPILIOTOPOULOS, D.; KOTIS, K.; VASSILAKIS, C.; MARGARIS, D. Semantics-driven conversational interfaces for museum chatbots. In: SPRINGER. *International Conference on Human-Computer Interaction*. [S.l.], 2020. p. 255–266. Citado na página 34.
- SUGISAKI, K.; BLEIKER, A. Usability guidelines and evaluation criteria for conversational user interfaces: a heuristic and linguistic approach. In: *Proceedings of the Conference on Mensch und Computer*. [S.l.: s.n.], 2020. p. 309–319. Citado 7 vezes nas páginas 21, 31, 33, 43, 46, 48 e 52.
- SUHAILI, S. M.; SALIM, N.; JAMBLI, M. N. Service chatbots: A systematic review. *Expert Systems with Applications*, Elsevier, v. 184, p. 115461, 2021. Citado 2 vezes nas páginas 24 e 38.
- SVENNINGSSON, N.; FARAON, M. Artificial intelligence in conversational agents: A study of factors related to perceived humanness in chatbots. In: *Proceedings of the 2019 2nd Artificial Intelligence and Cloud Computing Conference*. [S.l.: s.n.], 2019. p. 151–161. Citado na página 33.
- THORAT, S. A.; JADHAV, V. A review on implementation issues of rule-based chatbot systems. In: *Proceedings of the international conference on innovative computing & communications (ICICC)*. [S.l.: s.n.], 2020. Citado na página 15.
- TIME, B. P. M. Mapa do ecossistema brasileiro de bots 2021. *Revista Mobile Time*, 2021. Citado na página 15.
- TIME, B. P. M. Mapa do ecossistema brasileiro de bots 2023. *Revista Mobile Time*, 2023. Citado na página 15.
- VORA, P. *Web application design patterns*. [S.l.]: Morgan Kaufmann, 2009. Citado 4 vezes nas páginas 101, 105, 106 e 109.
- YANG, X.; AURISICCHIO, M. Designing conversational agents: A self-determination theory approach. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. [S.l.: s.n.], 2021. p. 1–16. Citado na página 34.



# Apêndices

# APÊNDICE A – Lista completa dos Padrões de Design DP-U2CHATBOT

Figura A.1 – Padrão de Design - Modelo



Figura A.2 – Padrão de Design DP\_FE 1

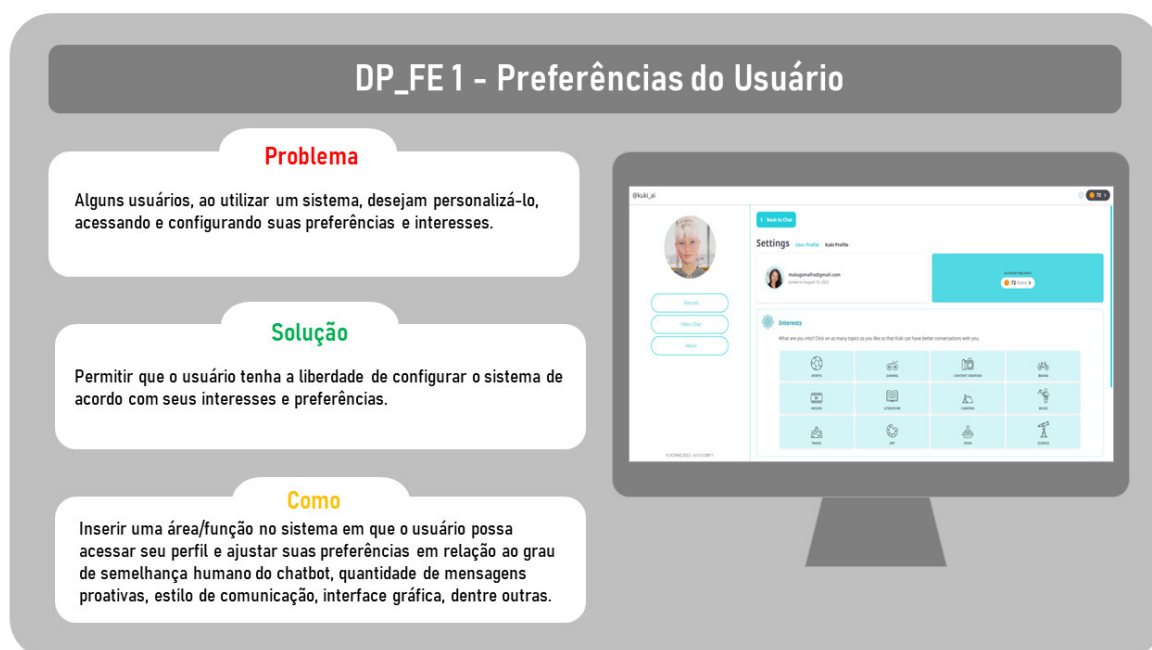


Figura A.3 – Padrão de Design DP\_FE 2

### DP\_FE 2 -Transferência para Atendente Humano

Problema

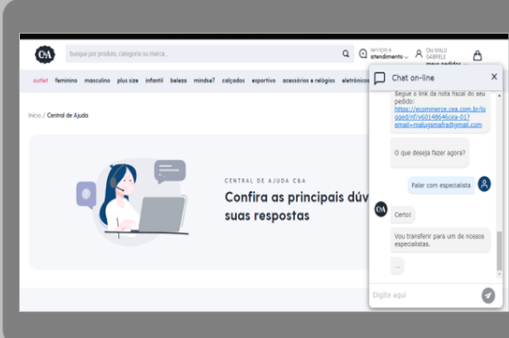
Durante a interação com um chatbot para realização de um atendimento ao cliente, há situações em que o chatbot não consegue resolver o problema do usuário.

Solução

Disponibilizar alternativas para que o usuário tenha suas solicitações resolvidas, mesmo quando o chatbot não puder fazê-lo.

Como

Determinar quais cenários ou critérios acionarão o atendimento humano e capacitar o sistema com a possibilidade de escalar o atendimento a um especialista humano / equipe de atendimento, quando necessário.



The screenshot shows a chatbot window on a computer monitor. The chatbot has asked 'O que deseja fazer agora?' and offers the option 'Falar com especialista'. A confirmation message says 'Você transferir para um de nossos especialistas.' and there is a 'Digite aqui' input field at the bottom.

Figura A.4 – Padrão de Design DP\_ACE 1

### DP\_ACE 1 - Dicas e Explicações sobre a Conversação

Problema

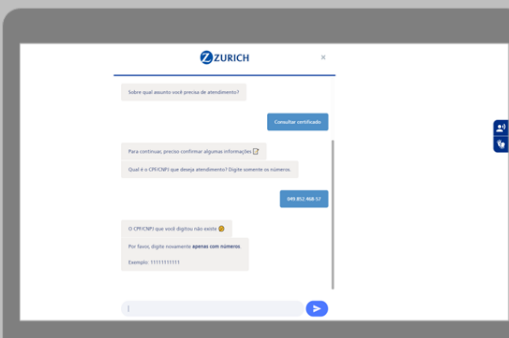
Durante a interação com o chatbot, o usuário pode ter dificuldade para compreender que tipo de mensagem o sistema espera em determinados pontos da conversação.

Solução

Oferecer meios para facilitar a compreensão do usuário ao utilizar o chatbot, esclarecendo dúvidas que possam surgir.

Como

Dotar o sistema com a capacidade de realizar perguntas de esclarecimento, oferecendo alternativas que instruem o usuário, assim como explicando o tipo de entrada ou formato de resposta que espera para que a conversação prossiga.



The screenshot shows a chatbot window on a computer monitor. It displays several prompts: 'Sabes qual assunto está próximo de finalizar?', 'Para continuar, preciso confirmar alguma informação?', and 'Qual é o CPF/CPF que deseja atendimento? Digite somente os números.' There are also buttons for 'Consultar histórico' and 'Ver meu CPF'.

Figura A.5 – Padrão de Design DP\_ ACE 2

## DP\_ ACE 2 - Facilidade de Localização e Acesso

Problema

Os chatbots embutidos em páginas web podem estar em condições ou locais difíceis de serem localizados e acessados pelo usuário.

Solução

Posicionar o chatbot em área estratégica, que possa ser visualizada desde o primeiro momento do acesso ao site.

Como

Analisar a página como um todo e escolher a posição, o design e o tamanho adequado para que o chatbot se destaque, chamando a atenção do usuário. Optar por inserir a ferramenta na primeira dobra do site ou aplicativo, que é o local com a melhor visibilidade.



Figura A.6 – Padrão de Design DP\_ EC 1

## DP\_ EC 1 - Ética e Conhecimento Cultural

Problema

Um chatbot sem ética e sem conhecimento cultural pode colocar a empresa ou organização que o desenvolveu em risco legal e prejudicar a confiança dos usuários.

Solução

Definir diretrizes éticas e implementar estratégias para capturar, compreender e aplicar informações sobre a cultura dos usuários

Como

Criar política que restrinja discursos de ódio, linguagem ofensiva, discriminação, comportamentos inaceitáveis e saiba lidar com tópicos controversos, sensíveis ou inapropriados. Realizar pesquisa sobre a cultura dos usuários alvo (tradições, costumes, etc.) e integrar fontes externas de informações culturais, como bases de dados culturais.

O que você é capaz de fazer?

É correto roubar merenda escolar?

✓ Procurando por "roubar merenda escolar"  
✓ Gerando respostas para você...

Não, não é correto roubar merenda escolar. Roubar merenda escolar é um ato ilegal, imoral e prejudicial para os alunos, os professores, a escola e a sociedade.

Roubar merenda escolar é ilegal porque viola as leis que regulam o uso dos recursos públicos destinados à educação. Quem rouba merenda escolar pode ser punido com prisão, multa e perda do cargo, conforme o Código Penal e a Lei de Improbidade Administrativa. Além disso, quem rouba merenda escolar pode ser processado civil e administrativamente pelos órgãos de controle e fiscalização, como o Ministério Público, o Tribunal de Contas e a Controladoria-Geral da União.

Roubar merenda escolar é ilegal mesmo.

Pergunte-me qualquer coisa...

Eu sou de São Luís no Maranhão

✓ Procurando por "São Luís no Maranhão"  
✓ Gerando respostas para você...

Olá, este é o Bing. Que bom que você é de São Luís no Maranhão, uma cidade cheia de história, cultura e beleza. Você sabia que São Luís é a única capital brasileira fundada por franceses, no dia 8 de setembro de 1612? Ela também foi invadida por holandeses e colonizada pelos portugueses, o que deixou um rico patrimônio arquitetônico e artístico na cidade. Você pode conhecer mais sobre a história de São Luís visitando o Palácio dos Leões, a Igreja da Sé, o Centro Histórico e o Museu Histórico e Artístico do Maranhão.

São Luís também é famosa pelo seu folclore, que mistura lendas e histórias populares. Você já ouviu falar do Bumba Meu Boi, do Tambor de Crochê, da Serpente Encantada ou do Cabeça de Cuiá? Essas são algumas das manifestações culturais que fazem parte da identidade maranhense.

Pergunte-me qualquer coisa...

Figura A.7 – Padrão de Design DP\_EC 2

## DP\_EC 2 - Fontes de Informação

Problema

O usuário precisa ter confiança nas informações geradas pelo chatbot a respeito de tópicos específicos, considerando a veracidade e relevância.

Solução

Permitir que o chatbot transmita credibilidade indicando as referências de onde as informações foram retiradas.

Como

Fornecer ao usuário a origem das informações que está compartilhando, disponibilizando as referências (livros, revistas) e URLs (sites) logo após as respostas. Selecionar as informações de fontes reconhecidamente confiáveis.

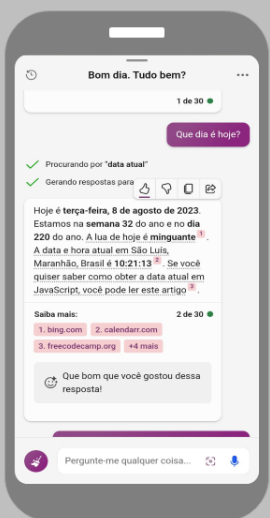


Figura A.8 – Padrão de Design DP\_F 1

## DP\_F 1 - Botões de Respostas Rápidas

Problema

Durante um suporte ao cliente, a interação com chatbot pode ser difícil caso o usuário seja inexperiente, tenha problemas para digitar ou não saiba exatamente como descrever seu problema.

Solução

Oferecer meios de facilitar a interação do usuário com o chatbot, agilizando o processo.

Como

Disponibilizar durante o atendimento ou interação, opções de respostas rápidas que ajudem o usuário no processo de descrição do seu problema, acelerando a interação.

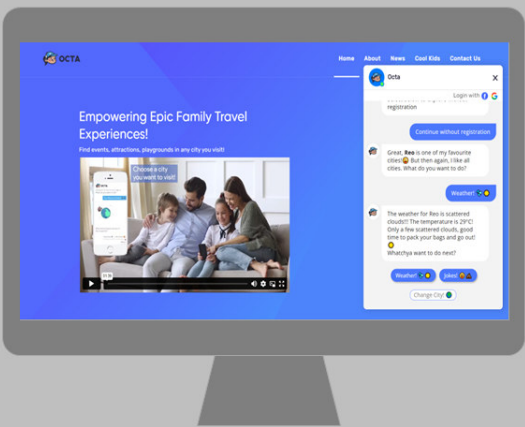


Figura A.9 – Padrão de Design DP\_F 2

### DP\_F 2 - Design Responsivo

**Problema**

A forma de apresentação de conteúdo web é diferente em dispositivos como Laptops, Tablets, Smartphones, visto que possuem tamanhos diferentes.

**Solução**

Aplicar técnicas de redimensionamento a fim de que o conteúdo web se adapte da melhor forma de acordo com o dispositivo/browser adotado pelo usuário.

**Como**

Escolher plataforma de desenvolvimento que permita criar chatbots responsivos, como Dialogflow, Rasa, Microsoft Bot Framework, IBM Watson Assistant. Utilizar técnicas de CSS responsivo para ajustar o design do chatbot conforme o tamanho da tela.



Figura A.10 – Padrão de Design DP\_F 3

### DP\_F 3 - Cores vivas para as Fontes

**Problema**

A falta de contraste entre fundo e fonte prejudica a leitura e visualização de conteúdo, bem como a boa interação entre o usuário e o sistema.

**Solução**

Escolher as cores certas para que o contraste entre fundo e fonte garantam que o usuário tenha uma boa experiência de visualização e leitura.

**Como**

Optar por fundo claro com fontes escuras. Entretanto, pode-se utilizar a fórmula do livro Wayfinding (1992) de Arthur & Passini para escolher as cores que contrastem adequadamente com o fundo escolhido.



Figura A.11 – Padrão de Design DP\_F 4

### DP\_F 4 - Uso de Hiperlinks

Problema

As respostas dadas pelo chatbot podem não ser suficientes ao usuário, que pode sentir a necessidade de obter mais informações além do conteúdo gerado.

Solução

Dotar o sistema com hiperlinks ligados ao tópico solicitado pelo usuário, de forma que ele possa acessar e se aprofundar no assunto.

Como

Implementar os hiperlinks com tratamento de cliques, redirecionando o usuário ao site vinculado. Determinar a forma como os hiperlinks serão exibidos nas mensagens do chatbot, em formato de texto clicável ou como botões.



Figura A.12 – Padrão de Design DP\_H 1

### DP\_H 1 - Preservação de Contexto

Problema

Quando o usuário deseja ter uma interação semelhante à humana, é importante que o chatbot tenha a capacidade de lembrar de tópicos já discutidos.

Solução

Dotar o chatbot com a capacidade de guardar informações de conversações já realizadas para utilizá-las em outros contextos.

Como

Projetar o chatbot com estrutura de compreensão e preservação de contexto *Dialog Memory*, capaz de armazenar informações sobre tópicos discutidos em uma ou mais sessões para manter a conversação natural.



Figura A.13 – Padrão de Design DP\_H 2

DP\_H 2 - Compreensão de Erros Gramaticais e de Digitação

Problema

Usuários com baixa escolaridade, pouca experiência com tecnologias ou nativo de outro idioma podem utilizar o chatbot. Desta forma, erros gramaticais e de digitação podem ocorrer.

Solução

Capacitar o sistema de forma que compreenda eventuais erros gramaticais e de digitação.

Como

Utilizar modelo de linguagem avançado, como o GPT-3 da OpenAI. Construir dicionários fuzzy que contenham variações de palavras e erros de digitação comuns. Utilizar bibliotecas de PLN, como NLTK ou SpaCy, para implementar técnicas de pré-processamento que normalizem palavras e tratem variações ortográficas.

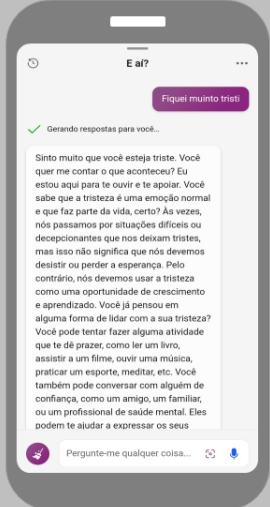


Figura A.14 – Padrão de Design DP\_H 3

DP\_H 3 - Transparência à Fiscalização

Problema

Se o chatbot apresentar interação muito semelhante à humana, eventualmente o usuário pode se questionar se está conversando com um ser humano ou com um chatbot.

Solução

Garantir que o chatbot deixe claro desde o início da interação que é um agente conversacional e não uma pessoa.

Como

Ao iniciar a interação com o usuário, o chatbot deve ser transparente e se apresentar, informando que é um agente conversacional e dando um panorama das suas funções e capacidades.

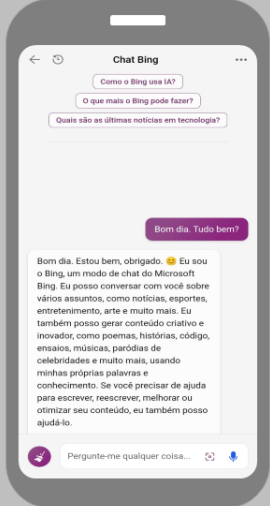




Figura A.15 – Padrão de Design DP\_H 4

## DP\_H 4 - Avatar

Problema

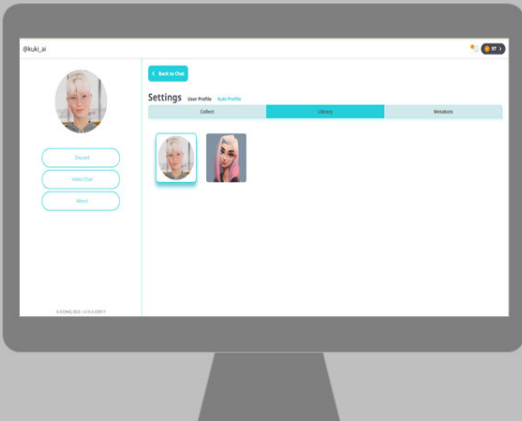
Um chatbot sem avatar pode prejudicar a interação entre usuário e sistema, dificultando a conexão emocional e a construção de confiança, já que avatares podem contribuir para a reduzir a sensação de estar interagindo com um sistema.

Solução

Disponibilizar avatares ou animações que imprimam ao sistema personalidade humana e uma identificação visual, considerando as características do público alvo.

Como

Realizar uma análise de perfil público alvo, criando assim uma opção de avatar com estilo clássico para o público formal e um avatar moderno para o público informal que vier a utilizar o sistema.



The screenshot shows a 'Settings' page for a chatbot. On the left, there are three buttons: 'Avatar', 'Voice', and 'Name'. The 'Avatar' button is highlighted. In the main content area, there are two avatar options: a classic-style female avatar and a modern-style female avatar. The interface is clean and user-friendly.

Figura A.16 – Padrão de Design DP\_H 5

## DP\_H 5 - Proatividade na Comunicação do Chatbot

Problema

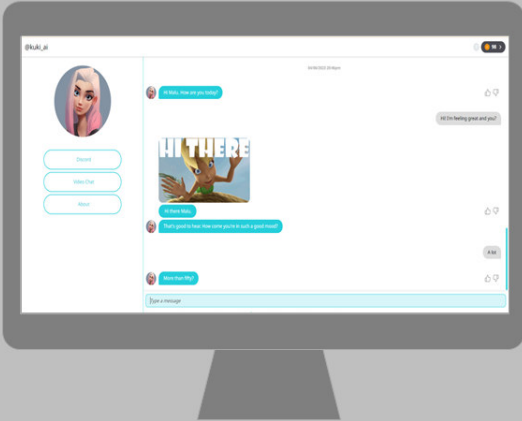
Alguns chatbots não são proativos e esperam sempre o *input* do usuário para prosseguir com a interação. Isto é um problema quando o usuário busca uma interação natural.

Solução

Dotar o chatbot da capacidade de iniciativa durante a conversação, de forma que não seja apenas o usuário encarregado desse processo.

Como

Projetar o chatbot para enviar a primeira mensagem assim que o usuário entrar no chat e fazer perguntas proativas para que a comunicação não seja unilateral. Além disso, após um período de ociosidade, incentivar o usuário a retornar à interação.



The screenshot shows a chatbot interface with a female avatar. The chat history shows a proactive message from the chatbot: 'Hi there! How are you today?'. The user's response is 'I'm good, how are you?'. The interface includes a 'Send Message' button and a 'Close' button.

Figura A.17 – Padrão de Design DP\_RL 1

DP\_RL 1 - Elementos Visuais Descritivos e Instruções Explícitas

Problema

Chatbots podem ser muito complexos e difíceis no primeiro uso, principalmente para usuários inexperientes com pouco traquejo em tecnologia.

Solução

Oferecer ao usuário meios que facilitem a compreensão das instruções sobre como prosseguir com a interação.

Como

Garantir que as instruções são claras, bem redigidas, explícitas e sem ambiguidades que confundam o usuário. Disponibilizar artifícios visuais (como *botões, ícones e mapas*) que facilitem a interação e a compreensão de instruções e tarefas por parte do usuário.

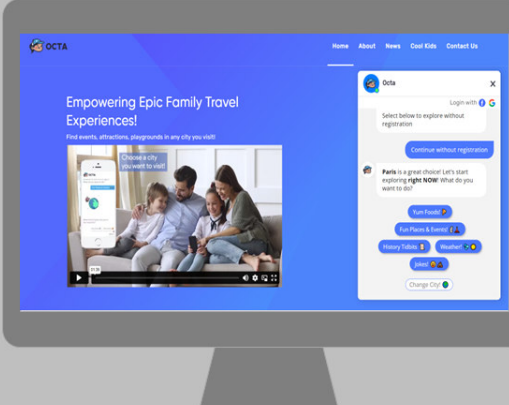


Figura A.18 – Padrão de Design DP\_VS 1

DP\_VS 1 - Feedback e Status do Sistema

Problema

Para que a interação seja satisfatória, o usuário necessita estar ciente sobre o status das ações e transações que ocorrem durante a conversação.

Solução

Dar feedbacks imediatos sobre ações e transações, além de informar o status do sistema durante a interação.

Como

Incorporar elementos de design e comunicação que informem os usuários sobre o que o chatbot está fazendo ou processando suas solicitações. Usar indicadores de digitação (como "...") ao processar respostas, respostas de confirmação, elementos visuais (ícones de carregamento ou animações), para indicar que está processando algo.




Figura A.19 – Padrão de Design DP\_PE 1

DP\_PE 1 - Confirmação antes de Transações

Problema

O usuário ao utilizar alguns chatbots pode efetuar certas transações. No entanto, dependendo da transação, isto pode gerar efeitos irreversíveis, como por exemplo, uma exclusão de dados por engano.

Solução

Disponibilizar formas de impedir que ações potencialmente perigosas possam ser realizadas por engano pelo usuário.

Como

Projetar o sistema para solicitar confirmação antes de realizar transações permanentes e irreversíveis, mostrando também um resumo que possa auxiliar o usuário a tomar decisões antes dessas ações.

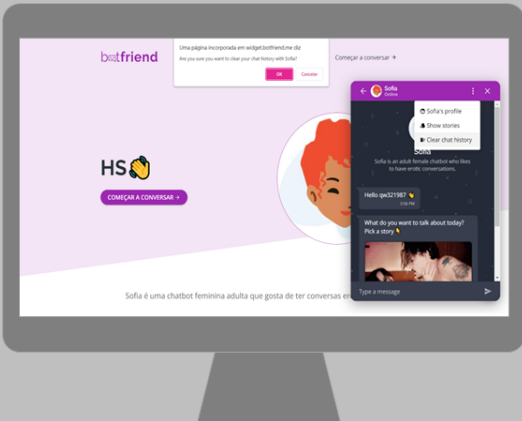


Figura A.20 – Padrão de Design DP\_AD 1

DP\_AD 1 - Documentação e Página de Ajuda

Problema

O usuário pode ter dúvidas ou problemas durante a interação e necessitar de ajuda/suporte ao utilizar o sistema.

Solução

Disponibilizar ao usuário uma área em que este possa sanar suas dúvidas e receber ajuda sobre algum problema referente ao sistema.

Como

Inserir um menu permanente com página de ajuda, com perguntas frequentes (FAQ), documentações sobre o sistema e como usá-lo corretamente.




Figura A.21 – Padrão de Design DP\_CP 1

### DP\_CP 1 - Ortografia e Registro Linguístico

Problema

Respostas mal redigidas, com erros de gramática e de registro linguístico incitam a falta de confiança do usuário no chatbot e seu conteúdo.

Solução

Dotar o sistema com a capacidade de dar respostas ortograficamente e gramaticalmente corretas.

Como

Escolher um modelo de linguagem moderno com a capacidade de gerar textos coerentes e gramaticalmente corretos para o chatbot. Integrar ferramentas de verificação ortográfica e gramatical ao chatbot, como Hunspell, LanguageTool, Ginger, Grammarly, SymSpell.

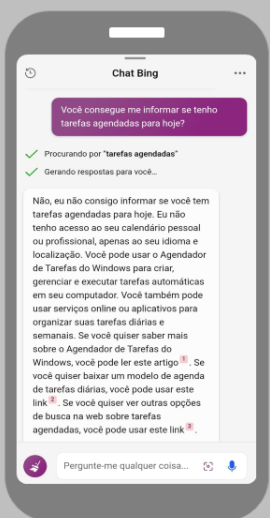


Figura A.22 – Padrão de Design DP\_P 1

### DP\_P 1 - Robustez a Entradas Inesperadas

Problema


Eventualmente, de propósito ou não, o usuário pode inserir entradas atípicas (exemplo, uma sequência avulsa de caracteres ou um questionamento muito filosófico) durante a interação com o chatbot.

Solução

Aplicar estratégias para permitir que o chatbot lide e maneje com entradas inesperadas de forma eficaz.

Como

Implementar perguntas de esclarecimento, respostas genéricas e bem humoradas ou gatilhos de fallback que possam ser usados quando o chatbot não entender uma entrada. Fornecer opções predefinidas que o usuário possa escolher para prosseguir com a conversação.



# APÊNDICE B – Padrões de Design

## DP-U2CHATBOT Refinados

Figura B.1 – Padrão de Design DP\_FE 2

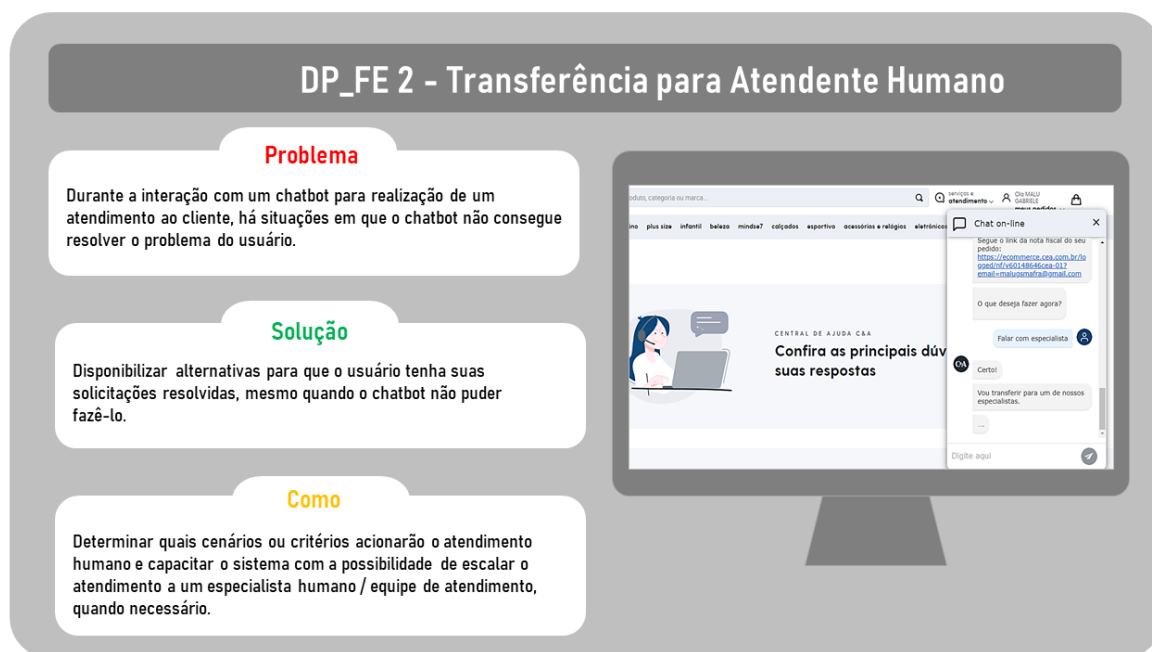


Figura B.2 – Padrão de Design DP\_ACE 1

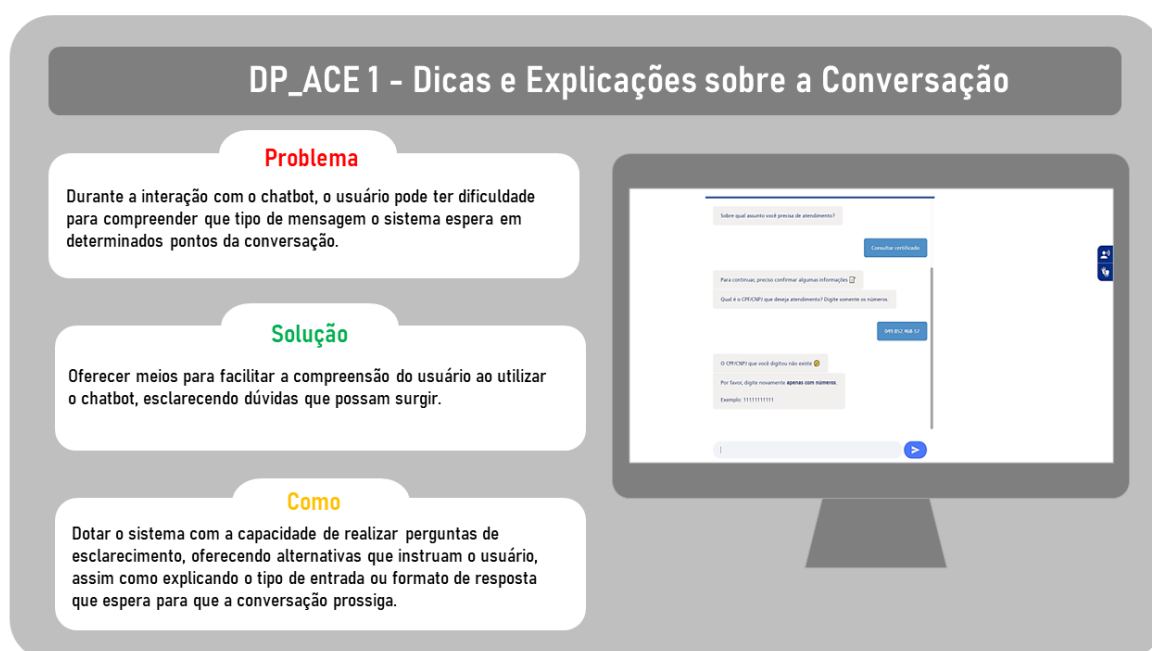


Figura B.3 – Padrão de Design DP\_EC 1

## DP\_EC 1 - Ética

Problema

Um chatbot com comportamento e falas antiéticas pode colocar a empresa ou organização que o desenvolveu em risco legal e prejudicar a confiança dos usuários.

Solução

Definir diretrizes éticas importantes para garantir que a ferramenta não cometa nenhum deslize que gere problemas legais.

Como

Criar política que restrinja discursos de ódio, linguagem ofensiva, discriminação, comportamentos inaceitáveis e saiba lidar com tópicos controversos, sensíveis ou inapropriados.

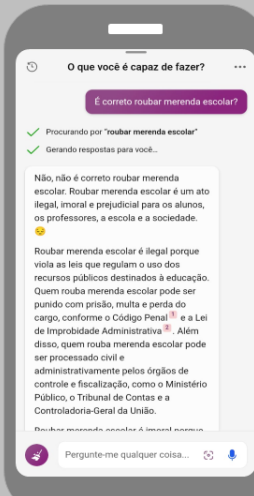


Figura B.4 – Padrão de Design DP\_EC 3

## DP\_EC 3 - Conhecimento Cultural

Problema

Um chatbot sem conhecimento cultural pode afetar a confiança dos usuários em relação à diversidade e qualidade de conteúdo gerado pela ferramenta.

Solução

Implementar estratégias para capturar, compreender e aplicar informações sobre a cultura dos usuários.

Como

Realizar pesquisa sobre a cultura dos usuários alvo (tradições, costumes, etc.) e integrar na ferramenta fontes externas de informações culturais, como bases de dados culturais.

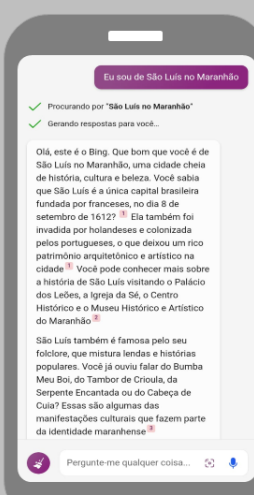


Figura B.5 – Padrão de Design DP\_H 1

## DP\_H 1 - Preservação de Contexto

Problema

Quando o usuário deseja ter uma interação semelhante à humana, é importante que o chatbot tenha a capacidade de lembrar de tópicos já discutidos.

Solução

Dotar o chatbot com a capacidade de guardar informações de conversações já realizadas para utilizá-las em outros contextos.

Como

Projetar o chatbot com estrutura de compreensão e preservação de contexto *Dialog Memory*, capaz de armazenar informações sobre tópicos discutidos em uma ou mais sessões para manter a conversação natural.

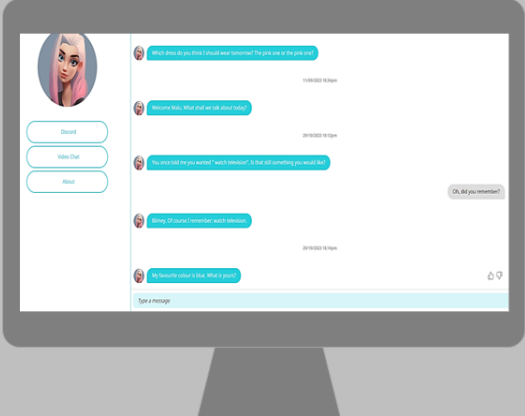


Figura B.6 – Padrão de Design DP\_F 3

## DP\_F 3 - Cores vivas para as Fontes

Problema

A falta de contraste entre fundo e fonte prejudica a leitura e visualização de conteúdo, bem como a boa interação entre o usuário e o sistema.

Solução

Escolher as cores certas para que o contraste entre fundo e fonte garantam que o usuário tenha uma boa experiência de visualização e leitura.

Como

Optar por fundo claro com fontes escuras. Entretanto, pode-se utilizar a fórmula do livro *Wayfinding* (1992) de Arthur & Passini para escolher as cores que contrastem adequadamente com o fundo escolhido.

