

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO INTERDISCIPLINAR DE NOVAS TECNOLOGIAS NA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA NA EDUCAÇÃO**

CLÓVIS DA SILVEIRA

**DESENVOLVIMENTO DE UM *FRAMEWORK* COM ESTRATÉGIAS DE
AVALIAÇÃO DE AGENTES CONVERSACIONAIS EDUCACIONAIS**

Porto Alegre

2022

CLÓVIS DA SILVEIRA

**DESENVOLVIMENTO DE UM *FRAMEWORK* COM ESTRATÉGIAS DE
AVALIAÇÃO DE AGENTES CONVERSACIONAIS EDUCACIONAIS**

Tese apresentada ao Programa de Pós-Graduação em Informática na Educação, do Centro Interdisciplinar de Novas Tecnologias na Educação, da Universidade Federal do Rio Grande do Sul, como requisito parcial para a obtenção do título de Doutor em Informática na Educação.

Orientadora: Prof.^a Dra. Liane Margarida Rockenbach Tarouco
Coorientador: Dr. Fabrício Herpich

Porto Alegre

2022

CIP - Catalogação na Publicação

Silveira, Clóvis da
DESENVOLVIMENTO DE UM FRAMEWORK COM ESTRATÉGIAS DE
AVALIAÇÃO DE AGENTES CONVERSACIONAIS EDUCACIONAIS /
Clóvis da Silveira. -- 2022.
203 f.
Orientadora: Liane Margarida Rockenbach Tarouco.

Coorientador: Fabrício Herpich.

Tese (Doutorado) -- Universidade Federal do Rio
Grande do Sul, Centro de Estudos Interdisciplinares em
Novas Tecnologias na Educação, Programa de
Pós-Graduação em Informática na Educação, Porto
Alegre, BR-RS, 2022.

1. Agentes Conversacionais. 2. Métricas de
Avaliação. 3. Chatbot. 4. Qualidade de Software. 5.
NBR-ISO IEC-25010. I. Margarida Rockenbach Tarouco,
Liane, orient. II. Herpich, Fabrício, coorient. III.
Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UFRGS com os
dados fornecidos pelo(a) autor(a).



SERVIÇO PÚBLICO FEDERAL
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO INTERDISCIPLINAR DE NOVAS TECNOLOGIAS NA EDUCAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA NA EDUCAÇÃO

**ATA SOBRE A DEFESA DE TESE DE DOUTORADO
CLÓVIS DA SILVEIRA**

Às nove horas do dia seis de julho de dois mil e vinte e dois, no endereço eletrônico <https://mconf.ufrgs.br/webconf/defesa-clovis-da-silveira> conforme a portaria 2291 de 17/03/2020 que suspende todas as atividades presenciais possíveis, nesta Universidade, reuniu-se a Comissão de Avaliação, composta pelos Professores Doutores: Milton Antonio Zaro, Eliane Pozzebon, Marcelo Augusto Rauh Schmitt para a análise da Defesa de Tese de Doutorado intitulada “**Desenvolvimento de Um Framework com Estratégias de Avaliação de Agentes Conversacionais Educacionais.**” do doutorando de Pós – Graduação em Informática na Educação Clóvis da Silveira sob a orientação da Prof.^a Dra. Liane Margarida Rockenbach Tarouco e coorientação do Prof. Dr. Fabrício Herpich. A Banca, reunida, após a apresentação e arguição, emite o parecer abaixo assinalado.

Considera a Tese Aprovada

sem alterações;

sem alterações, com voto de louvor;

e recomenda que sejam efetuadas as reformulações e atendidas as sugestões contidas nos pareceres individuais dos membros da Banca;

Considera a Tese Reprovada.

Considerações adicionais (a critério da Banca):

A tese e a defesa realizadas preenchem os requisitos para obtenção do título de Doutor em Informática na Educação. O texto é organizado de maneira adequada, com introdução, referencial teórico e metodologia claros. As conclusões são compatíveis com os resultados obtidos e também foi desenvolvido um produto tecnológico para que os desenvolvedores de ChatBots possam avaliá-los de forma automatizada.

Liane Margarida
Rockenbach
Tarouco

Assinado de forma digital
por Liane Margarida
Rockenbach Tarouco
Dados: 2022.07.06
18:06:24 -03'00'

Prof.^a Dra. Liane Margarida Rockenbach Tarouco
Orientadora

Prof. Dr. Fabrício Herpich
Coorientador

_____(videoconferência)_____
Prof. Dr. Milton Antonio Zaro
PPGIE

_____(videoconferência)_____
Prof.^a Dr.^a Eliane Pozzebon
UFSC/DEC

_____(videoconferência)_____
Prof. Dr. Marcelo Augusto Rauh Schmitt
IFRS/POA

*À Camila, pelo incentivo e apoio em todos os momentos.
À Isabelli, pelos momentos incríveis de alegria e descontração.*

AGRADECIMENTOS

A Deus, primeiramente, por ser a base das minhas conquistas, pela saúde, por me tornar perseverante e por sustentar o meu bom ânimo em todos os dias da minha vida.

À Orientadora, Professora Dra. Liane Margarida Rockenbach Tarouco, e ao Coorientador, Professor Dr. Fabrício Herpich, pelo conhecimento compartilhado, pela sabedoria em direcionar, cobrar e apontar o melhor caminho para o desenvolvimento deste estudo. Agradeço, ainda, pelo incentivo e, principalmente, pelo cuidado e sensibilidade com que conduziram este trabalho.

A todos os professores do Programa de Pós-Graduação em Informática na Educação, por terem contribuído para o meu desenvolvimento e conhecimento, e aos meus colegas de doutorado, pela parceria e pelo companheirismo. A vocês, o meu profundo respeito e reconhecimento.

Ao colégio e aos alunos do Curso Técnico em Informática para Internet, do Senac São Leopoldo, por apoiarem a investigação, para que a pesquisa pudesse ser concretizada.

Aos colegas de trabalho, pelas ideias e angústias compartilhadas.

A minha mãe, Maria de Lourdes, a minha avó Alvina, a minha família e aos meus amigos, pela compreensão nos momentos de ausência.

A minha amada esposa, Camila, pelas horas de estudo compartilhadas, por estar ao meu lado, pelo apoio e incentivo e, especialmente, por compartilhar dos meus sonhos.

A todas as pessoas que conviveram comigo nesse período de doutorado e que, de uma forma ou de outra, contribuíram para a realização deste sonho.

RESUMO

O crescimento das tecnologias de informação e comunicação tem demonstrado uma série de ferramentas desenvolvidas para as mais diversas finalidades, entre elas, as ferramentas que possibilitam o envio de mensagens instantâneas e proporcionam uma comunicação em tempo real, como é o caso dos agentes conversacionais, que permitem a participação e intervenção dos sujeitos, promovendo, dessa forma, a interação de um para um. Um agente conversacional mantém um diálogo com quem o utiliza, por meio de perguntas realizadas pelo estudante, mediante texto, simulando um tutor pedagógico que provê suporte ao ensino de conteúdo. De fato, avaliar e construir agentes conversacionais educacionais é um desafio, pois, atualmente, o uso de agentes demanda referências específicas para apoiar a avaliação de agentes conversacionais educacionais por conta de métricas de avaliação da qualidade de *software*. Nesse sentido, a pesquisa desenvolvida na presente tese resultou na construção de um *framework* com métricas de avaliação de *software*, passíveis de serem usadas na avaliação de *chatbots* em contexto educacional. O conjunto destas métricas foi pensado também como apoio para orientar a construção de um agente conversacional educacional. O *framework* envolveu métricas de avaliação de *software* da NBR-ISO IEC-25010 e métricas de trabalhos relacionados que abordam a avaliação de agentes conversacionais. Ademais, um aplicativo *web*, denominado *ChatAval*, foi desenvolvido com a finalidade de automatizar a definição das funcionalidades de categorias, métricas, avaliações e de apoiar a criação de questionário e entrega das avaliações com indicadores resultantes das respostas já calculadas automaticamente. O procedimento metodológico adotado neste estudo consiste em uma pesquisa exploratória e descritiva, caracterizada por procedimentos bibliográficos, documentais e estudo de caso. O estudo de campo envolveu alunos do Curso Técnico em Informática para Internet do Senac, unidade São Leopoldo-RS, com os quais foi trabalhado o uso dos agentes conversacionais METIS, ATENA e ALTEIA, que já vêm sendo estudados e desenvolvidos no contexto do Projeto AVATAR (Ambiente Virtual de Aprendizagem e Trabalho Acadêmico Remoto), da Universidade Federal do Rio Grande do Sul. O *framework* foi validado por um conjunto de especialistas. Neste trabalho, foram considerados especialistas os profissionais que atuam com Tecnologia da Informação e não têm vivências ou experiências na docência; professores de diversas áreas da educação, mas que usam as tecnologias no contexto de suas disciplinas; e profissionais formados nas áreas de Ciência da Computação, Sistemas de Informação, Licenciatura em Computação e outras formações, os quais atuam como docentes na área de Tecnologia da Informação. A coleta de dados envolveu três fases: levantamento do perfil dos sujeitos; atividades práticas de programação com os alunos, a fim de realizar um teste piloto nas funcionalidades do *ChatAval* a partir da avaliação dos agentes METIS, ATENA e ALTEIA; e a avaliação e validação das métricas que compõem o *framework* pelos especialistas. Os resultados das Fases 1 e 2 apontam que as funcionalidades do *ChatAval* foram validadas com sucesso. Já em relação ao resultado do *framework*, nenhum dos especialistas rejeitou ou sugeriu que fosse excluída uma determinada métrica, mas que fossem realizados pequenos ajustes em relação ao texto ou realocação para uma outra categoria, o que demonstra que as métricas com origem dos trabalhos relacionados, bem como as métricas da ISO/IEC 25010 atenderam, de forma satisfatória, a composição do *framework*.

Palavras-chave: Agentes Conversacionais. Métricas de Avaliação. *Chatbot*. Qualidade de *Software*. NBR-ISO IEC-25010.

ABSTRACT

The growth of information and communication technologies has demonstrated a series of tools developed for the most diverse purposes, among them, the tools that make it possible to send instant messages and provide real-time communication, as is the case of conversational agents, which allow the participation and intervention of subjects, thus promoting one-to-one interaction. A conversational agent maintains a dialogue with the user through questions asked by the student using text, simulating a pedagogical tutor that supports the teaching of content. Indeed, evaluating and building educational conversational agents is a challenge, as currently the use of agents demands specific references to support the evaluation of educational conversational agents due to *software quality assessment metrics*. In this sense, the research developed in this thesis resulted in the construction of a *framework* with *software evaluation metrics*, which can be used in the evaluation of chatbots in an educational context. The set of these metrics was also thought of as support to guide the construction of an educational conversational agent. The *framework* involved *software evaluation metrics* from NBR-ISO IEC-25010 and metrics from related works that address the evaluation of conversational agents. In addition, a *web* application called *ChatAval* was developed with the purpose of automating the definition of the functionalities of categories, metrics, evaluations, as well as to support the creation of a questionnaire and delivery of evaluations with indicators resulting from the answers already calculated automatically. The methodological procedure adopted in this study consists of an exploratory and descriptive research characterized by bibliographic and documentary procedures and a case study. The field study involved students from the Technical Course in Computer Science for the Internet at Senac, São Leopoldo-RS unit, with whom the use of Conversational Agents METIS, ATENA and ALTEIA, which have already been studied and developed in the context of the AVATAR Project (Virtual Learning Environment and Remote Academic Work), from the Federal University of Rio Grande do Sul. The *framework* was validated by a group of experts. In this work, professionals who work with Information Technology and do not have teaching experiences were considered specialists; teachers from different areas of education, but who use technologies in the context of their disciplines; and professionals trained in the areas of Computer Science, Information Systems, Degree in Computing and other training, who act as professors in the area of Information Technology. Data collection involved three phases: survey of the subjects' profile; practical programming activities with the students, in order to carry out a pilot test on *ChatAval*'s functionalities based on the evaluation of the agents METIS, ATENA and ALTEIA; and the evaluation and validation of the metrics that make up the *framework* by specialists. Results from Phases 1 and 2 indicate that the *ChatAval features* were successfully validated. Regarding the result of the *framework*, none of the experts rejected or suggested that a certain metric be excluded, but rather that there were small adjustments in relation to the text or reallocation to another category, which demonstrates that the metrics with origin of the related works, as well as the metrics of ISO/IEC 25010 satisfactorily met the composition of the *framework*.

Keywords: Conversational Agents. Evaluation Metrics. *Chatbot*. *Software Quality*. NBR-ISO IEC-25010.

LISTA DE FIGURAS

Figura 1.2 - Estrutura da tese.....	27
Figura 2.1 - Exemplos de código em AIML	31
Figura 2.2 - Arquitetura de um <i>chatbot</i> baseado no sistema ALICE	32
Figura 2.3 – Tela da Agente METIS	38
Figura 2.4 - Exemplo de integração de <i>chatbot</i> e AVATAR	40
Figura 2.5 - ISO/25010: 2011.....	53
Figura 2.6 - Características e subcaracterísticas e atributos de qualidade de <i>software</i>	60
Figura 4.1 – Métricas de avaliação para agentes conversacionais	95
Figura 4.2 – Manutenibilidade e Portabilidade: Características da NBR-ISO IEC-9126-1 que não aparecem nos estudos relacionados	97
Figura 4.3 - Percurso das métricas para desenvolvimento do <i>framework</i>	98
Figura 4.4 – Organograma das tecnologias envolvidas no protótipo	108
Figura 4.5 - Tela para responder um formulário sobre um agente conversacional	114
Figura 4.6 - Tela com os dados de resultado de formulários.....	114
Figura 4.7 - Tela com mais detalhes do resultado de um formulário	115
Figura 5.1 – Distribuição da turma presencial x on-line	119
Figura 5.2 – Aulas em diferentes espaços do Senac.....	120
Figura 5.3 – Ambiente que os alunos tinham acesso para gerenciar as classes	121
Figura 5.4 – Trecho de código elaborado por aluno.....	121
Figura 5.5 – Exemplo de inserção do agente conversacional no <i>site</i> desenvolvido.....	122
Figura 5.6 – Compilação de diferentes sites com o agente conversacional integrado	123
Figura 5.7 – Exemplo de <i>site</i> hospedado no <i>GitHub</i> pelo aluno.....	123
Figura 5.8 – Configuração do cadastro das perguntas no <i>ChatAval</i>	127
Figura 5.9 – Questionário gerado automaticamente pelo <i>ChatAval</i>	128
Figura 5.10 - Avaliação das métricas para a avaliação do <i>software</i>	129
Figura 5.11 – Percurso para a composição das métricas do <i>framework</i>	133
Figura 5.12 – Grupo de especialistas que validaram o <i>framework</i>	135
Figura 5.13 – Resultados métricas categoria Funcionalidade	136
Figura 5.14 - Resultados métricas categoria Eficiência	138
Figura 5.15 - Resultados métricas categoria Usabilidade	140
Figura 5.16 - Resultados métricas categoria Confiabilidade.....	142
Figura 5.17 - Resultados métricas categoria Segurança.....	144

Figura 5.18 - Resultados métricas categoria Manutenção.....	145
Figura 5.19 – Resultados métrica categoria Portabilidade	146
Figura 5.20 – Resultados métricas categoria Satisfação	148
Figura 5.21 – Resultados métricas categoria Cobertura de Contexto	150
Figura 5.22 – Resultados métricas categoria Controle de Usuários.....	151
Figura 5.23 – Tela inicial do <i>ChatAval</i> com as categorias validadas pelos especialistas	159
Figura 5.24 – Tela <i>ChatAval</i> com parte das métricas (Subcategorias).....	160
Figura 5.25 – Quantidade de métricas (subcategorias) para cada categoria.....	161
Figura 5.26 - Cadastro de um formulário no <i>ChatAval</i> com as métricas validadas	161

LISTA DE TABELAS

Tabela 2.1 - Qualidade em uso	57
Tabela 3.1 - Organização trimestres e conteúdos	88
Tabela 3.2 – Organização dos conteúdos desenvolvidos na Fase 2.1	90
Tabela 3.3 - Organização dos dias e horários desenvolvidos na Fase 2.2.....	93
Tabela 4.1 – Comparação de métricas de avaliação para agentes conversacionais x métricas da NBR-ISO IEC-9126-1 e NBR-ISO IEC-25010	96
Tabela 4.2 – Compilação das métricas ISO/IEC 9126, 25010, Figura 4.1 e trabalhos relacionados	98
Tabela 4.3 – Categorias, subcategorias e perguntas	104
Tabela 4.4 – Pesos da Tabela <i>Likert</i>	116
Tabela 5.1 – Percurso para a construção e avaliação dos agentes	125
Tabela 5.2 – Compilação das métricas utilizadas e analisadas pelos especialistas	133
Tabela 5.3 – Transcrição das avaliações dos especialistas - categoria Funcionalidade	137
Tabela 5.4 - Transcrição das avaliações dos especialistas - categoria Eficiência	139
Tabela 5.5 - Transcrição das avaliações dos especialistas - categoria Usabilidade	140
Tabela 5.6 - Transcrição das avaliações dos especialistas - categoria Segurança.....	144
Tabela 5.7 – Transcrição das avaliações dos especialistas - categoria Manutenção	145
Tabela 5.8 – Transcrição das avaliações dos especialistas - categoria Portabilidade.....	147
Tabela 5.9 – Transcrição das avaliações dos especialistas - categoria Satisfação	148
Tabela 5.10 – Transcrição das avaliações dos especialistas – cat Cobertura de Contexto.....	150
Tabela 5.11 – Transcrição das avaliações dos especialistas – cat Controle de Usuários	151
Tabela 5.12 – Compilação das métricas ISO/IEC 9126, 25010, Figura 4.1 e trabalhos relacionados após a avaliação e validação dos especialistas	152
Tabela 5.13 – Mapeamento das métricas para o público do <i>ChatAval</i>	156

LISTA DE QUADROS

Quadro 1 - Atributos de qualidade de <i>chatbots</i> e agentes conversacionais	63
--	----

LISTA DE ABREVIATURAS E SIGLAS

AGATA	<i>Automatic Generation of AIML from Text Acquisition</i>
AIML	<i>Artificial Intelligence Markup Language</i>
ALICE	<i>Artificial Linguistic Internet Computer Entity</i>
ALTEIA	Agente Lexical de apoio a Tarefas Educacionais usando Inteligência Artificial
ATENA	Agente Tutor para Ensino e Navegação no Ambiente
AVA	Ambiente Virtual de Aprendizagem
AVATAR	Ambiente Virtual de Aprendizagem e Trabalho Acadêmico Remoto
CSS	<i>Cascading Style Sheets</i>
EAD	Educação a Distância
FAQ	<i>Frequently Asked Questions</i>
HTML	<i>HyperText Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
IA	Inteligência Artificial
IEC	<i>International Electrotechnical Commission</i>
ISO	<i>International Organization for Standardization</i>
METIS	Mediadora de Educação em Tecnologia Informática e Socializadora
NBR	Denominação de norma da Associação Brasileira de Normas Técnicas (ABNT)
NPC	<i>Non-Player Character</i>
PPGIE	Programa de Pós-Graduação em Informática na Educação
SGBDR	Sistema de Gestão de Bases de Dados Relacionais
TIC	Tecnologia de Informação e Comunicação
UFRGS	Universidade Federal do Rio Grande do Sul
ZDP	Zona de Desenvolvimento Proximal

SUMÁRIO

1 INTRODUÇÃO.....	16
1.1 JUSTIFICATIVA E MOTIVAÇÃO	16
1.2 PROBLEMA DE PESQUISA	25
1.3 OBJETIVOS DA PESQUISA	25
1.3.1 Objetivo geral.....	26
1.3.2 Objetivos específicos.....	26
1.4 ESTRUTURA DA TESE.....	26
2 REFERENCIAL TEÓRICO	29
2.1 AGENTES CONVERSACIONAIS: CARACTERÍSTICAS E FUNCIONALIDADES	29
2.1.1 Agente METIS	37
2.1.2 Agente ATENA	39
2.1.3 Agente ALTEIA.....	41
2.1.4 Tendências em <i>chatbots</i>	43
2.2 MÉTRICAS DE AVALIAÇÃO DE <i>SOFTWARES</i>	46
2.2.1 Métricas da ISO/IEC 9126-1	47
2.2.2 Métricas de qualidade externas.....	50
2.2.3 Métricas de qualidade internas	51
2.2.4 Métricas da ISO/IEC 25010:2011	52
2.2.5 Qualidade em uso	57
2.2.6 Escolha das métricas de avaliação de <i>softwares</i>	58
2.3 TRABALHOS RELACIONADOS	60
2.3.1 Estudo 1 - Projeto e avaliação do agente conversacional SARA	61
2.3.2 Estudo 2 - Avaliando a qualidade dos <i>chatbots</i>	63
2.3.3 Estudo 3 - Avaliação de agentes conversacionais incorporados.....	64
2.3.4 Estudo 4 - Conversas reais com inteligência artificial	66
2.3.5 Estudo 5 - Uma avaliação do sistema do agente conversacional	67
2.3.6 Estudo 6 - Métricas para o <i>Chatbot Analytics in 2019</i>	69
2.3.7 Estudo 7 – <i>Chatbot Analytics</i> : as métricas essenciais que é preciso acompanhar	71
2.3.8 Estudo 8 - Medindo a eficácia do <i>Chatbot</i>	72
2.3.9 Estudo 9 - <i>ChatEval</i> : Uma ferramenta para avaliação do <i>Chatbot</i>	73
2.3.10 Estudo 10 - Análise de expressões lexicais.....	74
2.3.11 Estudo 11- Métricas para o desenvolvimento de agentes conversacionais.....	74
2.3.12 Estudo 12 - Conversar com agente humano.....	76
2.3.13 Outros trabalhos de avaliação de <i>chatbot</i>	76
2.4 COMPARAÇÃO DAS MÉTRICAS DOS TRABALHOS SELECIONADOS	79
3 PROCEDIMENTOS METODOLÓGICOS	85
3.1 CENÁRIO DO ESTUDO DE CASO	87
3.2 PERFIL DEMOGRÁFICO.....	88
3.3 FASES DA PESQUISA	89
3.3.1 Fase 1 - Perfil dos alunos do Curso Técnico em Informática	89
3.3.2 Fase 2 - Conhecendo e avaliando os agentes Aonversacionais	90
3.3.3 Fase 3 - Participação dos Especialistas.....	93

4 PERCURSO DE CONSTRUÇÃO DO <i>FRAMEWORK</i>	95
4.1 <i>CHATAVAL</i> : APLICATIVO <i>WEB</i> PARA AUTOMATIZAR O <i>FRAMEWORK</i>	107
5 RESULTADOS	118
5.1 RESULTADOS DA FASE 1 - PERFIL DOS ALUNOS DO CURSO TÉCNICO EM INFORMÁTICA.....	118
5.2 RESULTADOS DA FASE 2 - CONHECENDO E AVALIANDO OS AGENTES CONVERSACIONAIS.....	119
5.2.1 Resultados Fase 2.1 - Práticas de programação	119
5.2.2 Resultados Fase 2.2 - Experiências dos alunos com os agentes conversacionais	124
5.3 RESULTADOS DA FASE 3 - PARTICIPAÇÃO DOS ESPECIALISTAS.....	132
5.3.1 Apresentação das métricas analisadas pelos especialistas	133
5.3.2 Apresentação do perfil dos especialistas	134
5.3.3 Resultados da análise e validação das métricas	135
5.3.4 Público do <i>ChatAval</i>	156
5.4 VALIDAÇÃO DO <i>FRAMEWORK</i> COM O APLICATIVO <i>CHATAVAL</i>	158
6 CONCLUSÃO	163
6.1 TRABALHOS FUTUROS.....	168
6.2 CONTRIBUIÇÕES	169
REFERÊNCIAS	170
APÊNDICES	177

1 INTRODUÇÃO

1.1 JUSTIFICATIVA E MOTIVAÇÃO

A ascensão das Tecnologias de Informação e Comunicação (TICs) tem demonstrado relevantes aportes para os procedimentos de ensino e aprendizagem. Entre as contribuições, destaca-se uma variedade de ferramentas desenvolvidas para as mais diversas finalidades, que proporcionam comunicação em tempo real, como é o caso dos agentes conversacionais, os quais permitem a participação e intervenção dos sujeitos, promovendo a interação de um para um.

Nesse sentido, as TICs podem contribuir, de maneira significativa, a partir do uso de *chatbots*. O termo *chatbot* foi proposto por Michael Maulding, em 1994, para referenciar os sistemas de conversação com usuários humanos, por meio de linguagem natural (SHAWAR; ATWELL, 2007). Os *chatbots*, também denominados agentes conversacionais, *chatterbots*, agentes conversacionais ou sistemas de diálogo, são recursos que reproduzem conversa entre pessoas por meio de uma comunicação textual entre um utilizador humano, que disponibiliza a entrada, e o agente que retorna, fornecendo respostas ou elaborando novas perguntas. Cabe salientar que, neste estudo, optou-se por utilizar os termos *chatbot* e agentes conversacionais como sinônimos.

Conforme Wallace (2003), desenvolvedor da solução para dar suporte à implementação de *chatbots*, no caso do *software* do *chatbot Alicebot*, o serviço a ser provido compreende a consulta enviada e a pesquisa em uma estrutura, buscando resposta para mandar ao usuário. A pesquisa inclui elementos da consulta (palavras-chave) e seleciona a melhor alternativa de resposta, com base em alguns critérios. A base de conhecimento que apoia o funcionamento deste *chatbot* é estruturada pelo uso de uma linguagem de marcação chamada AIML (*Artificial Intelligence Markup Language*). Esta linguagem permite configurar e projetar o retorno do *chatbot*, usando recursos elevados, sem precisar haver familiaridade com a linguagem de programação. Em sua pesquisa intitulada “The Elements of AIML Style”, Wallace (2003) apresenta os recursos iniciais para se desenvolver um *chatbot*, por meio da linguagem AIML, com vistas às possibilidades de perguntas e resoluções, apresentando as *tags* (marcações) da linguagem.

Contribuições de Satu, Parvez e Mamun (2015) referem alguns *chatbots* na linguagem AIML com propósitos distintos. Entre as finalidades listadas nesta pesquisa, estão as aplicações relativas ao *e-learning*, ao governo eletrônico, ao modelo de diálogo, ao humor, ao

gerenciamento de rede, à arquitetura modular adaptativa, entre outros. Nestas finalidades, os *chatbots* não apenas fornecem recursos importantes, como também interagem com os clientes, podendo apresentar soluções aos seus problemas por conta da máquina de inferência que consulta a base de conhecimento escrita em AIML. Em tais casos, o *chatbot* substitui a ação humana no atendimento de perguntas que costumam ser apresentadas. O uso deste em lugar de FAQ (*Frequently Asked Questions*) é um dos mais populares, sendo viável obter respostas por meio de diálogos com a máquina, usando linguagem natural (DALE, 2016).

Leonhardt et al. (2005) apontam como características dos *chatbots* a capacidade de tornar a relação de alunos mais contínua com o computador, que permanece 24 horas disponível. Esta interação é importante como método educacional, de acordo com Bloom (1984) em seu estudo “The 2 sigma problem”. Nele, o autor estabelece uma diferença de desempenho entre alunos que receberam atendimento individualizado e os que foram atendidos em sala de aula. Vale ressaltar que a assistência individual por humanos tem alto valor associado, razão pela qual Bloom e outros pesquisadores perscrutaram artifícios, usando tecnologia, a fim de deixar a atividade de aprendizagem mais conversacional. Bloom instigou pesquisadores a achar formas de instrução grupal tão eficientes quanto as aulas individuais, pois a tecnologia pode representar soluções de tutoria, sem o alto custo de fornecer um tutor humano para cada aluno.

Contar com um tutor humano por 24 horas, durante sete dias por semana, para realizar tarefas de interação é inviável financeiramente. No entanto, soluções automatizadas, como é o caso dos agentes conversacionais, possibilitam, via Internet, tornar tais recursos disponíveis ininterruptamente, podendo prover soluções a menor custo. Nesse sentido, o *chatbot* é uma resposta no quesito dos tutores inteligentes. Apesar de ser um tanto limitado, tem capacidade para aprimorar o potencial acadêmico dos discentes, proporcionando um recurso que aparenta o que seria fornecido por um tutor individual.

De fato, o uso de *chatbots* no meio educativo aparece de forma muito significativa, considerando-se que, em muitos estudos, são apresentados como relevantes, por revelarem resultados positivos. Moraes, Santos e Redecker (2016) apresentam uma experiência em sala de aula com a utilização de *chatbots* como método de aprendizado em uma turma de Inteligência Artificial do curso de Sistemas de Informação. Conclui-se, a partir disso, que houve um maior comprometimento dos discentes nas atividades propostas, além de ganhos em suas aprendizagens.

Ainda visando a algo positivo quanto ao uso do *chatbot* no meio educativo, Santos, Roland e Silva (2019) apresentam esforços direcionados para ensinar atividades de

aprendizagem, utilizando um agente conversacional denominado *Dory*. Este tem foco na recomendação de aplicativos educacionais relacionados às matérias escolares, com o propósito de complementar a formação dos estudantes. Deste modo, professores do ensino fundamental participaram do processo de criação das atividades que envolveram a avaliação e a seleção dos aplicativos a serem recomendados. As conclusões do estudo apontaram o potencial de aplicabilidade do *chatbot* na conjuntura em questão.

Paschoal, Chicon e Falkembach (2017) implementaram o *chatbot Ubibot*, que foi desenvolvido com o propósito de ajudar estudantes do curso superior em Ciências da Computação, especificamente da matéria de Engenharia de *Software*. A principal funcionalidade do *Ubibot* é auxiliar nos estudos dos alunos, de modo a adaptar suas mensagens, considerando o contexto do estudante, com a intenção de oferecer acesso propício à aprendizagem ubíqua. A pesquisa tem o intuito de apresentar a concepção, a implementação e a análise de um agente conversacional integrado ao Moodle. Os resultados do estudo demonstram que o *chatbot* conseguiu satisfazer os objetivos esperados, já que contribuiu para a aprendizagem dos estudantes. Além disso, alguns apontamentos dos estudantes revelam a necessidade de ampliação do total de bases de conhecimento e tratamento de algumas sentenças de interação.

Com base nos relatos dessas pesquisas, verifica-se que a intensificação da utilização de agentes conversacionais que servem de apoio na aprendizagem em diversos contextos educacionais. Também se observa que praticamente na unanimidade dos estudos apresentados, o *chatbot* sempre aparece com resultados positivos quanto ao amparo ao aprendizado, indiferentemente das especificidades em que foi desenvolvido, das tecnologias utilizadas, ou até mesmo da forma de utilização, como complemento em um Ambiente Virtual de Aprendizado, já em uso nas pesquisas dos autores referidos.

Fica claro, portanto, que os agentes conversacionais são importantes no contexto educacional ao auxiliarem os estudantes por intermédio de diálogos, em que se pode apoiar e elucidar questionamentos de uma maneira bastante ágil, cordial e eficaz. Ressalta-se que o *chatbot* fica à disposição dos alunos 24 horas por dia, durante toda a semana, a menos que ocorra um problema relacionado à Internet, ou ao servidor em que o *chatbot* está hospedado. Isso é bastante relevante quando comparado a uma turma com aulas presenciais composta de muitos alunos com apenas um professor para esclarecer dúvidas em um período bem pequeno (uma ou duas vezes por semana).

Vale registrar que os agentes conversacionais desenvolvidos para serem utilizados neste meio têm sua referência de construção no auxílio de um professor ou especialista no assunto em um contexto educacional específico (CLEMENTE, 2016; SANTANA, 2017).

No geral, percebe-se que diferentes estudos científicos relacionados a agentes conversacionais fazem questionários e apresentam a opinião dos alunos relativa à experiência entre eles e o *chatbot*. Por meio deles, busca-se entender se o agente conversacional apresentou respostas às perguntas dos discentes, se foi prestativo durante os diálogos etc. Há, portanto, uma preocupação em construir um *chatbot* para auxiliar na aprendizagem, mais especificamente, para responder corretamente aos diálogos durante interação com os alunos.

Nesse sentido, cabe recuperar o conceito sociointeracionista de Vygotsky (1978) sobre o fato de o aprendizado consistir em um procedimento social, já que o desenvolvimento do aluno sofre influência de seus pares no processo de evolução cognitiva. A Zona de Desenvolvimento Proximal (ZDP) é o conceito base desta teoria, sendo representada como a adjacência entre o nível de desenvolvimento real, que se costuma conceituar por meio da decifração individual de problemas, e o nível de desenvolvimento potencial, fomentado a partir da resolução de problemas por intermédio de um adulto ou em conjunto com outros mais capacitados. Baseando-se neste conceito, as interações sociais passam a ser consideradas importantes para a elaboração de conhecimento por parte dos alunos, em especial, as que permitem a conversa, a colaboração, a troca de informações mútuas e o embate de diferentes pontos de vista.

De acordo com Dillenbourg (1999), os pares não aprendem porque são dois, mas porque realizam atividades que estimulam mecanismos de aprendizagem particulares. Isso inclui atribuições ou instrumentos realizados de forma individual, uma vez que a cognição própria não é abolida em interação entre pares. Além disso, a relação entre os indivíduos proporciona atividades extras, como explicações, desacordos e regulação mútua, as quais desencadeiam mecanismos cognitivos complementares.

Neste âmbito, a aprendizagem entre os pares também merece atenção. A concepção dos conteúdos para materiais de ensino, segundo Mayer (2001), envolve: 1) esforço que contribui para a elaboração dos esquemas, derivada da natureza do conteúdo a ser aprendido; 2) atividades que beneficiam o objetivo da aprendizagem; e 3) esforço requerido para processar o conteúdo em decorrência de sua estruturação e organização, derivada do material instrucional.

Adicionalmente, a presente pesquisa considera a noção de aprendizagem multimídia proposta por Mayer e Moreno (2003), que trata da aprendizagem a partir de palavras e imagens, cujas palavras incluem o discurso oral e escrito, e as imagens, que podem ser estáticas, como

as fotos ou dinâmicas em vídeos. Assim, o princípio multimídia enuncia que as pessoas aprendem mais com a utilização de imagens e palavras, do que somente com o emprego isolado de um ou de outro. Neste sentido, Araujo et al. (2014) acreditam que ambientes educacionais que envolvem recursos multimídia propiciam situações que facilitam a elaboração de significados, à medida que oferecem ao aprendiz ferramentas poderosas para utilizar em uma atividade individual ou colaborativa.

Os diversos cenários integrados de aprendizagem exigem um gerenciamento que Dillenbourg (2013) define como orquestração:

Orquestração refere-se à maneira como um professor gerencia, em tempo real, atividades de várias camadas em um contexto de múltiplas restrições. Muitos cenários pedagógicos integram atividades individuais, como, por exemplo, leitura, trabalho em equipe (exemplo, solução de problemas) e atividades em toda a classe (por exemplo, palestras). Algumas dessas atividades são baseadas em computadores, mas outras não; alguns estão face a face enquanto outros estão on-line. Essa integração pedagógica é espelhada pela integração técnica de diferentes ferramentas (simulações, testes, *wikis* etc.) distribuídas em *laptops*, sensores, *tablets*. Estes cenários integrados exigem formas de gerenciamento conhecidas como orquestração (DILLENBOURG, 2013, p. 1).

Destaca-se, portanto, que a criação de cenários eficazes de suporte à aprendizagem é e continuará sendo relevante, e a utilização de *chatbots* para este fim constitui uma importante alternativa. A aplicabilidade de um agente conversacional pode tornar mais natural a procura de referências e o elucidar de dúvidas e questões, resultando em um elemento de apoio mais alinhado aos princípios de aprendizagem multimídia de Mayer e Moreno (2003) e às orientações relativas ao *design* instrucional, referidos por Dillenbourg (2013), que favoreçam a aprendizagem.

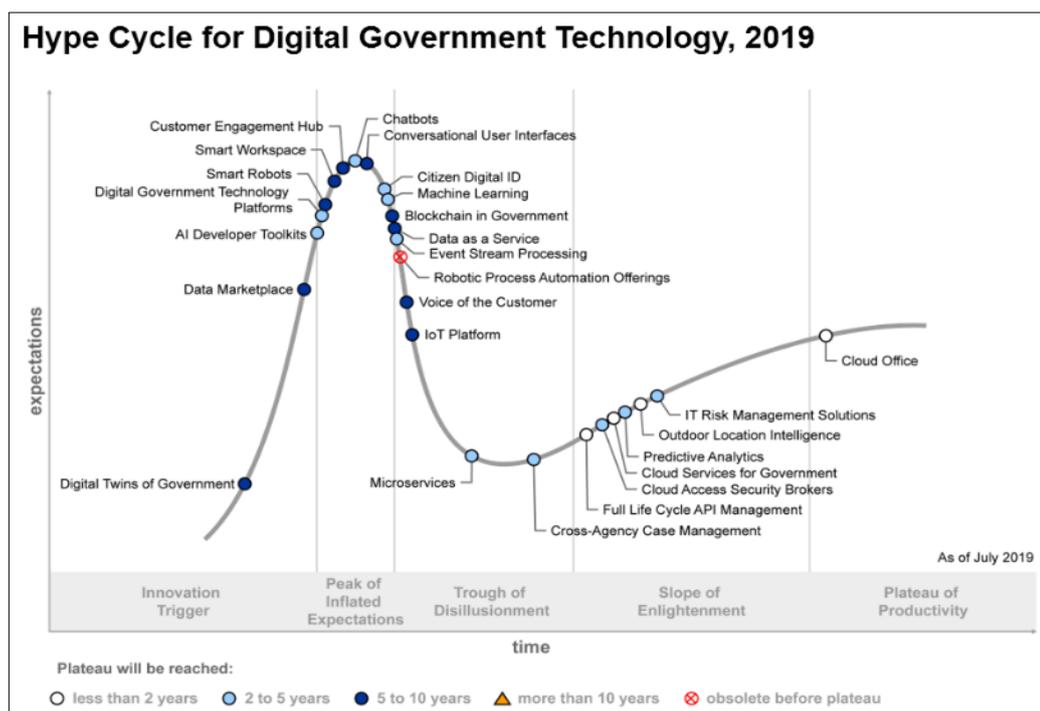
Estudos de Harsley et al. (2017) e Cho e Lim (2017) apontam que a cooperação, comparada com o esforço individual, costuma resultar em conquistas significativas, maior tempo de retenção do conteúdo e produção de pensamento crítico. Além disso, colabora com o aumento da criatividade na solução de problemas e, em um maior grau, desenvolve a persistência do estudante, reduzindo o tempo necessário para o aprendizado.

No geral, as interações são estabelecidas entre pares humanos, tais como: colegas, tutores e professores. No entanto, existem casos em que a interação entre os pares (estudantes e professores) é escassa, difícil ou até mesmo inviável. Em cursos de Educação a Distância (EAD), por exemplo, limitações geográficas, administrativas e financeiras não permitem que haja condições de dispor de um interlocutor para que o desempenho colaborativo ocorra sempre que for conveniente para o discente. Mesmo em cursos presenciais, há algumas limitações na

interação, quando, por exemplo, um professor pode não dar conta de atender a todas as dúvidas dos alunos no período da aula, salientando-se que os estudantes podem ter dúvidas a qualquer momento. Por essa razão, ter um meio para esclarecer essas dúvidas pode ser muito útil para o desenvolvimento do aprendizado de um aluno.

Diante do exposto, é possível verificar a crescente utilização do *chatbot* no relatório *Hype Cycle for Digital Government Technology*, de 2019 (GARTNER'S, 2019), ciclo anual de perspectivas sobre as tecnologias, no qual é apontado o crescimento de *chatbots* no intervalo de dois a cinco anos e interfaces de usuários conversacionais com o utilizador no período de cinco a dez anos. A Figura 1.1 resume as expectativas em relação às tecnologias em desenvolvimento e aos prazos de amadurecimento esperados.

Figura 1.1 - *Hype Cycle for Digital Government Technology*



Fonte: Gartner (2019).

Apesar de os *chatbots* apresentarem tendência de crescimento (GARTNER, 2019) e vantagens potenciais, também é possível observar insuficiências na percepção da linguagem natural. Além disso, a análise relacionada ao retoque é contínua, conforme apontam estudos como os de Dale (2017), Abushawar e Atwell (2015), Fryer et al. (2017), Mou e Xy (2017). Assim, a inserção de um *chatbot* denota inúmeros incitamentos, tais como: dificuldade de colocar informações na base de dados, concepção das muitas formas que o utilizador pode

desfrutar para elaborar um questionamento e expansão da chance de retorno para além do conteúdo da base de conhecimento, incluindo outras bases externas e um método semântico, que proporciona encontrar conceitos correspondentes ou relacionados.

Entre os desafios dos *chatbots*, encontra-se a possibilidade de aumentar a base de conhecimento, com a finalidade de se adquirir uma melhor interação entre os estudantes e o agente conversacional. Na visão de Wallace (2003), é preciso uma grande quantidade de códigos AIML para um *chatbot* ter um bom funcionamento e, conseqüentemente, de muito tempo e desempenho para a elaboração do código, levando-se em conta o trabalho constante de criação de códigos novos para aumentar a base de diálogos do *chatbot* (WALLACE, 2003). Para que se tenha uma base de conhecimento consistente, McNeal e Newyear (2013) sugerem que sejam necessárias, aproximadamente, 60.000 categorias, pois deve-se considerar que cada pergunta ou conceito, na base de conhecimento, requer múltiplas categorias para corresponder às permutações da questão e garantir uma resposta adequada ou correta.

Neste sentido, Krassmann et al. (2017) desenvolveram o *FastAIML*¹ ao observarem a dificuldade de se conseguir dispositivos de autoria, a fim de auxiliar desenvolvedores, docentes e pesquisadores da área da Educação a inserir novas informações na base de conhecimento de *chatbots* educacionais. A finalidade do *FastAIML* é automatizar a criação de arquivos no estilo de marcação AIML. Assim, por intermédio de uma interface simples, qualquer pessoa pode digitar uma possível pergunta que será realizada pelos estudantes e digitar quais são as possíveis respostas. A partir disso, o *FastAIML* vai gerar os códigos já na sintaxe da expressão de marcação AIML automaticamente, que poderá ser salva e enviada à base de conhecimento de agentes conversacionais da área da educação.

Também com o propósito de atuar diante dos desafios em aumentar a base de conhecimento entre *chatbots* e estudantes, Silveira (2019) desenvolveu a *EasyAIML*, um gerador de códigos AIML *desktop*. Com este *chatbot*, os códigos são criados de forma automatizada, por meio de *frameworks* com as *tags* necessárias mediante entradas do usuário. O *EasyAIML* é uma aplicação *desktop* desenvolvida em Java que permite ao usuário elaborar códigos automaticamente, usando campos que serão preenchidos conforme o que for desejado ensinar ao *chatbot*. Após preencher os campos e mandar o programa criar o código AIML, basta clicar em um botão para conceber o código AIML e retorná-lo ao usuário com opções de salvar o código gerado no formato *.aiml*. O objetivo principal do *EasyAIML* é reduzir o tempo

¹ O *FastAIML* está disponível em: <<http://avatar.cinted.ufrgs.br/fastaiml/>>.

utilizado, programando em AIML, e facilitar o desenvolvimento de escrita de códigos por pessoas que não tenham conhecimentos em programação, contribuindo, dessa forma, para o aumento da base de conhecimento de agentes conversacionais.

Ainda com o propósito de ajudar os usuários a elaborarem mais rapidamente as informações para a base de dados dos agentes conversacionais, Krassmann e Flach (2018) desenvolveram o AGATA (*Automatic Generation of AIML from Text Acquisition*). Este, assim como o *FastAIML* e o *EasyAIML*, consiste em um sistema de criação de base de conhecimento em AIML, capaz de criar automaticamente arquivos na linguagem de marcação AIML, mas com o diferencial de criar a base de conhecimento a partir da investigação e tratamento de *corpus* linguístico. Sua funcionalidade compõe a procura de palavras-chave que podem conter termos simples ou compostos a partir de um repertório apontado pelo utilizador.

Cumprе salientar que, ao desenvolver o AGATA, Krassmann e Flach (2018) deixaram uma contribuição para solucionar os desafios para aumentar os dados na base de conhecimento de agentes conversacionais de forma descomplicada e ágil, visto que os códigos são criados automaticamente na linguagem AIML. A maior contribuição dos autores está relacionada ao agente conversacional poder adquirir seu conhecimento a partir do que aparece em um livro de história, física, português, em artigos científicos, bem como em outro documento que contenha dados singulares sobre determinado assunto e utilizá-los para futuros diálogos com os estudantes.

Percebe-se que em diversos estudos relacionados a *chatbots* existe uma ausência, no que se refere aos parâmetros e métricas para a avaliação. É preciso, então, que, além de responder corretamente as questões dos alunos, sejam investigadas quais funcionalidades são necessárias para aprimorar o desempenho dos agentes conversacionais, a fim de garantir sucesso (CLEMENTE, 2016; SANTANA, 2017; PASCHOAL; CHICON; FALKEMBACH, 2017).

De outro modo, alguns parâmetros e métricas são desenvolvidos especificamente para averiguar algo relacionado à funcionalidade técnica do agente conversacional. Neste âmbito, um estudo que investigou métricas de avaliações de *chatbot*, realizado por Goh, Kumar e Choon (2016), destacou a crescente necessidade de métricas para avaliar e medir a qualidade das respostas produzidas pelos sistemas de agentes conversacionais, baseadas em diferentes abordagens e domínios. A avaliação foi realizada em três sistemas de conversação diferentes, avaliadas nos agentes ELIZA (psicoterapeuta rogeriana), ALICE (entidade linguística artificial para computadores na Internet) e AINI (identidade artificial de linguagem natural).

Outro foco de análise de agentes conversacionais, envolvendo parâmetros e métricas, é proposto por Hussain, Sianaki e Ababneh (2019), que discutem e apresentam a classificação de *chatbots*, as técnicas de *design* usadas em *chatbots* anteriores e modernos, e como as duas principais categorias lidam com o contexto da conversa.

Registra-se que os *chatbots* anteriores apenas criaram uma ilusão de inteligência, empregando técnicas simples de correspondência de padrões e de processamento de *strings* (cadeias de caracteres) para contato com os utilizadores, usando modelos baseados em regras e modelos generativos. Nesse sentido, os autores apresentam um aporte teórico e classificação de agentes conversacionais, mas não apresentam nenhuma proposta de melhoria ou projeto viável.

Já o estudo proposto por Yu et al. (2016) foca em uma abordagem que utiliza dados de colaboração coletiva para aumentar o banco de dados de respostas, concentrando-se naquelas que as pessoas julgam inadequadas, para que, no final, se possa criar um conjunto de dados de respostas de bate-papo mais apropriadas, com o objetivo de identificar e substituir as que são particularmente inadequadas. Os resultados desta pesquisa apontam que a versão com o banco de dados expandido foi classificada significativamente melhor em relação à adequação do nível de resposta e da capacidade geral de envolver os usuários. Verifica-se, portanto, que foram qualificadas e analisadas as questões que envolvem o banco de questões em agentes conversacionais.

Observa-se que os benefícios do interesse no uso dos agentes conversacionais no meio educativo apresentam-se em diversos estudos envolvendo o progresso de *chatbots* para as mais diversas finalidades educacionais que, no final, revelam resultados sobre o sucesso do uso e sobre o quanto os *chatbots* contribuem para o aprendizado (SANTOS; ROLAND; SILVA, 2019). Okonkwo e Ade-Ibijola (2021) também destacam que o uso de *chatbots* constitui uma inovação tecnológica que aumenta o interesse do estudante na aprendizagem, além de ensejar a aquisição de habilidades cognitivas.

A crescente utilização dos *chatbots*, no âmbito educacional, apresenta muitos resultados a partir de parâmetros de avaliação, que permitem: avaliar e medir a qualidade das respostas produzidas pelo sistema (GOH; KUMAR; CHOON, 2016); classificar os *chatbots* e as técnicas usadas para lidar com o contexto da conversa (HUSSAIN; SIANAKI; ABABNEH, 2019). Além disso, pode-se examinar estudos envolvendo a base de dados, com a finalidade de apresentar respostas mais apropriadas para os diálogos (YU; XU; BLACK, 2016).

Levando-se em consideração as pesquisas apresentadas, que tratam do uso do agente conversacional no âmbito educacional, faz-se necessário aprofundar os estudos em parâmetros de avaliação de *chatbots* para uso no contexto educacional, com o propósito de apoiar os

profissionais de todos os segmentos da educação, identificando-se aí a lacuna de pesquisa da presente tese. Geralmente, os agentes conversacionais educacionais são desenvolvidos com finalidades específicas, tal como responder a conteúdos de Física, Metodologias Ativas ou outros. No entanto, não são apresentadas as métricas utilizadas para a avaliação dos respectivos agentes.

Diante disso, salienta-se que, ao se desenvolver um *framework* com recursos e métricas coerentes para validar a avaliação de qualquer agente conversacional no meio educacional, este estudo de tese irá preencher a lacuna que se traduz em avaliação de *chatbots* educativos. Adicionalmente, as métricas propostas nesta tese terão sua relevância para avaliar agentes conversacionais educacionais dos mais diferentes aspectos de avaliação, sem contar que as métricas apresentadas também poderão contribuir como referência para a elaboração de agentes conversacionais.

1.2 PROBLEMA DE PESQUISA

Tendo em vista os constantes avanços em relação ao uso dos agentes conversacionais no contexto educacional, como demonstrado, percebe-se que não há uma padronização para avaliá-los, uma vez que há poucas ferramentas e estudos tanto para avaliar, como para construir os *chatbots* educacionais, o que se constitui como um verdadeiro desafio. Constata-se, assim, o ensejo de examinar a área que avalia o *chatbot* para uso educacional, a partir da exploração e aprofundamento de métricas de avaliação já existentes.

Assim, considerando-se o exposto, a questão que norteia a proposta desta pesquisa é a seguinte: a elaboração de um *framework* enriquecido com métricas de avaliação de *softwares* poderia contribuir para a avaliação de agentes conversacionais educacionais?

1.3 OBJETIVOS DA PESQUISA

1.3.1 Objetivo geral

O objetivo desta tese consiste em desenvolver um *framework* com métricas de avaliação de *software*, a fim de elencar o que um agente conversacional educacional deve ter para ser utilizado como referência para a avaliação de *chatbots* de contexto educacional.

1.3.2 Objetivos específicos

Para alcançar o objetivo geral desta tese, foram estabelecidos objetivos específicos, os quais orientam, de forma precisa, a construção deste trabalho:

- Elicitar conceitos de agentes conversacionais no contexto educativo e identificar as abordagens tecnológicas utilizadas e as estratégias pedagógicas incorporadas;
- Investigar métricas exploradas em agentes conversacionais no contexto educacional;
- Investigar métricas de avaliação de *softwares*;
- Desenvolver um aplicativo para automatizar as funcionalidades das métricas e automatizar os processos de avaliação do *framework*;
- Validar as métricas de avaliação de agentes conversacionais voltados à educação com especialistas.

1.4 ESTRUTURA DA TESE

O desenvolvimento desta pesquisa abrange as seguintes etapas: resgate teórico, desenvolvimento de métricas de avaliação de agentes conversacionais e análise dos resultados.

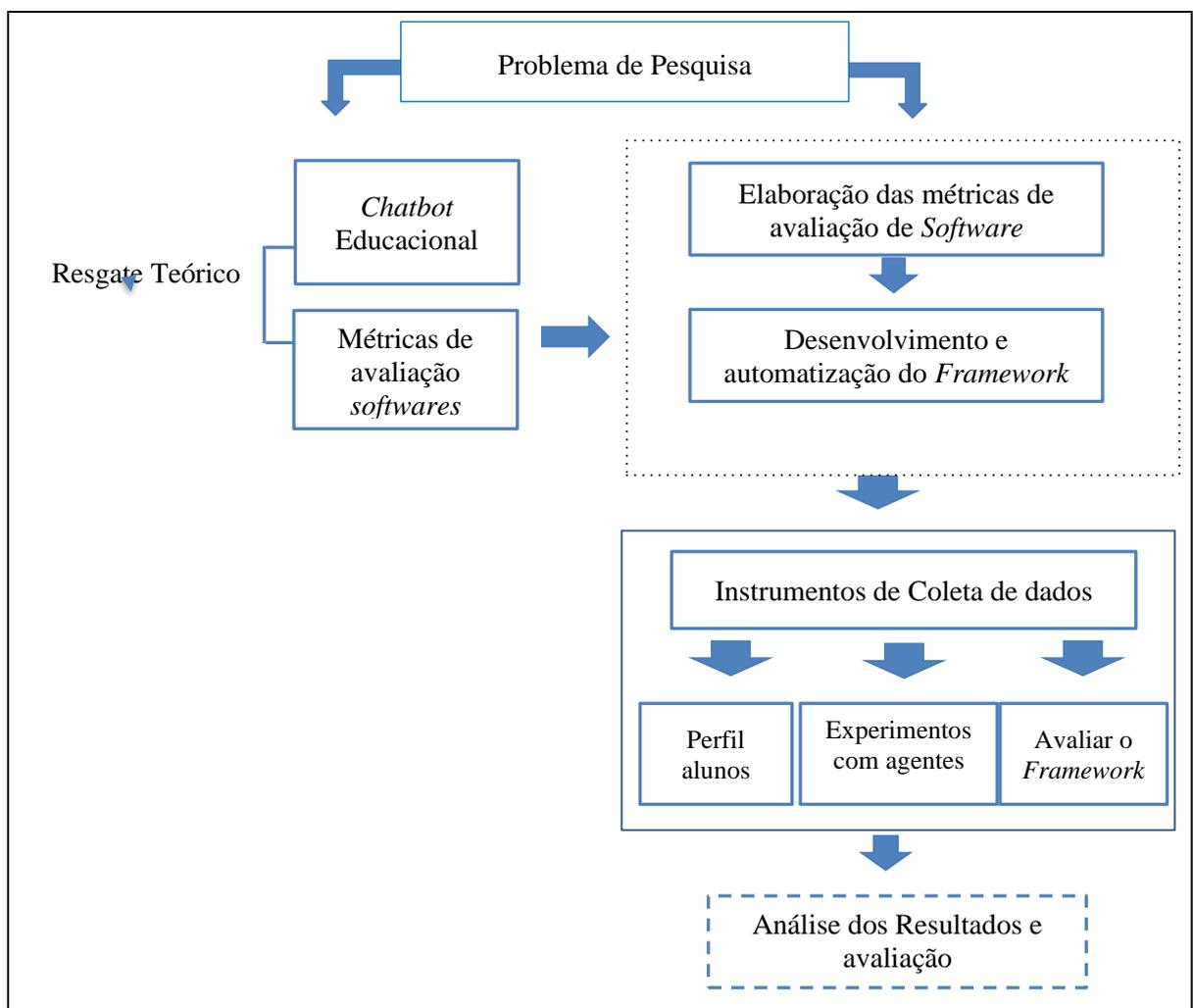
O resgate teórico versa sobre os temas centrais do problema de pesquisa: *chatbot* educacional e métricas de avaliação de *softwares* e agentes conversacionais. O trabalho de proposição de métricas de avaliação de *softwares* envolve a criação de um *framework* à luz dos principais autores e normas de avaliação que abordam métricas e avaliações de *softwares* e de agentes conversacionais, bem como outros modelos e estudos que tratam deste tema.

Já o apanhamento de dados compreende três fases. Na primeira, o pesquisador aplica um questionário aos alunos com questões relacionadas ao conhecimento e às experiências deles com a utilização de agentes conversacionais, de modo a verificar se conhecem especificamente os *chatbots* ATENA, METIS e ALTEIA. Na segunda fase, ocorrem as práticas e os experimentos com os discentes do Técnico em Informática para Internet do Senac/RS (Campus São Leopoldo), que aconteceram no contexto do Curso Técnico em Informática para Internet de forma integrada aos conteúdos programáticos. Já a terceira fase envolve a participação de

especialistas em agentes conversacionais e de desenvolvedores de *chatbots*, os quais realizam a validação das métricas do *framework* proposto.

A análise compreende a apresentação dos dados coletos, por meio dos instrumentos elaborados, e a avaliação dos resultados do estudo. Os resultados visam apontar se os indicadores de métricas de *softwares* sugerem o que um agente conversacional educativo pode e deve ter para o desenvolvimento e adequações de futuros *chatbots* educacionais. A figura 1.2 representa a estrutura da presente de tese.

Figura 1.2 - Estrutura da tese



Fonte: Elaborado pelo autor (2021).

A descrição e o detalhamento individual das etapas apresentadas na figura 1.2 encontram-se distribuídos ao longo de quatro capítulos. O segundo capítulo apresenta o referencial teórico de agentes conversacionais e métricas de avaliação de *softwares*. O terceiro

capítulo indica os procedimentos metodológicos adotados para o desenvolvimento desta pesquisa. O capítulo 4 demonstra o percurso de construção do *framework* para avaliação de *chatbots* e apresenta o aplicativo *ChatAval*, desenvolvido para a automatização do *framework*. O capítulo 5 apresenta os dados coletados na pesquisa e os analisa à luz do referencial teórico. Por fim, a conclusão registra os pontos conclusivos da tese e aponta possibilidades para desenvolvimento de trabalhos futuros.

A contribuição desta pesquisa consiste na elaboração de um *framework* automatizado, com o apoio de um aplicativo desenvolvido para automatizar a avaliação e construção de agentes conversacionais educacionais, com base em métricas de avaliação de *softwares*.

2 REFERENCIAL TEÓRICO

Este capítulo apresenta uma revisão bibliográfica sobre os principais tópicos abordados nesta tese, com foco nos agentes conversacionais, nas suas características e funcionalidades, nas métricas de avaliação e nos trabalhos relacionados.

2.1 AGENTES CONVERSACIONAIS: CARACTERÍSTICAS E FUNCIONALIDADES

Os agentes conversacionais, também denominados *chatbots* (*chatterbot*, assistentes virtuais ou agentes conversacionais), são aplicações que representam diálogo, por meio de um contato em linguagem natural (por texto ou voz) entre um humano, que estabelece a entrada, e o agente que responde a ele, apresentando respostas ou elaborando perguntas. Os *chatbots* já vêm sendo utilizados e analisados há algum tempo em variadas áreas, tais como atendimento a clientes suporte técnico *on-line*.

Esses agentes inteligentes são tipicamente usados em sistemas de diálogos com vários propósitos práticos, incluindo serviços ao consumidor e aquisição de informações (ABSHAWAR; ATWELL, 2005). Quando eles têm a competência de manter um diálogo com um humano, são considerados *chatbots*. Um *chatbot* que está dentro de um meio inteligente de aprendizagem deve ter controle do conteúdo e ter informações sobre as necessidades do usuário, para fornecer apoio e suporte na aprendizagem dos conceitos e nas atividades a serem deliberadas (KONZEN et al., 2011). Além disso, possui uma base de conhecimento que permite oferecer respostas a perguntas sobre um determinado assunto e mantém uma conversa com o utilizador, por meio de perguntas realizadas através de texto fornecido pelo estudante, simulando um tutor pedagógico que provê suporte ao ensino de conteúdos (PASCHOAL, 2017).

Entre os *chatbots* iniciais, construídos de 1964 até 1966, está o ELIZA. Ele foi desenvolvido para apontar a superficialidade na interação entre homem e máquina e foi capaz de responder as perguntas de um ser humano com um linguajar natural. Simulava diálogos utilizando um método de correspondência de padrões (*pattern matching*) e troca, ou seja, modificava os questionamentos recebidos em novos questionamentos devolvidos ao usuário, incentivando-o a permanecer se manifestando. Esta tática passava aos usuários que conversavam com ELIZA a ilusão de que estavam conversando com outro humano e não com um *chatbot* (WEIZENBAUM, 1966).

O *chatbot* ELIZA foi desenvolvido com o propósito de simular o comportamento de um psicólogo rogeriano, o qual utiliza um questionamento derivado das respostas para as perguntas do paciente (ROGERS, 1951). Seu funcionamento, segundo Clemente (2016), envolvia os fundamentos de identificar palavras-chave, encontrar um contexto, fazer transformações no texto em caso de necessidade, elaborar respostas na ausência de termos-chave, e concluir o bate-papo usando *scripts* para tal processamento. Nessa época, as bases de conhecimento ainda eram pequenas e não havia linguagens ou modelos apropriados.

Outro exemplo de *chatbot* é o PARRY, que intenciona aparentar uma pessoa com esquizofrenia. Ele foi implementado usando um molde de conduta de uma pessoa com a doença, baseando-se em convicções, conceituações e crenças. Além de ter uma estratégia conversacional, foi um programa mais sério e avançado do que o ELIZA (COLBY, 1981).

Desenvolvida por Michael Mauldin na Universidade Carnegie Mellon, a agente conversacional denominada JULIA teve a função de ajudar os utilizadores do sistema *TinyMUD* (*Multi-User Dungeons*). Sua principal característica foi a habilidade de lembrar informações registradas sobre as ações do usuário (MAULDIN, 1994). Este *chatbot* também apresentava a aptidão de aprender algumas coisas com o usuário, dando uma maior sensação de inteligência (ZDRAVKOVA, 2000).

Um dos programas recentes mais notáveis é A.L.I.C.E. (*Artificial Linguistic Internet Computer Entity*), também denominado como *Alicebot*, que utiliza a linguagem de marcação AIML (*Artificial Intelligence Markup Language*). A.L.I.C.E é um projeto pertencente à área da inteligência artificial, que começou na década de 90, com um método de código aberto preservado por uma comunidade ativa. O sistema é composto por uma máquina de inferência² conversacional e pela base de conhecimento em AIML (AGUIAR; VITORINO, 2015).

A linguagem AIML permite a estruturação das informações inseridas na base de conhecimento de *chatbots* baseados na tecnologia A.L.I.C.E. Conforme Augello et al. (2012), a linguagem AIML é específica para sua função como um agente conversacional e apoia-se em técnicas de correspondência de padrões sem qualquer capacidade de raciocínio.

A AIML foi amplamente disseminada pela comunidade de desenvolvedores do *software* livre *Alicebot* e formou a base para o primeiro *Alicebot*, também denominado A. L. I. C. E. - *Artificial Linguistic Internet Computer Entity* (WALLACE, 2003). No início, Wallace,

² Máquina (ou motor) de inferência é o componente do sistema que aplica os fatos e regras e heurística que compõem a base de conhecimento no processo de resolução do problema.

Tomabechi e Aimless (2003) previram *chatbots* atuando como livros falantes para crianças, ou para ensino de língua estrangeira e ensino em geral. Desde então, o uso desses agentes conversacionais como ferramentas educacionais tem sido continuamente explorado (KERLY; HALL; BULL, 2007).

Segundo Leonhardt et al. (2003), a AIML alicerça-se em padrões de entrada do usuário, que são comparados aos padrões descritos na base de conhecimento, selecionando as melhores respostas como saída. A estrutura da linguagem AIML é constituída de categorias, as quais consistem de dois elementos ao menos: o *pattern*, que corresponde a uma frase que o utilizador pode escrever (basicamente uma questão ou pergunta); e o *framework*, que armazena as respostas utilizadas durante os diálogos. Esta linguagem permite configurar e programar as respostas do agente conversacional. A figura 2.1 apresenta um tipo de código em AIML com algumas *tags* bastante utilizadas.

Figura 2.1 - Exemplos de código em AIML

```
<aiml>
  <category>
    <pattern>
      Olá
    </pattern>
    <template>
      Olá!? Seja muito bem vindo! Como posso lhe ajudar?
    </template>
  </category>

  <category>
    <pattern>
      Hoje passei só para te dar um oi! Tchau!
    </pattern>

    <template>
      Até mais! Precisando de algo estou aqui para lhe ajudar?
    </template>
  </category>
</aiml>
```

Fonte: Elaborado pelo autor (2021).

Constatam-se, na figura 2.1, as seguintes *tags*:

- *<aiml>* identifica o início e o fim de um documento AIML;
- *<category>* identifica o início e o fim de uma “unidade de conhecimento”;
- *<pattern>* representa a entrada do usuário - uma possível pergunta;
- *<framework>* representa a resposta mais apropriada para o padrão da *tag <pattern>*.

Como se pode observar no exemplo apresentado na figura 2.1, se o usuário digitasse “Olá”, ele receberia como resposta a frase “Olá!? Seja muito bem-vindo! Como posso lhe ajudar?”.

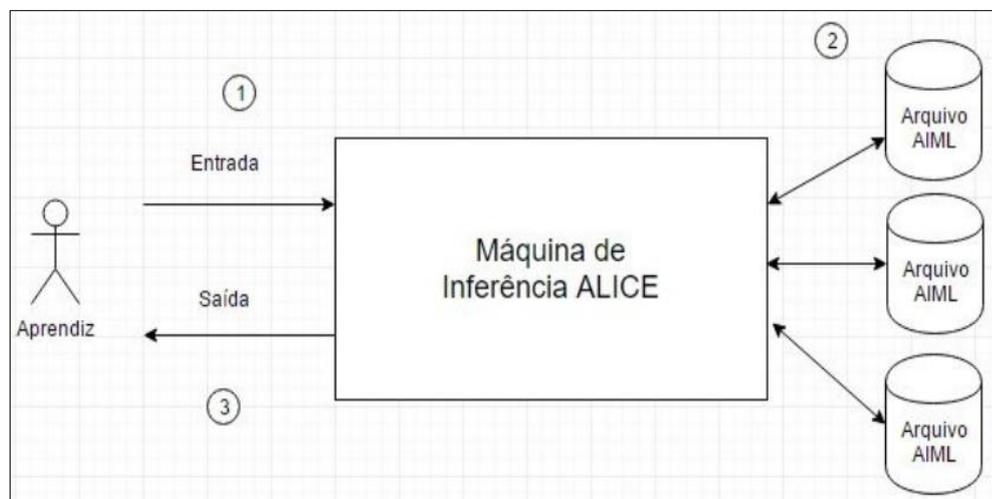
Há mais algumas *tags* utilizadas no linguajar AIML:

- `<random>` permite mais de uma saída, de forma aleatória, para um mesmo `<pattern>`, possibilitando que possíveis respostas possam ser colocadas dentro de uma `tag ` diferente; todas elas contidas pela `tag <random>`;
- `<srail>` possibilita a utilização de várias entradas em uma mesma saída;
- `<set>` possibilita definir e dar valor às variáveis dentro do arquivo AIML;
- `<get>` permite recuperar a informação contida em variáveis.

Para além dessas, segundo Leonhardt et al. (2003), existem mais de 20 *tags* e variações adicionais, compondo o linguajar AIML, que são responsáveis por fornecer a necessária desenvoltura para o *chatbot* propor uma resposta à mensagem recebida.

O *software* do agente conversacional compreende a consulta enviada, busca em uma base de conhecimento uma resposta que contenha elementos da consulta e seleciona, de acordo com alguns pré-requisitos, uma resposta para mandar ao usuário. Para poder realizar sua função com qualidade, o agente conversacional precisa contar com uma extensa base de conhecimento para apoio nas intervenções/reações. Também deve ser desenvolvido segundo algumas características de desenvolvimento de *software*, para que possa ainda contribuir plenamente para o processo de comunicação entre os sujeitos e o agente. A figura 2.2 demonstra a arquitetura de um *chatbot* baseado no sistema ALICE.

Figura 2.2 - Arquitetura de um *chatbot* baseado no sistema ALICE



Conforme a figura 2.2, no item 1, o aprendiz informa o texto que deseja enviar para o *chatbot*. Esta entrada é processada pelo sistema, gerando uma notação possível de ser usada para consulta na base de conhecimentos. No item 2, a consulta é feita sobre a base AIML em todos os arquivos disponíveis neste formato. No instante em que o sistema identificar qual a melhor resposta (*tag* <framework>) para o padrão apresentado (*tag* <pattern>), o *chatbot* retorna (item 3) ao aprendiz a frase encontrada.

Entre as potencialidades dos agentes conversacionais, Ghose e Barua (2013) destacam a competência em recuperar informações de forma instantânea e simples, sem que o discente tenha que procurar ou navegar em várias páginas da *web* para encontrar respostas a perguntas frequentes. Isso possibilita que o *chatbot* tenha aptidão para funcionar como um aliado virtual para o discente e, assim como ocorre com humanos, pode realizar assistências que interferem na Zona de Desenvolvimento Proximal - ZDP.

Desse modo, agentes conversacionais têm propensão para atuar de diversas formas, tais como: a) dar orientação sobre tarefas a serem realizadas (especialmente importantes em EAD, cujos estudantes relatam, com certa frequência, que se sentem perdidos e um tanto desorientados); b) responder as perguntas relacionadas aos conceitos e aos conteúdos que estão sendo estudados; c) proporcionar questionamentos aos estudantes de forma a instigar reflexão e meta-conhecimento; e d) sugerir alguns sites e vídeos com conteúdo adicional relativo ao contexto referido. Diante dessas estratégias, é pertinente criar um ambiente de aprendizagem que se baseia na teoria sociointeracionista de Vygotsky (1978), capaz de facilitar a aprendizagem colaborativa usando os agentes conversacionais como substitutos de aliados ou tutores reais. Desse modo, as relações sociais passam a compor elementos importantes para a produção de informações de forma colaborativa, por meio da conversa, da troca de informações e do embate de opiniões divergentes.

Mesmo que o *chatbot* esteja sendo utilizado há algum tempo em variados domínios, tais como atendimento a clientes e suporte técnico on-line (ABSHAWAR; ATWELL, 2015), ressalta-se que sua utilização vem se expandindo principalmente no contexto educacional (JACOB Jr. et al., 2011). É utilizado especialmente em formas de aprendizagem de Educação a Distância (EAD) e on-line (*e-learning*), chegando também a ser implementado de modo a agregar recursos e funcionalidades em ambientes 3D (SGOBBI; TAROUCO; REATEGUI, 2017). Wollny et al. (2020) realizaram uma revisão sistemática de literatura e constataram que os quatro principais objetivos no uso educacional de *chatbots* são aprimoramento de habilidades, eficiência na educação, motivação dos estudantes e disponibilidade contínua do

recurso. Okonkwo e Ade-Ibijola (2021) constataram que os agentes conversacionais têm sido usados como recurso educacional em atividades de ensino e aprendizagem (oferta de conteúdo, responder perguntas, obter ajuda individualizada) e, também, na administração, avaliação, assessoria, pesquisa e desenvolvimento.

Tibola et al. (2014) informam que um estudante de *e-learning* pode ficar decepcionado se não houver um suporte apropriado que o ajude em determinada tarefa, sendo, portanto, aconselhável fomentar algum suporte humano para esclarecer, ajudar e dar *feedback*. *Chatbots* são relevantes nesse sentido, pois, conforme Jacob Jr. et al. (2011), trazem informações em forma de respostas diretas a perguntas de estudantes. Contudo, apesar de todas as vantagens que o *chatbot* proporciona, esses sistemas ainda apresentam limitações e desafios, como ampliar a base de conhecimento, entregar respostas adequadas às perguntas e proporcionar algumas funcionalidades adicionais. Porém, o esforço compensa, pois o uso do *chatbot* oportuniza inúmeros benefícios à área da educação, especialmente com a oferta de informações na forma de conversas em linguagem natural, tornando Ambientes Virtuais de Aprendizagem (AVA) mais amigáveis. Inegavelmente, o uso dos agentes conversacionais na educação, segundo Fryer et al. (2017) e Crown et al. (2011), oportuniza inúmeros benefícios, sobretudo com a apresentação de informações na forma de conversas, tornando-se um aliado que proporciona uma comunicação engendrada para o aprendizado.

Neste ínterim, a educação vem cada vez mais se apoiando nos recursos tecnológicos para realizar a mediação estudante-professor. Alguns cursos EAD já podem ser totalmente realizados on-line, por meio dos ambientes virtuais de aprendizagem. A fim de torná-los mais interativos, muitas abordagens educacionais disponibilizam junto a estes ambientes (de forma paralela, por meio da *web*) ou integram a eles, inserindo-os no próprio código, os agentes conversacionais, tais como: a) Prof. Elektra, com foco no ensino de física e redes de computadores (LEONHARDT et al., 2005); b) Blaze, criado para aprimorar habilidades cognitivas de estudantes durante a resolução de problemas matemáticos (AGUIAR; VITTORINO, 2011); e c) Doroty, que auxilia no treinamento de usuários aprendizes em redes de computadores (LEONHARDT et al., 2005).

Graesser et al. (2005) apresentam um exemplo de como a aprendizagem colaborativa pode ser oportunizada com o uso de agentes conversacionais, ao introduzir o AutoTutor, que fornece comentários ao aluno (*feedback* positivo, neutro e negativo). Segundo os autores, o *chatbot* inquire o estudante para obter mais informações, solicita que dê sugestões e preencha informações, identifica e corrige respostas incorretas, responde às perguntas e resume as respostas.

Ademais, os *chatbots* estão inseridos em redes sociais, sites de empresas, sistemas de recomendação e na educação como ferramentas de apoio e motivação à aprendizagem (MANFIO et al., 2014). O uso de *chatbots*, conforme Dale (2016), compõe uma das maiores tendências tecnológicas da atualidade, facilitando contato com usuários de grandes corporações, como *Apple* (Siri), *Amazon* (Alexa), *Microsoft* (Cortana) e *Google* (Now).

Há, portanto, apoio para contribuir na concepção de *chatbots*, de novas plataformas de desenvolvimento para ajudar no fomento de agentes conversacionais que disponibilizam o modelo *Software as a Service* (SaaS)³, como, por exemplo, *Pandorabots*, *Chatfuel*, *Botsify*, *Watson Conversation* (IBM), *Rebot.me*, *Imperson*, *Wit.ai* (Facebook), *Api.ai* (Google), dividindo as atribuições entre o provedor de serviços e o cliente (RADZIWILL; BENTON, 2017). A partir do cadastramento e do *login* de acesso, o usuário pode criar e personalizar seu próprio *chatbot* on-line.

Outros serviços que auxiliam na incorporação de agentes conversacionais são as soluções PaaS (*Plataform as a Service*)⁴ disponíveis para *download*, como *RiveScript*⁵ e *Superbot*, as quais apresentam *frameworks* para programação do *chatbot* em linguagem de *script* ou AIML. O *Superbot* é um kit de desenvolvimento oferecido pela própria comunidade A.L.I.C.E. (*Artificial Linguistic Internet Computer Entity*), proporcionando um mote para a construção ágil da base de conhecimento por meio de planilhas.

Além disso, os *chatbots* também podem ser utilizados em outros contextos, como, por exemplo, nos mundos virtuais imersivos. Para Greis e Reategui (2010), a concepção de mundos virtuais pressupõe atributos tanto de imersão como de interação e navegação. Os agentes possibilitam, ainda, , experimentos, o uso de espaços virtuais, como salas e laboratórios, entre outros, com diversos usuários agrupados num mesmo espaço (ambiente) ao mesmo tempo, além de interface gráfica (fornece um ambiente virtual tridimensional); imediaticidade (promove interação em tempo real); interatividade (proporciona opções de conexão e de criação de objetos e conteúdo); e socialização e comunidade (possibilita a formação de grupos e comunidades com possíveis interesses em comum).

³ No modelo SaaS, o fornecedor do *software* se responsabiliza por toda a estrutura necessária à disponibilização do sistema (servidores, conectividade, cuidados com segurança da informação), e o cliente utiliza o *software* via internet, pagando um valor pelo serviço. O usuário não administra as características individuais da aplicação, exceto configurações específicas. Os desenvolvedores se concentram em atualização e não em infraestrutura.

⁴ PaaS (Plataforma como serviço) é um ambiente de desenvolvimento e implantação completo na nuvem, com recursos que permitem fornecer tudo, de aplicativos simples baseados em nuvem a sofisticados aplicativos empresariais habilitados para a nuvem. Adquirem-se os recursos necessários por meio de um provedor de serviços de nuvem em uma base pré-paga, sendo possível acessá-los por uma conexão com a Internet segura.

⁵ Maiores informações no site: <https://www.rivescript.com/>

Assim, o uso dos *chatbots* pode ampliar as funcionalidades dos mundos virtuais imersivos, fomentando ainda mais os recursos e tecnologias. O mesmo acontece com um agente conversacional, o que permite sua utilização para diálogos no processo de ensino e aprendizagem de qualquer conteúdo, abrangendo os ensinos fundamental, médio, técnico, superior e a formação continuada.

Como exemplo de aplicação prática no contexto da educação, há o Projeto AVATAR⁶ (Ambiente Virtual de Aprendizagem e Trabalho Acadêmico Remoto), da Universidade Federal do Rio Grande do Sul (UFRGS), o qual busca investigar, testar e promover a capacitação para o uso de laboratórios virtuais em ambientes imersivos. Este projeto tem o intuito de implementar formas de oportunizar a aprendizagem a partir de laboratórios virtuais com mundos virtuais, por meio do uso da plataforma *OpenSimulator* (HERPICH et al., 2017). Tem seu foco especialmente no ensino de ciências, um campo conceitual que possui uma demanda por laboratórios para realizar experimentos práticos feitos por discentes. Mas existe a necessidade de direcionamento das atividades aos estudantes, assim como acontece nos laboratórios reais, de modo que saibam o que pode ser feito nesses espaços, possibilitando, assim, explicações e orientações relativas aos experimentos e aos conceitos subjacentes, a fim de instigar reflexão relacionada ao que foi observado na realização dos experimentos.

Destaca-se que, no Projeto AVATAR, são utilizados *Non Player Characters* – NPC, que podem ser instanciados em qualquer ponto do mundo virtual e em diferentes condições, quando o avatar do usuário (ou a personagem que representa um determinado usuário) se aproxima de determinada localização, ou quando alguma ação do usuário ocorre, por exemplo, se terminou de realizar um experimento. Se toda a programação necessária para a habilidade de reação do agente tivesse que ser criada a partir do seu início, usando a linguagem de *scripts*, a tarefa seria imensa. Nesse sentido, investigou-se a ligação entre o NPC, presente no mundo virtual, com um *chatbot* externo baseado na máquina de inferência *ALICEbot*.

Junto aos laboratórios, foram desenvolvidos diferentes agentes conversacionais para uso no mundo virtual imersivo e mesmo em páginas *web*, que poderiam ser acessadas diretamente pelos estudantes por intermédio de um link em um navegador de internet em qualquer computador, ou por meio de ambientes virtuais de aprendizagem, como o *Moodle*.

⁶ O Projeto AVATAR visa estimular o desenvolvimento de objetos multimídia interativos para o ambiente imersivo *Open Simulator*, que é um *software* livre e implementa um ambiente aberto no qual é possível adicionar cenários compostos de artefatos virtuais que podem ser manuseados pelos usuários. Maiores informações no site do Projeto disponível em: <http://www.ufrgs.br/avatar>.

No âmbito do Projeto AVATAR, inicialmente, foi desenvolvida a ATENA⁷ (Agente Tutora para Ensino e Navegação no Ambiente de Física), que é uma agente conversacional desenvolvida para o projeto AVATAR, e pode ser acessada diretamente pela interface *web*. Duas outros agentes conversacionais foram adicionalmente desenvolvidas: HIGIA (Individual Habitat Individual e Guia Interativo de Atitudes), usada em um projeto com o objetivo de motivação de indivíduos para a prática de atividade física (SGOBBI; TAROUCO; REATEGUI, 2017), e a Agente METIS (Mediadora de Educação em Tecnologia Informática e Socializadora) usada via interface *web* almejando deixar mais fácil e mais real essa interação.

Isto posto, cabe registrar que a presente tese, apresentada ao Programa de Pós-Graduação em Informática na Educação, está vinculada à linha de pesquisa de Ambientes Informatizados e Ensino a Distância, na qual o Projeto AVATAR vem sendo desenvolvido. Considerando que o objetivo da tese consiste em criar um *framework* com sugestões de métricas de avaliação de *software*, a fim de elencar o que um agente conversacional educacional pode e deve ter para ser relevante, com vistas a ser utilizado como referência para a avaliação e construção de *chatbots* desenvolvidos para o contexto educacional, serão utilizados os três agentes conversacionais utilizados no Projeto AVATAR: METIS, ATENA e ALTEIA, com o propósito de investigar, sugerir e propor futuras melhorias. Acredita-se, ainda, que o *framework* desenvolvido contribuirá não apenas com o resultado do estudo da presente tese, mas também com novas sugestões de métricas para o desenvolvimento dos agentes conversacionais já utilizados no Projeto AVATAR. Assim, as próximas subseções apresentam os respectivos agentes.

2.1.1 Agente METIS

A agente conversacional METIS (Mediadora de Educação em Tecnologia Informática e Socializadora) foi desenvolvida com a finalidade de dialogar com os discentes, por meio de uma interface que simula uma conversação on-line (*chat*) entre pessoas. A implementação da METIS fundamenta-se no sistema A.L.I.C.E., desenvolvido por Wallace (2003), o qual organiza a base de conhecimento, notação baseada em XML (*eXtensible Markup Language*), AIML. Foi desenvolvida utilizando o *software Open Source Program-O* e possui uma interface

⁷ Acesso ao agente ATENA em: <http://avatar.cinted.ufrgs.br/atenaweb/gui/jquery/>

em *Cascading Style Sheets* (CSS), possibilitando que a tela seja adaptada de acordo com o dispositivo, o que facilita a interação de alunos até mesmo por um aparelho celular.

A base de dados conversacionais da METIS foi construída a partir de perguntas e respostas elaboradas por especialistas na área da Educação, mídias e tecnologias. Já a base de conhecimento do sistema foi criada por especialistas em mídias e informática na educação, de modo a proporcionar um conhecimento mínimo dos conteúdos e respostas positivas às dúvidas apresentadas pelos estudantes.

A METIS possui, na sua base de conhecimento, um conjunto de dados relacionados com tecnologias educacionais⁸. Conforme a figura 2.3, pode-se constatar que o aluno envia uma mensagem, em que, neste exemplo que se apresenta, está questionando METIS: “você poderia me explicar o que é licença de *software*?”, e a agente responde com uma mensagem de texto contendo um conceito para o conteúdo e apresenta ainda um vídeo relacionado à resposta. Esta estrutura está organizada no formato de um *frame* disponibilizado na própria janela de *chat* com tamanho pré-definido. Quando as respostas da agente são muito longas ou possuem muitos recursos, a janela de resposta apresenta uma barra de rolagem em sua lateral direita, permitindo que o usuário tenha acesso a todo o conteúdo apresentado pelo agente conversacional. No que se refere ao conteúdo de vídeos, registra-se que podem ser visualizados em tamanho miniatura, o que é exibido como padrão dentro da área de resposta da agente, conforme está demonstrado na figura 2.3. Mas podem também ser assistidos em formato de tela cheia, clicando-se na opção “Tela inteira”, que se encontra disponível no vídeo.

Figura 2.3 – Tela da Agente METIS



Fonte: METIS (2020).

⁸ A agente METIS pode ser acessada em: <http://avatar.cinted.ufrgs.br/metis/>

A METIS foi desenvolvida de modo que outros recursos multimídia sejam agregados às respostas, por meio de *links* para repositórios digitais, ou de bibliotecas eletrônicas, consultas na *web*, áudio e imagens. Neste sentido, uma vez que haja a possibilidade de vincular diversos tipos de recursos para serem compartilhados com os usuários, acredita-se que este modelo de agente conversacional - com respostas envolvendo diversos recursos de mídia - pode encorajar o estudante a procurar mais informações para além das informações disponíveis na base de conhecimento da METIS, atuando, assim, como um companheiro virtual.

Alguns experimentos envolvendo as conversações entre a METIS e alunos facilitaram a realização de inspeções nos registros das interações (*logs*), possibilitando detectar perguntas não respondidas pelo sistema e avaliá-las, considerando sua relevância, para adicionar informações na base de conhecimento. No caso da agente METIS, foi realizado um processo de aprimoramento, por meio do qual todas as perguntas realizadas pelos usuários foram avaliadas por humanos especialistas, com vistas a identificar e aprimorar a qualidade das respostas.

2.1.2 Agente ATENA

O Agente Tutor para Ensino e Navegação no Ambiente – ATENA⁹, que também faz parte do Projeto AVATAR, foi desenvolvido para conversação na área de física, possui capacidade de diálogo e acompanha o estudante em seu percurso no cenário imersivo virtual. A figura 2.4 apresenta a imagem do ATENA em momento de interação com um usuário.

Figura 2.4 - Agente Conversacional ATENA



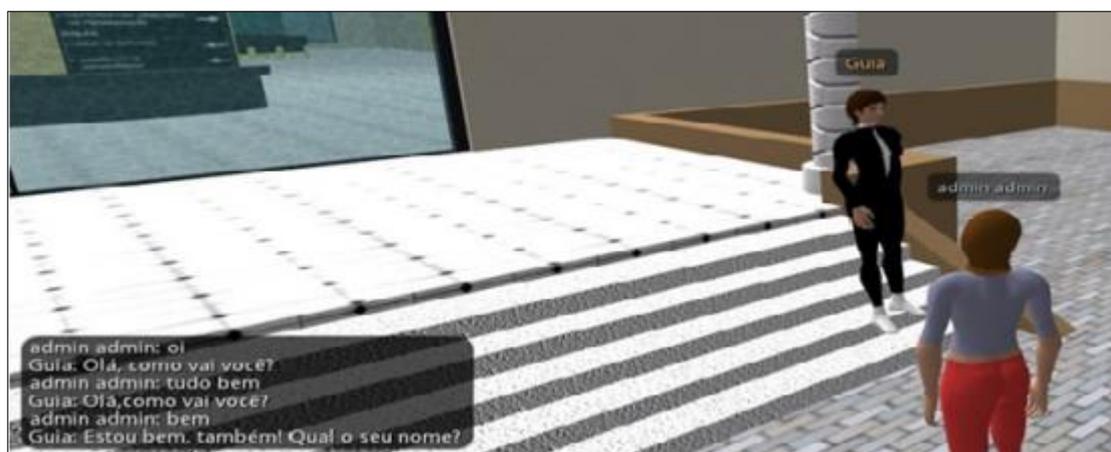
Fonte: ATENA (2020).

⁹ Disponível em: <http://avatar.cinted.ufrgs.br/atenaweb/gui/jquery/?>

Na figura 2.4, observa-se que foi questionado ao ATENA “o que é física?” e o agente apresentou uma resposta em forma de texto e uma imagem relacionada ao questionamento.

No âmbito do Projeto AVATAR, Sgobbi, Herpich e Nunes (2019) utilizaram a conexão de um *Non-Player Character* – NPC, ou seja, uma personagem programável e não manipulável por um avatar em Mundos Virtuais Imersivos. A figura 2.5 demonstra uma ilustração entre a integração do mundo virtual e o ATENA.

Figura 2.4 - Exemplo de integração de *chatbot* e AVATAR



Fonte: Sgobbi, Herpich e Nunes (2019).

Conforme o exemplo, diálogos como “Oi!”, “Olá, como vai você?”, entre os demais apresentados, podem vir de um agente conversacional. Ao programá-lo, é possível definir as opções de respostas, com base nas perguntas, para ampliar a base de conhecimento e, conseqüentemente, aperfeiçoar o diálogo de agentes conversacionais integrados aos mundos virtuais. Ainda conforme a figura 2.5, houve configurações e programação de *scripts* para estabelecer a conexão entre NPC e ATENA para que, em determinadas situações previstas e programadas no mundo virtual, ATENA pudesse surgir no modo de agente conversacional e auxiliar os alunos em suas dúvidas.

A figura 2.6, a seguir, apresenta algumas linhas do *script* para verificar a presença e a ausência do avatar, a fim de estabelecer o diálogo e conexão entre NPC e ATENA.

Figura 2.6 - Exemplo de *script* do avatar e *chatbot*

```

IInstantMessage(idAvatar, "Olá "+avatarName+", seja bem-vindo ao Projeto AVA-
TAR.");
    IISleep(1);
IInstantMessage(idAvatar, "Eu sou a Atena e estou aqui para o auxiliar com os experi-
mentos de física.");
    IISleep(2);
IInstantMessage(idAvatar, "Converse comigo no canal " + escutaCanal + " (ao escre-
ver uma mensagem utilize o prefixo '/' + escutaCanal + "mensagem' sem as aspas).");
    } else {
        vector userPos = IIDetectedPos(0);
        vector npcPos = osNpcGetPos(idNPC);
        vector diff = userPos - npcPos;
        diff.z = 0;
        vector initial = <1.0,0.0,0.0>;
        osNpcSetRot(idNPC, IIRotBetween(initial,diff));
    }

```

Fonte: Sgobbi, Herpich e Nunes (2019).

2.1.3 Agente ALTEIA

O Agente Lexical de apoio a Tarefas Educacionais usando Inteligência Artificial – ALTEIA¹⁰ é um *chatbot* especialista em Metodologias Ativas, Sala de Aula invertida e ferramentas de autoria. Possui capacidade de entregar respostas aos estudantes em relação a conceitos e responder a diversos questionamentos relativos à sua especialidade.

A figura 2.7 apresenta a tela inicial do agente ALTEIA.

Figura 2.7 - Agente Conversacional ALTEIA

Alteia

Fale:

Usuário: o que são metodologias ativas?
 ALTEIA: Eu gosto de Metodologias Ativas! Nelas o aluno passa a ser o protagonista no seu aprendizado, retendo conteúdo por meio de técnicas de ensino tanto dentro da sala de aula como fora, estimulando sua autonomia e participação. Existem diversos tipos de metodologias ativas.
 Para saber mais clique [aqui](#).

Fonte: ALTEIA (2020).

¹⁰ Disponível em: <http://avatar.cinted.ufrgs.br/alteia/gui/plain/index.php#end>.

Observa-se que, ao acessar o *site* inicial do ALTEIA, o usuário envia uma pergunta em forma de texto ao agente conversacional: “o que são metodologias ativas?”. Na sequência, o *chatbot* não apenas apresentou o retorno, também em formato de texto, como estabeleceu um pequeno diálogo com antes da resposta: “Eu gosto de Metodologias Ativas!”. Além disso, após apresentar a resposta à pergunta realizada, o ALTEIA sugeriu um *link* que pode direcionar o estudante a novas possibilidades de pesquisas.

Atualmente, o Projeto do agente conversacional ALTEIA está em vias de desenvolvimento, mas possui uma base de conhecimento capaz de dialogar utilizando frases, perguntas e afirmações, além de saber qual é a data atual e de poder realizar alguns cálculos. O *site*¹¹ apresenta um vídeo, que aborda as principais funcionalidades e recursos do ALTEIA para auxiliar os estudantes a realizarem suas perguntas. Também apresenta algumas de suas limitações, conforme pode ser constatado na figura a seguir.

Figura 2.8 - Vídeo explicativo da agente conversacional ALTEIA



Fonte: ALTEIA (2020).

No vídeo explicativo, conforme a figura 2.8, os alunos são orientados a realizar perguntas simples e diretas, considerando que a referida agente ainda está aprendendo e, a partir da ampliação de sua base - e devido a este fato -, frases complexas ainda confundem a busca de respostas. O vídeo explicativo também: 1) orienta os alunos a utilizar palavras-chave; 2) apresenta dicas de como formular e consultar nos diálogos com o ALTEIA; 3) orienta a fazer uma pergunta por vez; 4) informa que se deve utilizar palavras-chave em uma frase apenas; 5)

¹¹ As principais informações do Projeto Alteia podem ser consultadas no *site*: <http://penta3.ufrgs.br/Avatar/Alteia/index.html>

informa, também, que quanto mais conversar com a agente, mais ela vai aprender. Consideram-se relevantes as informações do vídeo para os estudantes que irão conversar com o ALTEIA, afinal, a agente conversacional tem em sua programação a função de buscar e armazenar informações em sua base de conhecimento, a fim de entregar respostas conforme as demandas dos estudantes.

A primeira utilização do agente ALTEIA foi em janeiro de 2020 em um curso de Metodologias Ativas para docentes da Universidade Federal do Rio Grande do Sul. Uma das tarefas atribuídas a eles foi a de avaliar o conhecimento deste agente sobre este campo conceitual, após terem assistido aos vídeos e inspecionado os materiais disponibilizados no curso sobre Sala de Aula Invertida e Metodologias Ativas. Depois de terem conversado com o ALTEIA, a base de conhecimento foi avaliada, pois todas as perguntas e as respectivas respostas foram registradas no *log* de acesso. A partir do respectivo *log*, a equipe de suporte pôde constatar quais respostas o *chatbot* teve dificuldades para responder. Após o primeiro curso, a base de conhecimento foi ajustada para incrementar, criar e ampliar novas informações, de modo a torná-la cada vez mais completa. Há um trabalho contínuo de verificação das perguntas realizadas na base de conhecimento e, caso o ALTEIA não tenha a informação, um novo conhecimento será adicionado.

2.1.4 Tendências em *chatbots*

Ao longo da seção 2.1, procurou-se trazer conceitos de *chatbots*, resgatando desde os primeiros *chatbots* construídos, como o ELIZA, em 1964, passando pela máquina de inferência da A.L.I.C.E, que iniciou na década de 90, e a linguagem AIML. Além disso, apresentaram-se os agentes METIS, ATENA e ALTEIA que foram utilizados neste estudo.

Cabe registrar, entretanto, que desde os primeiros *chatbots* houve diversas mudanças. Recentes tendências em *chatbots* podem ser vistas como uma resposta direta à aceitação de grandes empresas de tecnologia, especificamente a inclusão do Siri como parte do sistema operacional da *Apple*, em 2011, a promoção do Alexa pela *Amazon* desde 2014, a chegada do Cortana para computadores da *Microsoft*, em 2015, e o *Watson* da *IBM*, em 2010.

A utilização dos *chatbots* tem sido bem aceita e estima-se que, em 2019, mais de 50% dos consumidores americanos e alemães usaram *chatbots* pelo menos uma vez, com números ainda maiores no Reino Unido ou na França, o que demonstra que seu uso está se tornando comum. Os gastos da Ásia/Pacífico (excluindo o Japão) em sistemas de Inteligência Artificial atingiram US\$ 6,2 bilhões em 2019, registrando um aumento de quase 54% em relação a 2018.

Os gastos com sistemas de inteligência artificial têm uma previsão de aumento de US\$ 21,4 bilhões até 2023, com uma taxa de crescimento anual composta de 39,6% (PRESS, 2019).

Atualmente, os *chatbots* são utilizados nas mais diversas áreas de aplicação, envolvendo desde o atendimento ao cliente, saúde, educação, apoio ao trabalho, além de um acentuado interesse na pesquisa de *chatbots* no meio acadêmico. Paralelamente a isso, o crescente número de pesquisas sobre *chatbots* tem um caráter interdisciplinar marcante, abrangendo áreas como informática, gestão e marketing, ciência da mídia e comunicação, linguística, filosofia, psicologia e sociologia, engenharia, *design* e interação humano-computador (FOLSTAD et al., 2021).

Nesse sentido, à medida que a utilização de *chatbots* aumenta nas mais diversas áreas, há a necessidade de definir direções para que novos estudos e iniciativas sejam sistematicamente construídos. O *framework* proposto nesta pesquisa apresenta métricas para avaliação de agentes conversacionais e tem foco predominantemente educacional. No entanto, poderá também ser utilizado como apoio para a avaliação e construção de *chatbots* para uso em outras áreas de aplicação.

Os *chatbots* atuais contêm uma grande variedade de tecnologias e serviços nos mais diversos níveis, envolvendo abordagens simples baseadas em regras, sistemas baseados em dados estatísticos ou abordagens neurais (MCTEAR, 2020). Por outro lado, Folstad et al. (2021) apontam algumas tendências para estudos futuros e desenvolvimentos de *chatbots*, divididas em seis tópicos: 1) Usuários e implicações; 2) Experiência e design do usuário do *chatbot*; 3) Estruturas e plataformas de *chatbot*; 4) *Chatbots* para colaboração; 5) Democratização dos *chatbots* - *chatbots* para todos; 6) Ética e privacidade em *chatbots*.

Outra tendência envolve a empatia das respostas. Segundo Acheampong (2020), o desafio em possibilitar empatia e respostas emocionalmente ajustadas envolve a capacidade do agente em detectar o estado emocional do humano, e ser capaz de fornecer uma resposta emocional adequada relacionada à consulta de uma pessoa durante suas conversas. Essa tendência envolvendo a empatia nas interações será um grande passo na conversação, pois, ao verificar algumas tecnologias, os diferentes modelos de linguagem propostos como arquitetura de *chatbots* ainda são incapazes de imitar corretamente a conversa humana devido à abordagem incorreta da modelagem do diálogo. Na linguagem AIML, por exemplo, existe uma tentativa de resolver problemas de conversação com uma abordagem de pergunta e resposta previamente cadastrada na base de conhecimento. Assim, ao realizar uma determinada pergunta para o agente conversacional, ele busca em sua base de dados respostas correlacionadas com os termos-chave da pergunta e tenta selecionar a resposta mais adequada. Mas existe o risco de a

base de dados ainda não ter a informação sobre a questão que está sendo abordada. Este, no entanto, não é o raciocínio por trás da conversa humana, que não avança simplesmente uma entrada e uma saída prevista, pois é fundamental levar em consideração uma série de etapas anteriores, como o contexto subjacente da conversa e as informações que estão sendo compartilhadas entre os participantes. Para poder trabalhar a empatia das respostas, de acordo com o estado emocional do humano, é necessário usar um sistema mais complexo de tratamento das entradas do usuário. Casas et al. (2021) descrevem a criação de um *chatbot* que pode conversar com empatia. Os autores usaram um classificador de emoções (DeepMoji) e um gerador de respostas (Generative Pre-trained Transformer (GPT) model) implementado em 2 CPUs de alta performance Tesla v100 GPUs.

Nesse sentido, observa-se que, além das questões técnicas de compreensão e geração de linguagem natural, bons agentes conversacionais devem ter conhecimento das características humanas, observar as emoções dos usuários e fornecer empatia em suas respostas para engajá-los. Por outro lado, de acordo com estudos de Clark et al. (2019), os recursos sociais relatados como fundamentais em uma conversação humano-humano, como compreensão e terreno comum, confiança, escuta ativa e humor, são necessários para conversas entre humanos e o agente conversacional.

Constata-se que em relação aos estudos relacionados à empatia nas conversas, a variedade de recursos que ocorreram nos últimos anos se deu a partir do poder computacional e o compartilhamento de tecnologias e *softwares* de código aberto, possibilitando, assim, fazer com que os programas de *chatbots* se tornassem cada vez mais comuns. Para Caldarini, Jaf e Macgarry (2022), o desenvolvimento de recentes técnicas de Inteligência Artificial e Processamento de Linguagem Natural tornaram os *chatbots* mais fáceis de implementar, mais flexíveis em termos de aplicação e manutenibilidade e cada vez mais capazes de imitar a conversação humana. No entanto, a interação entre o humano e o *chatbot* não é perfeita.

Por fim, um aspecto da implementação dos *chatbots* é sua avaliação. Avaliar sistemas de conversação tem se mostrado uma tarefa desafiadora, uma vez que a conversação humana atende a diferentes objetivos e funções. Dependendo do objetivo do *chatbot*, as métricas usadas para avaliar o diálogo podem ser bem diferentes. Um *chatbot* de assistente pessoal, por exemplo, será avaliado principalmente com base na eficácia da interação, podendo envolver métricas que respondam a questões como: “O *chatbot* realizou a tarefa com sucesso?”, ou, “O diálogo entre o agente conversacional e o usuário foi eficiente?”. Já um *chatbot* companheiro poderá ser avaliado em sua capacidade de manter a conversa e de envolver os usuários.

O estudo de Caldarini, Jaf e Macgarry (2022) aponta que há diversas pesquisas que tratam de *chatbots*, apresentando e discutindo questões de tecnologias e implementação, bem como questões relacionadas aos diálogos de conversação. Porém, a avaliação não é analisada em profundidade em artigos recentes, concentrando-se mais nas estruturas que abarcam o contexto do *chatbot*, em vez de comparar várias métricas de avaliação. Assim, a falta de uma proposta de referência comum sobre a avaliação de *chatbots* limita o teste correto e a comparação de diferentes modelos. Neste sentido, a proposta da tese de se criar um *framework* com métricas de avaliação de agentes conversacionais educacionais contribuirá, de forma significativa, com as questões de avaliação de agentes conversacionais, especificamente, os agentes conversacionais educacionais.

A próxima seção aborda as métricas de avaliação de *softwares* e apresentar estudos relacionados que envolvem métricas de avaliação de agentes conversacionais educacionais.

2.2 MÉTRICAS DE AVALIAÇÃO DE *SOFTWARES*

Esta seção tem o propósito de apresentar métricas de avaliação de *softwares*. Considerando o uso de agentes conversacionais na área da Educação e o objetivo da tese de criar um *framework* com sugestões de métricas de *software*, a fim de sugerir o que um *chatbot* educacional pode e deve ter para ser interessante, e para poder ser utilizado como referência para a avaliação e construção de *chatbots* desenvolvidos para o contexto educacional, a presente tese utiliza elementos da *International Organization Standardization* (ISO) e da *International Electrotechnical Commission* (IEC).

Como organismos normalizadores de relevância internacional, por meio de comitês técnicos estabelecidos e reconhecidos no setor de *software*, a ISO e a IEC são responsáveis por editar e definir a qualidade de um *software*, a fim de que as características de um produto *software* seja capaz de satisfazer necessidades explícitas e implícitas. As necessidades explícitas são requisitos que determinam as condições de utilização dos produtos e informam os objetivos, funções e desempenho esperado. Já as implícitas são funcionalidades fundamentais para o usuário e precisam ser declaradas em razão de possíveis consequências. Os padrões ISO são internacionalmente aceitos por especialistas e são considerados como uma fórmula que apresenta a melhor maneira de fazer algo.

Desenvolver ou selecionar produtos de *software* de alta qualidade é de primordial importância em toda a sociedade contemporânea que utiliza as Tecnologias de Informação e

Comunicação. No contexto educativo, desenvolver e selecionar *softwares* educacionais com o uso de agentes conversacionais não seria diferente, pois especificá-los e avaliá-los são fatores fundamentais para garantir qualidade adequada e para que o *software* venha a ser desenvolvido.

Nesse sentido, a ISO/IEC 25010 é uma norma ISSO, disponibilizada em 2011, que trata da qualidade do produto de *software*, e foi desenvolvida especificamente para dar suporte às necessidades de métricas de avaliação de *software*. Substituiu a norma ISO/IEC 9126 e passou a contemplar as características “segurança” e “compatibilidade”.

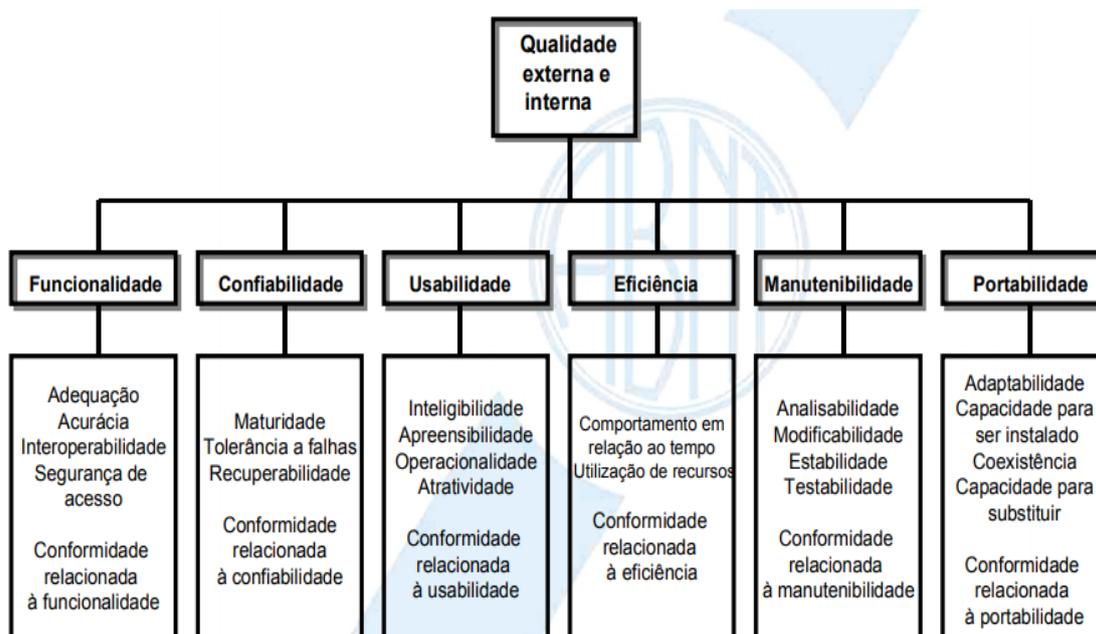
No estudo da presente tese, são apresentadas as métricas da NBR ISO/IEC 9126-1 e, posteriormente, as duas novas métricas da ISO/IEC 25010, pois o levantamento e o conhecimento das respectivas métricas são fundamentais para a composição e elaboração do *framework* proposto nesta pesquisa.

2.2.1 Métricas da ISO/IEC 9126-1

O modelo de qualidade ISO/IEC 9126-1 apresenta seis características de qualidade (Funcionalidade, Confiabilidade, Usabilidade, Eficiência, Manutenibilidade e Portabilidade), as quais são, por sua vez, subdivididas em subcaracterísticas.

A figura 2.9 demonstra o esquema das categorias e subcategorias propostas pela NBR ISO/IEC 9126-1.

Figura 2.9 - Modelo de qualidade para qualidade externa e interna NBR ISO/IEC 9126-1



A seguir, apresenta-se a descrição de cada característica e de suas respectivas subcaracterísticas, conforme o manual NBR ISO/IEC 9126-1.

Funcionalidade: Capacidade do produto de *software* de prover funções que atendam às necessidades explícitas e implícitas, quando o *software* estiver sendo utilizado sob condições especificadas.

- **Adequação:** Capacidade do produto de *software* de prover um conjunto apropriado de funções para tarefas e objetivos do usuário especificados.
- **Acurácia:** Capacidade do produto de *software* de prover, com o grau de precisão necessário, resultados ou efeitos corretos ou conforme acordados.
- **Interoperabilidade:** Capacidade do produto de *software* de interagir com um ou mais sistemas especificados.
- **Segurança de acesso:** Capacidade do produto de *software* de estar de acordo com normas, convenções ou regulamentações previstas em leis e prescrições similares relacionadas à funcionalidade.
- **Conformidade relacionada à funcionalidade:** Capacidade do produto de *software* de estar de acordo com normas, convenções ou regulamentações previstas em leis e prescrições similares relacionadas à funcionalidade.

Confiabilidade: Capacidade do produto de *software* de manter um nível de desempenho especificado, quando usado em condições especificadas.

- **Maturidade:** Capacidade do produto de *software* de evitar falhas decorrentes de defeitos no *software*.
- **Tolerância a falhas:** Capacidade do produto de *software* de manter um nível de desempenho especificado em casos de defeitos no *software* ou de violação de sua interface especificada.
- **Recuperabilidade:** Capacidade do produto de *software* de restabelecer seu nível de desempenho especificado e recuperar os dados diretamente afetados no caso de uma falha.
- **Conformidade relacionada à Confiabilidade:** Capacidade do produto de *software* de estar de acordo com normas, convenções ou regulamentações relacionadas à confiabilidade.

Usabilidade: Capacidade do produto de *software* de ser compreendido, aprendido, operado e atraente ao usuário, quando usado sob condições especificadas.

- **Inteligibilidade:** Capacidade do produto de *software* de possibilitar ao usuário compreender se o *software* é apropriado e como ele pode ser usado para tarefas e condições de uso específicas.
- **Apreensibilidade:** Capacidade do produto de *software* de possibilitar ao usuário aprender sua aplicação.
- **Operacionalidade:** Capacidade do produto de *software* de possibilitar ao usuário operá-lo e controlá-lo.
- **Atratividade:** Capacidade do produto de *software* de ser atraente ao usuário.
- **Conformidade relacionada à confiabilidade:** Capacidade do produto de *software* de estar de acordo com normas, convenções, guias de estilo ou regulamentações relacionadas à usabilidade.

Eficiência: Capacidade do produto de *software* de apresentar desempenho apropriado, relativo à quantidade de recursos usados, sob condições especificadas.

- **Comportamento em relação ao tempo:** Capacidade do produto de *software* de fornecer tempos de resposta e de processamento e taxas de transferência apropriados, quando o *software* executa suas funções, sob condições estabelecidas.
- **Utilização de recursos:** Capacidade do produto de *software* de usar tipos e quantidades apropriados de recursos, quando o *software* executa suas funções sob condições estabelecidas.
- **Conformidade relacionada à eficiência:** Capacidade do produto de *software* de estar de acordo com normas e convenções relacionadas à eficiência.

Manutenibilidade: Capacidade do produto de *software* de ser modificado. As modificações podem incluir correções, melhorias ou adaptações do *software* devido a mudanças no ambiente e nos seus requisitos ou especificações funcionais.

- **Analisabilidade:** Capacidade do produto de *software* de permitir o diagnóstico de deficiências ou causas de falhas no *software*, ou a identificação de partes a serem modificadas.
- **Modificabilidade:** Capacidade do produto de *software* de permitir que uma modificação especificada seja implementada.

- **Estabilidade:** Capacidade do produto de *software* de evitar efeitos inesperados decorrentes de modificações no *software*.
- **Testabilidade:** Capacidade do produto de *software* de permitir que o *software*, quando modificado, seja validado.
- **Conformidade relacionada à manutenibilidade:** Capacidade do produto de *software* de estar de acordo com normas ou convenções relacionadas à manutenibilidade.

Portabilidade: Capacidade do produto de *software* de ser transferido de um ambiente para outro.

- **Adaptabilidade:** Capacidade do produto de *software* de ser adaptado para diferentes ambientes especificados, sem necessidade de aplicação de outras ações ou meios além daqueles fornecidos para essa finalidade pelo *software* considerado.
- **Capacidade para ser instalado:** Capacidade do produto de *software* para ser instalado em um ambiente especificado.
- **Coexistência:** Capacidade do produto de *software* de coexistir com outros produtos de *software* independentes, em um ambiente comum, compartilhando recursos comuns.
- **Capacidade para substituir:** Capacidade do produto de *software* de ser usado em substituição a outro produto de *software* especificado, com o mesmo propósito e no mesmo ambiente.
- **Conformidade relacionada à portabilidade:** Capacidade do produto de *software* de estar de acordo com normas ou convenções relacionadas à portabilidade.

Ressalta-se, por fim, que os requisitos de qualidade dependem da necessidade de uso e podem ser especificados pelas métricas externas e, também, por métricas internas.

2.2.2 Métricas de qualidade externas

As métricas de qualidade externa incluem requisitos derivados das necessidades de qualidade dos usuários, como os requisitos de qualidade em uso. Os requisitos de qualidade externa são usados como meta para validação em vários estágios durante o desenvolvimento. Assim, as métricas externas oferecem a usuários, avaliadores, executores de teste e

desenvolvedores os benefícios de poderem avaliar a qualidade do produto de *software* durante seu teste ou operação.

É conveniente salientar que os requisitos de qualidade externa, abrangidos pelas características de qualidade definidas nesta parte da NBR ISO/IEC 9126, podem estar declarados na especificação de requisitos de qualidade usando métricas externas; e convém que sejam traduzidos em requisitos de qualidade interna e que sejam usados como critérios quando da avaliação do produto.

Métricas externas empregam medidas de um produto de *software* provenientes de capacidades do comportamento do sistema do qual o *software* é uma parte, por meio de teste, operação e observação do *software* executável ou do sistema. Antes de se adquirir ou utilizar um produto de *software*, convém que ele seja avaliado utilizando-se métricas baseadas nos objetivos de negócio e relacionadas ao uso, exploração e gestão do produto num ambiente técnico e organizacional especificado.

2.2.3 Métricas de qualidade internas

As métricas de qualidade internas especificam o nível de qualidade requerido sob o ponto de vista interno do produto. São usadas para especificar as propriedades dos produtos intermediários, podendo incluir modelos estáticos e dinâmicos, outros documentos e código-fonte. Oferecem a usuários, avaliadores, executores de teste e desenvolvedores os benefícios de poderem avaliar a qualidade do produto de *software* e considerar questões relativas à qualidade bem antes de um determinado *software* ser executado.

Requisitos de qualidade interna podem ser usados como metas para validação em vários estágios de desenvolvimento, podendo ser utilizados para definir estratégias e critérios de avaliação e de verificação. Isto pode incluir o uso de métricas adicionais (por exemplo, reusabilidade), as quais estão fora do escopo da NBR ISO/IEC 9126. Convém que requisitos de qualidade interna sejam definidos quantitativamente usando métricas internas.

Os detalhes da qualidade do *software* podem ser melhorados durante a implementação do código, revisão e teste, mas a natureza fundamental da qualidade do produto de *software*, representada pela qualidade interna, mantém-se inalterada, a menos que seja reprojetaada.

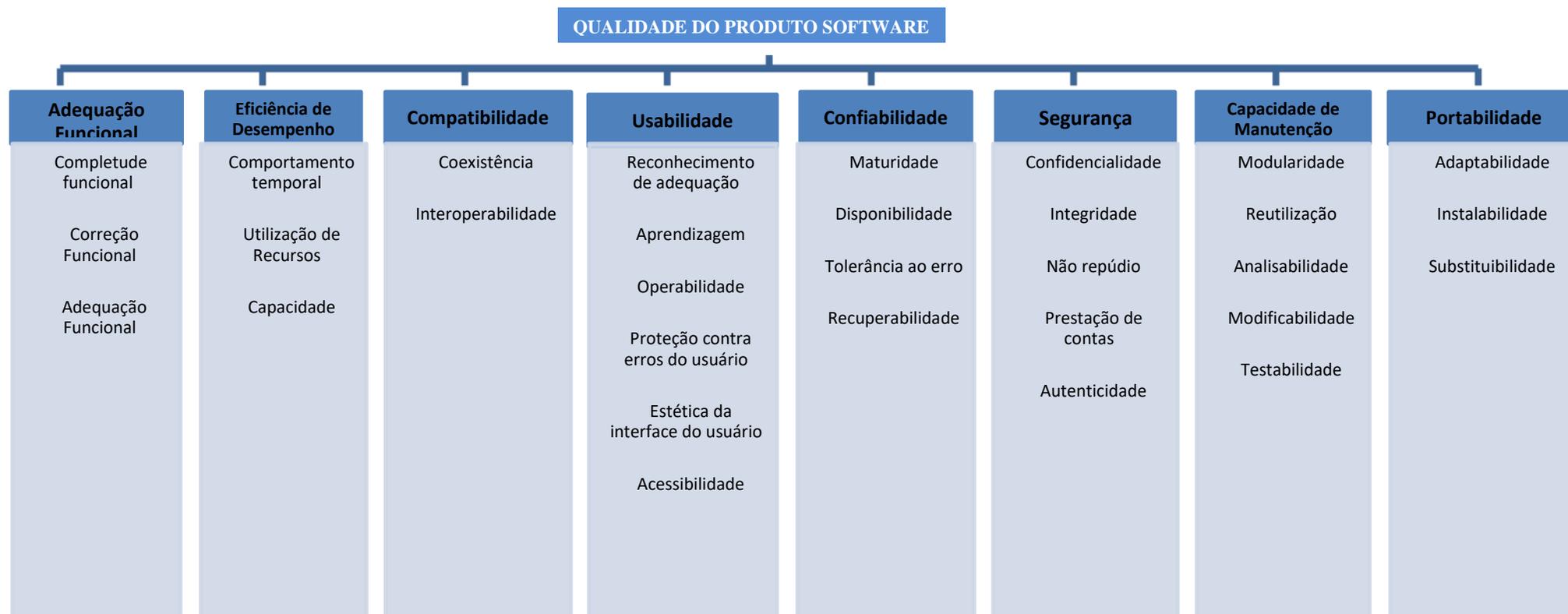
2.2.4 Métricas da ISO/IEC 25010: 2011

O modelo de qualidade ISO/IEC 25010: 2011 define quais as características que todo *software* deve ter de forma a alcançar um nível muito alto de qualidade. Este Padrão Internacional revisa a ISO / IEC 9126-1 e incorpora as mesmas características de qualidade de *software* com algumas alterações, tais como:

- 1 - O escopo dos modelos de qualidade foi estendido para incluir sistemas de computador e qualidade em uso de uma perspectiva de sistema;
- 2 - A cobertura de contexto foi adicionada como uma característica de qualidade em uso, com subcaracterísticas de completude e flexibilidade de contexto;
- 3 - A segurança foi adicionada como uma característica, em vez de uma subcaracterística de funcionalidade, com subcaracterísticas de confidencialidade, integridade, não repúdio, responsabilidade e autenticidade;
- 4 - Compatibilidade (incluindo interoperabilidade e coexistência) foi adicionada como uma característica. As seguintes subcaracterísticas foram adicionadas: integridade funcional, capacidade, proteção contra erros do usuário, acessibilidade, disponibilidade, modularidade e capacidade de reutilização.
- 5 - As subcaracterísticas de conformidade foram removidas, já que a conformidade com leis e regulamentos faz parte dos requisitos gerais do sistema, em vez de especificamente parte da qualidade. Muitas características e subcaracterísticas receberam nomes mais precisos.

A ISO/IEC 25010:2011 apresenta oito características de qualidade (Adequação Funcional, Eficiência de Desempenho, Compatibilidade, Usabilidade, Confiabilidade, Segurança, Capacidade de Manutenção e Portabilidade), as quais são, por sua vez, subdivididas em subcaracterísticas, conforme se pode observar na figura a seguir.

Figura 2.5 - ISO/25010: 2011



Fonte: Modelo ISO/IEC 25010:2011.

Descreve-se, na sequência, as características e as subcaracterísticas da ISO/IEC 25010 apresentadas na figura anterior. Para isso, o autor desta tese traduziu o texto¹² após consultar o site oficial da ISO 25000 STANDARDS¹³.

Adequação Funcional: Esta característica representa o grau em que um produto ou sistema fornece funções que atendem às necessidades declaradas e implícitas, quando usado sob condições especificadas. Esta característica é composta pelas seguintes subcaracterísticas:

- **Completude Funcional:** Grau em que o conjunto de funções cobre todas as tarefas e objetivos do usuário especificados.
- **Correção Funcional:** Grau de correção funcional em que as funções fornecem os resultados corretos com o grau de precisão necessário.
- **Adequação Funcional:** Grau em que as funções facilitam a realização de tarefas e objetivos especificados.

Eficiência de Desempenho: Essa característica representa o desempenho em relação à quantidade de recursos usados nas condições estabelecidas. Esta característica é composta pelas seguintes subcaracterísticas:

- **Comportamento temporal:** Grau em que os tempos de resposta e processamento e taxas de rendimento de um produto ou sistema, quando no desempenho de suas funções, atendem aos requisitos.
- **Utilização de Recursos:** Grau em que as quantidades e tipos de recursos utilizados por um produto ou sistema, ao realizar suas funções, atendem aos requisitos.
- **Capacidade:** Grau em que os limites máximos do produto ou sistema, parâmetro atendem aos requisitos.

Compatibilidade: O grau em que um produto, sistema ou componente pode trocar informações com outros produtos, sistemas ou componentes e/ou executar suas funções necessárias enquanto compartilha o mesmo ambiente de *hardware* ou *software*. Esta característica é composta pelas seguintes subcaracterísticas:

¹² Tradução livre do autor.

¹³ Disponível em: <https://iso25000.com/index.php/en/iso-25000-standards/iso-25010?start=6>. Acesso em: outubro de 2021.

- **Coexistência:** Grau em que um produto pode executar suas funções necessárias de forma eficiente, enquanto compartilha um ambiente e recursos com outros produtos, sem impacto prejudicial em qualquer outro produto.
- **Interoperabilidade:** Grau em que dois ou mais sistemas, produtos ou componentes podem trocar informações e usar informações que foram trocadas.

Usabilidade: O grau em que um produto ou sistema pode ser usado por usuários específicos para atingir objetivos com eficácia, eficiência e satisfação em um contexto de uso. Esta característica é composta pelas seguintes subcaracterísticas:

- **Reconhecimento de adequação:** Grau em que os usuários podem reconhecer se um produto ou sistema é apropriado para suas necessidades.
- **Aprendizagem:** Grau em que um produto ou sistema permite ao usuário aprender como usá-lo com eficiência em situações de emergência.
- **Operabilidade:** Grau em que um produto ou sistema é fácil de operar, controlar e é apropriado para usar.
- **Proteção contra erros do usuário:** Grau de proteção contra erros do usuário em que um produto ou sistema protege os usuários contra erros.
- **Estética da interface do usuário:** Grau em que uma interface de usuário permite uma interação agradável e satisfatória para o usuário.
- **Acessibilidade:** Grau em que um produto ou sistema pode ser usado por pessoas com a mais ampla gama de características e recursos, a fim de atingir um objetivo específico em um contexto de uso especificado.

Confiabilidade: Grau em que um sistema, produto ou componente executa funções específicas sob condições especificadas por um período especificado. Esta característica é composta pelas seguintes subcaracterísticas:

- **Maturidade:** Grau em que um sistema, produto ou componente atende às necessidades de confiabilidade em operação normal.
- **Disponibilidade:** Grau em que um produto ou sistema está operacional e acessível quando necessário para uso.
- **Tolerância ao erro:** Grau em que um sistema, produto ou componente opera conforme pretendido, apesar da presença de *hardware* ou falhas de *software*.

- **Recuperabilidade:** Grau em que, em caso de interrupção ou falha, um produto ou sistema pode recuperar os dados afetados diretamente e restabelecer o estado desejado do sistema.

Segurança: Grau em que um produto ou sistema protege informações e dados, para que pessoas ou outros produtos ou sistemas tenham grau de acesso aos dados apropriados para seus tipos e níveis de autorização. Esta característica é composta pelas seguintes subcaracterísticas:

- **Confidencialidade:** Grau em que o protótipo garante que os dados são acessíveis apenas para aqueles autorizados a ter acesso.
- **Integridade:** Grau em que um sistema, produto ou componente impede o acesso não autorizado ou modificação de programas de computador ou dados.
- **Não repúdio:** Grau em que se pode provar a ocorrência de ações ou eventos, de modo que não podem ser repudiados mais tarde.
- **Prestação de contas:** Grau em que as ações de uma entidade podem ser rastreadas exclusivamente para a entidade.
- **Autenticidade:** Grau em que a identidade de um sujeito ou recurso pode ser comprovada como sendo reivindicado.

Capacidade de Manutenção: Esta característica representa o grau de eficácia e eficiência com que um produto ou sistema pode ser modificado para melhorá-lo, corrigi-lo ou adaptá-lo às mudanças do ambiente e dos requisitos. Esta característica é composta pelas seguintes subcaracterísticas:

- **Modularidade:** Grau em que um sistema ou programa de computador é composto de componentes discretos, de modo que uma mudança para um componente tem impacto mínimo em outros componentes.
- **Reutilização:** Grau em que um ativo pode ser usado em mais de um sistema ou na construção de outros ativos.
- **Analisabilidade:** Grau de eficácia e eficiência com o qual é possível avaliar o impacto sobre um produto ou sistema de uma alteração pretendida em uma ou mais de suas partes, ou para diagnosticar um produto quanto a deficiências ou causas de falhas, ou para identificar peças a serem modificadas.

- **Modificabilidade:** Grau em que um produto ou sistema pode ser modificado, de forma eficiente, sem a introdução de feitos ou degradação da qualidade do produto existente.
- **Testabilidade:** Grau de eficácia e eficiência com o qual os critérios de teste podem ser estabelecidos para um sistema, produto ou componente e testes podem ser realizados para determinar se esses critérios foram atendidos.

Portabilidade: Grau de eficácia e eficiência com que um sistema, produto ou componente pode ser transferido de um *hardware*, *software* ou outro ambiente operacional ou de uso para outro. Esta característica é composta pelas seguintes subcaracterísticas:

- **Adaptabilidade:** Grau em que um produto ou sistema pode ser eficaz e eficientemente adaptado para diferentes ou em evolução *hardware*, *software* ou outros ambientes operacionais ou de uso.
- **Instalabilidade:** Grau de eficácia e eficiência em que um produto ou sistema pode ser instalado com sucesso e/ou desinstalado em um ambiente especificado.
- **Substituibilidade:** Grau em que um produto pode substituir outro produto de *software* especificado para a mesma finalidade no mesmo ambiente.

2.2.5 Qualidade em uso

A qualidade de uso do produto *software* permite elencar metas específicas, como eficácia, segurança e satisfação, de acordo com o contexto que deverá ser avaliado sob a perspectiva do usuário. Os atributos de qualidade em uso da ISO/IEC 25010 são categorizados em cinco características: Eficácia, Eficiência, Satisfação, Livre de riscos, Cobertura de Contexto, como representado tabela a seguir.

Tabela 2.1 - Qualidade em uso

Características	Subcaracterísticas	Definição
Eficácia		Precisão e integridade com as quais os usuários alcançam objetivos específicos.
Eficiência		Recursos gastos em relação à precisão e integridade com que os usuários atingem as metas.
Satisfação	Utilidade	Grau em que um usuário está satisfeito com a percepção de sua realização de objetivos pragmáticos, incluindo os resultados do uso e as consequências do uso.

	Confiança	Grau em que um usuário ou outra parte interessada tem confiança de que um produto ou sistema se comportará conforme pretendido.
	Prazer / satisfação	Grau em que um usuário obtém prazer em atender às suas necessidades pessoais.
	Conforto	Grau de satisfação do usuário com o conforto físico.
Livre de riscos	Mitigação de Risco Econômico	Grau em que um produto ou sistema mitiga o risco potencial para a situação financeira, operação eficiente, propriedade comercial, reputação ou outros recursos nos contextos de uso pretendidos.
	Mitigação de riscos de saúde e segurança	Grau em que um produto ou sistema mitiga o risco potencial para as pessoas nos contextos de uso pretendidos.
	Mitigação de risco ambiental	Grau em que um produto ou sistema mitiga o risco potencial para a propriedade ou o meio ambiente nos contextos de uso pretendidos.
Cobertura de Contexto	Completeness do Contexto	Grau em que um produto ou sistema pode ser usado com eficácia, eficiência, isenção de riscos e satisfação em todos os contextos de uso especificados.
	Flexibilidade	Grau em que um produto ou sistema pode ser usado com eficácia, eficiência, isenção de risco e satisfação em contextos além daqueles inicialmente especificados nos requisitos.

Fonte: ISO/IEC 25010.

2.2.6 Escolha das métricas de avaliação de *softwares*

Não há uma regra para a determinação da seleção de métricas, pois a seleção da categoria e da subcategoria proposta pela ISO/IEC 25010 está relacionada aos objetivos de avaliação de cada *software*, que cada um deseja avaliar. Assim, tendo em vista o objetivo deste estudo, que consiste em desenvolver um *framework* com métricas de avaliação de agentes conversacionais educacionais, justifica-se a apresentação das métricas nesta subseção como possibilidades de escolhas para a avaliação. É possível optar por uma única característica e uma única subcaracterística propostas no modelo de qualidade para qualidade externa e interna relacionadas com a intenção de avaliação, ou seja, a avaliação de um determinado agente conversacional estará com foco na Eficiência e Confiança.

Cabe mencionar que o uso de métricas para avaliação de agentes conversacionais é importante para avaliar, apontar e sugerir diversos aspectos para fins de comparação entre os agentes, de comparação das tecnologias e de possibilidades e apontamentos para melhorias futuras, com base em métricas de avaliação de *softwares* e métricas inspiradas em trabalhos e estudos internacionais que investigaram os *softwares*.

Nesta seção, são apresentadas as métricas baseadas na NBR ISO/IEC 25010 e em estudos internacionais, as quais serão a base para a construção do *framework* proposto na seção

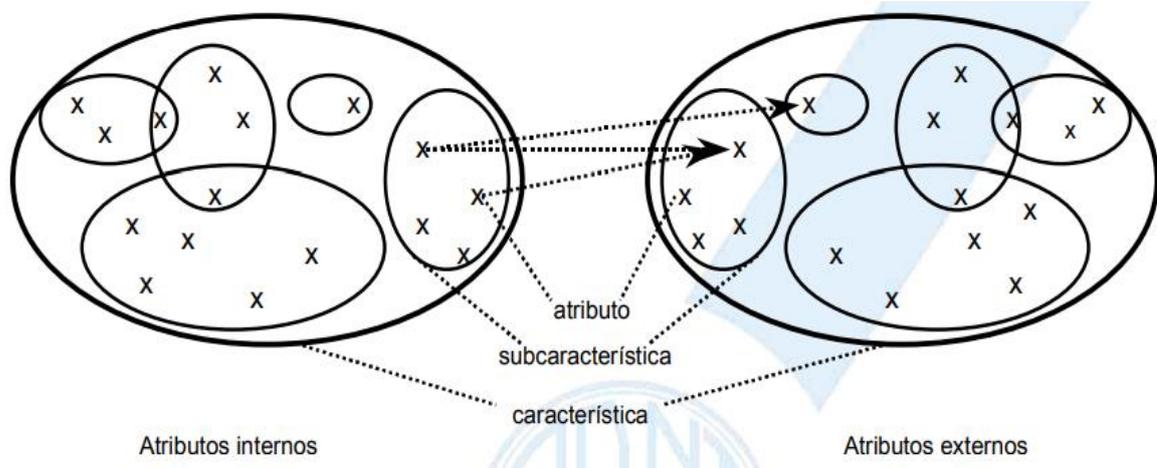
4 desta tese. De acordo com a NBR ISO/IEC 25010, há uma variedade de requisitos que devem ser considerados no momento de uma avaliação. Assim, as necessidades podem ser especificadas a partir de uma característica ou de mais critérios, baseados no modelo da NBR ISO/IEC 25010, como:

- Um determinado usuário optar por um *software*, com base nas métricas de qualidade do produto;
- Uma avaliação de um *software* ser realizada com base nas medidas externas de funcionalidade, usabilidade, eficiência;
- Uma equipe de suporte poder avaliar o *software*, considerando as métricas de manutenibilidade;
- Um analista de sistemas avaliar os *softwares*, tendo como guia as métricas de portabilidade;
- Um desenvolvedor avaliar o *software* em relação aos critérios, utilizando medidas internas de uma ou várias características de qualidade.

Observa-se que não há uma regra específica para examinar um determinado produto de *software*, mas um grupo de características e subcaracterísticas que apontam as métricas de avaliação. Assim, segundo o manual da NBR-ISO-9126-1, costumeiramente, três níveis são utilizados para avaliação: métricas internas, externas e de qualidade. Porém, atender aos critérios de medidas internas, em geral, não é suficiente para garantir o atendimento aos critérios de medidas externas; assim como atender aos critérios de medidas externas não é suficiente para garantir o atendimento aos critérios de qualidade em uso.

O fato é que alguns níveis de determinados atributos internos instigam os de alguns atributos externos, de maneira que há tanto um aspecto externo quanto um aspecto interno na maior parte das características. A título de exemplo, confiabilidade pode ser dimensionada, externamente, observando-se o número de falhas, num dado período de execução, durante um experimento de uso do *software* e, internamente, inspecionando-se as especificações detalhadas e o código-fonte para avaliar-se o nível de tolerância a falhas. Os atributos internos são tidos como indicadores dos atributos externos, como se pode observar na figura 2.11.

Figura 2.6 - Características e subcaracterísticas e atributos de qualidade de *software*



Fonte: NBR-ISO IEC-9126-1.

Conforme contata-se na figura 2.11, um atributo interno pode influenciar uma ou mais características, e uma característica pode ser influenciada por mais de um atributo. No exemplo da figura, todos os atributos de qualidade do produto de *software* são classificados numa estrutura hierárquica, em árvore, de características e subcaracterísticas. O nível mais alto desta estrutura consiste em características de qualidade e o nível mais baixo consiste em atributos de qualidade de *software*. A hierarquia não é perfeita, pois alguns atributos podem contribuir para mais de uma subcaracterística.

Como se pode verificar, a seção 2.2 procurou apresentar as métricas para avaliação e desenvolvimento do produto *software*, com base na NBR-ISO 9126-1 e na NBR-ISO 25010, pois as métricas de avaliação de *software*, sejam internas, externas ou de qualidade poderão ser utilizadas para todo e qualquer desenvolvimento e avaliação de qualquer produto de *software*. Assim, justifica-se a relevância em trazer para o presente estudo os respectivos parâmetros e métricas de avaliação de *softwares* desenvolvidos e utilizados pela NBR-ISO IEC-9126-1 e ISO/IEC 25010.

A próxima seção trata especificamente de trabalhos e estudos relacionados à avaliação de agentes conversacionais.

2.3 TRABALHOS RELACIONADOS

Considerando a proposta desta tese de criar um *framework* com sugestões de métricas de *software*, a fim de sugerir o que um *chatbot* educacional pode e deve ter para ser interessante,

de modo que possa ser utilizado como referência para a construção de *chatbots* desenvolvidos para o contexto educacional, esta subseção apresenta trabalhos e estudos relacionados à utilização, ao desenvolvimento e aos apontamentos de métricas, envolvendo, especificamente, o contexto de agentes conversacionais.

2.3.1 Estudo 1 - Projeto e avaliação do agente conversacional SARA

O estudo proposto por Niculescu et al. (2014) tem como foco o *design* e a avaliação do SARA, um agente conversacional para o domínio turístico. Este *chatbot* apresenta grande número de características diferentes e únicas: interação do diálogo falado, orquestração do diálogo, informações dependentes do contexto, um avatar animado e suporte para diferentes tipos de diálogo, ou seja, bate-papo, resposta a perguntas específicas e gerais, além de diálogos orientados para tarefas.

O SARA foi construído como uma resposta à crescente demanda por assistência turística pessoal e oferece uma solução confortável para quem deseja explorar a cidade e não tem guia turístico por perto. Atualmente, o agente possui duas implementações: como cliente da *web* e como aplicativo de celular. Para avaliar as interfaces do agente conversacional SARA, foi utilizada a avaliação heurística. Trata-se de um método qualitativo de inspeção que envolve especialistas em usabilidade, passando pela interface e executando várias tarefas, para julgar se a interface do usuário segue os princípios estabelecidos de usabilidade (heurística). Quatro especialistas participaram da avaliação heurística. Eles receberam algumas tarefas para executar, fazer saudações, questionar sobre os locais de seu interesse, recomendação de restaurantes, perguntas gerais sobre clima, moeda, horário de funcionamento e solicitação de aplicativo para chamar um local de interesse específico. Além disso, estavam livres para brincar com a aplicação e experimentá-la enquanto desejassem.

A avaliação usando heurísticas trouxe ao estudo de Niculescu et al. (2014) muitas informações sobre problemas de usabilidade e interação de falhas de *design*, bem como apontamentos para novas sugestões de melhorias. Em relação à visibilidade do status do sistema, uma preocupação era que o aplicativo móvel perdesse sua conexão com o servidor, apresentando ao usuário um erro congelado na página. Entre os erros, foi apontado que a página contém um botão de atualização, o qual parece não funcionar e os especialistas alertaram que os usuários provavelmente ficariam confusos se não soubessem o que está acontecendo com o aplicativo. Também foi indicado o problema de desativação do discurso do avatar. Neste caso,

os especialistas recomendam destacar o balão de fala em cores para sinalizar que o avatar está respondendo à pergunta.

Além disso, funcionalidades do sistema, como as informações de provisão na forma de mapas, informações da *web* e fotos, podem ser confusos para novos usuários. Por exemplo, um especialista estava se perguntando se o sistema pode responder perguntas sobre instruções, se estiver no modo de imagem. Outro especialista estava confuso sobre como usar uma funcionalidade chamada de *scanner*, em que ficou na dúvida em saber se tinha que pressionar uma determinada tecla para iniciar o processo de digitalização. Esta informação poderia ser fornecida em uma seção de **ajuda e documentação** para ser apresentada ao usuário no momento da instalação do aplicativo pela primeira vez. Sobre a **correspondência entre sistema e mundo real**, o sistema deve levar em conta o fato de que os usuários esperam que as informações sobre direção sejam apresentadas semelhantes ao *Google Maps*. Atualmente, o aplicativo apenas indica o caminho a ser seguido, mas não oferece informações adicionais sobre como chegar a um local de ônibus, metrô etc.

Uma característica positiva elogiada pelos especialistas foi a capacidade do sistema para enviar mensagens de texto e fazer chamadas telefônicas; uma funcionalidade que lembrava uma boa recepcionista em serviço. Já o controle e a liberdade do usuário são limitados, pois interrupções não são permitidas enquanto o avatar está falando, ou enquanto o *website* está carregado. Isso acontece porque os usuários não têm a opção de um botão de desfazer, interromper ou reiniciar uma consulta quando o sistema demora muito para responder. Um dos especialistas também comentou sobre um problema de consistência menos crítico, que trata das expressões faciais de avatar, pois, segundo ele, o avatar deve mostrar um rosto triste quando não tem resposta e sorrir quando encontra o caminho certo. Caso contrário, verifica-se a inconsistência entre os comportamentos apresentados diante do resultado do diálogo.

Segundo três especialistas que avaliaram o respectivo trabalho relacionado, novos usuários provavelmente se perguntariam sobre a maneira com que devem falar com o sistema, isto é, se eles podem usar frases completas ou apenas palavras. Constatou-se a necessidade em explicar aos usuários como falar / interagir com o aplicativo quando eles o usam pela primeira vez. Por fim, os *feedbacks* obtidos no resultado do estudo dos referidos autores apontam sugestões de melhorias nos aplicativos em termos de serviço, desempenho e usabilidade. Assim, este estudo avaliou as métricas de usabilidade, funcionalidade e confiabilidade.

2.3.2 Estudo 2 - Avaliando a qualidade dos *chatbots*

O estudo proposto por Radziwill e Benton (2017) apresenta abordagens de avaliação da qualidade e desenvolve um método de avaliação da qualidade baseado em atributos. O objetivo dos autores foi examinar a literatura para fornecer uma revisão abrangente dos atributos de qualidade para *chatbots* e agentes conversacionais, e identificar abordagens apropriadas de garantia da qualidade. Eles extraíram atributos de qualidade de 32 artigos. O diferencial desta pesquisa são as abordagens de avaliação da qualidade, as quais são revisadas mediante um método baseado em eficiência, eficácia e satisfação.

As propostas de Radziwill e Benton (2017) são apresentadas no quadro 1, que relaciona as categorias, as subcategorias e os respectivos atributos de qualidade.

Quadro 1: Atributos de qualidade de *chatbots* e agentes conversacionais

Categoria Eficiência		
Subcategoria	Atributo de qualidade	Referências utilizadas por Radziwill e Benton (2017):
<ul style="list-style-type: none"> ● Performance 	<ul style="list-style-type: none"> ● Degradação ● Robustez à manipulação ● Robustez para entrada inesperada ● Evitar declarações inapropriadas e seja capaz de executar controle de danos ● Alocação eficaz de funções, fornece canais de escalação apropriados para humanos 	<ul style="list-style-type: none"> ● Cohen & Lane (2016) ● Thieltges (2016) ● Kluwer (2011) ● Morrissey e Kirakowski (2013) ● Staven (2017)
Categoria Eficácia		
Subcategoria	Atributo de qualidade	Referências utilizadas por Radziwill e Benton (2017):
<ul style="list-style-type: none"> ● Funcionalidade 	<ul style="list-style-type: none"> ● Síntese precisa da fala ● Interpreta comandos com precisão ● Usar graus adequados de formalidade, registro linguístico ● Precisão linguística dos resultados ● Executar tarefas solicitadas ● Facilitar transações e acompanhar relatórios de status ● Facilidade de uso geral ● Envolver-se na solução de problemas <i>on-the-fly</i> ● Contém amplitude de conhecimento, é flexível na interpretação 	<ul style="list-style-type: none"> ● Kuligowska (2015) ● Eeuwen (2017) ● Morrissey e Kirakowski (2013) ● Wallace (2003) ● Ramos (2017) ● Eeuwen (2017) ● Solomon (2017) ● Cohen & Lane (2016)
<ul style="list-style-type: none"> ● Humanidade 	<ul style="list-style-type: none"> ● Passar no teste de Turing ● Não precisa passar no teste de Turing ● Transparente à inspeção, divulga sua identidade no <i>chatbot</i> ● Incluir erros para aumentar o realismo ● Interação convincente, satisfatória e natural 	<ul style="list-style-type: none"> ● Weizenbaum (1966); Wallace (2003) ● Ramos (2017) ● Bostrom e Yudkowski (2014) ● Coniam (2014)

	<ul style="list-style-type: none"> ● Capaz de responder a perguntas específicas ● Capaz de manter a discussão temática 	<ul style="list-style-type: none"> ● Morrissey e Kirakowski (2013)
Categoria Satisfação		
Subcategoria	Atributo de qualidade	Referências utilizadas por Radziwill e Benton (2017):
<ul style="list-style-type: none"> ● Afetividade 	<ul style="list-style-type: none"> ● Dar saudações, transmitir personalidade ● Dar dicas de conversação ● Fornecer informações emocionais por meio de tom, inflexão e expressividade ● Exalar calor e autenticidade ● Tornar as tarefas mais divertidas e interessantes ● Divertir-se / ou permitir que o participante desfrute da interação ● Ler e responder ao humor do participante humano 	<ul style="list-style-type: none"> ● Morrissey e Kirakowski (2013) ● Pauletto et al. (2013) ● Solomon (2017) ● Eeuwen (2017) ● Ramos (2017) ● Meira e Canuto (2015)
<ul style="list-style-type: none"> ● Ética e Comportamento 	<ul style="list-style-type: none"> ● Respeito, inclusão e preservação da dignidade (vinculado à escolha do conjunto de treinamento) ● Ética e conhecimento cultural dos usuários ● Proteger e respeitar a privacidade ● Não-recepção ● Sensibilidade às preocupações sociais e de segurança ● Confiabilidade (ligada à qualidade percebida) ● Consciência das tendências e contexto social 	<ul style="list-style-type: none"> ● Neff e Nagy (2016) ● Applin e Fischer (2015) ● Eeuwen (2017) ● Isaac e Bridewell (2014) ● Miner et al. (2016) ● Herzum et al. (2002) ● Vetter (2002)
<ul style="list-style-type: none"> ● Acessibilidade 	<ul style="list-style-type: none"> ● Responder a sugestões sociais ou a falta delas ● Poder detectar significado ou intenção ● Atender às necessidades diversas, como tempo extra de resposta e interface de texto 	<ul style="list-style-type: none"> ● Morrissey e Kirakowski (2013) ● Wilson et al. (2017) ● Radziwill e Benton (2017)

Fonte: Radziwill e Benton (2017).

Salienta-se que as referências citadas na terceira coluna do Quadro 1 não foram pesquisadas diretamente pelo autor da presente tese, mas, sim, foram indicadas por Radziwill e Benton (2017) e as respectivas referências podem ser encontradas no artigo referenciado neste estudo de tese.

Registra-se que as métricas extraídas dos trabalhos relacionados apresentados no quadro 1, conforme pesquisa de Radziwill e Benton (2017), foram utilizadas para o desenvolvimento do *framework* da respectiva tese, e foram comparadas com as métricas da NBR-ISO IEC9126-1 e com as métricas apresentadas nos demais trabalhos relacionados.

2.3.3 Estudo 3 - Avaliação de agentes conversacionais incorporados

O estudo de Brandão, Reis e Rocha (2013) propõe uma plataforma que permite avaliar o desempenho de um agente conversacional incorporado e a influência de vários elementos multimídias distintos. Usando o "Jogo de Imitação", conforme descrito por Alan Turing, como

ponto de partida e motivação, a plataforma permite a configuração de recursos que influenciam o ambiente virtual.

Para avaliar a influência dos recursos multimídia para complementar o agente conversacional, foi desenvolvida uma plataforma que pode ser acessada remotamente via Internet. Os usuários podem se registrar remotamente na plataforma e ter diferentes sessões de bate-papo inspiradas nas definidas por Turing. Foi usado o robô ALICE, com conjuntos padrão AIML. A plataforma permite a configuração de várias características quanto à aparência e comportamento da interface de usuário. A aparência da interface para cada sessão de conversa pode ser selecionada de um conjunto predefinido. Todas as interfaces têm uma introdução e pesquisa de usuários correspondentes para fornecer um uso guiado e coletar informações experimentais de acordo com o tipo de interface.

No experimento realizado, foi determinado que o tempo de conversação deveria ser curto. Depois do registro, os usuários são convidados a participar de 6 sessões de 2 minutos cada. A primeira conversa da sessão é com um moderador humano e a segunda com um agente conversacional, seguidas de sessões com humano e agente e, finalmente, mais duas sessões, uma com moderador e outra com agente.

O experimento contou com 52 usuários e o tema da conversa foi o campo de Inteligência Artificial. De acordo com as informações do registro, os usuários foram divididos, sendo 34 usuários do sexo masculino e 18 do sexo feminino; todos com um conhecimento médio de Tópicos de Inteligência Artificial. O uso da plataforma resultou em 132 consultas, cada uma representando uma sessão de bate-papo. Houve 264 minutos de conversa, das quais 202 estavam entre um humano e o agente. Apenas 31 sessões estavam entre 2 humanos. Isso contabiliza 751 linhas de conversação e um total de 115 usos do controle de identificação. Com o uso desse controle, os usuários indicaram que estavam falando com um agente por 79 vezes e indicaram que acreditavam que estavam falando com um humano por 36 vezes.

Os resultados possibilitaram verificar a influência positiva dos elementos de multimídia usados com agentes conversacionais. Apesar disso, não foi possível verificar um aumento na satisfação do usuário relacionado diretamente com os elementos adicionados à interface, mas pôde-se manter os níveis de satisfação até os estágios finais da participação do usuário.

Quanto ao uso das emoções, avatares que expressam emoções claramente melhora a experiência de uso, incentivando o usuário a ser mais expressivo. Mesmo que esse resultado tenha sido obtido usando interfaces que reproduzem o jogo de imitação, a plataforma foi projetada para inserir facilmente novas interfaces e configurações.

Por certo, uma possibilidade interessante seria desenvolver uma interface e base de conhecimento em que o agente atuaria como cliente atendente e comparar com os dados existentes. O aplicativo desenvolvido é um passo à frente em ter uma plataforma para avaliar o desempenho dos agentes conversacionais, pois permite o uso da internet para coletar dados e foi implementado para que possa ser reutilizado por outros pesquisadores.

Neste estudo relacionado, foram abordadas especificamente as métricas Eficiência e Testes de conversação, métricas essas estudadas para a proposta do desenvolvimento do *framework* para avaliar agentes conversacionais educacionais.

2.3.4 Estudo 4 - Conversas reais com inteligência artificial

Hill, Ford e Farreras (2015) analisaram como a comunicação muda quando as pessoas conversam com um agente inteligente e não com outro ser humano. Foram comparadas 100 conversas por mensagens instantâneas e 100 trocas com o popular *chatbot Cleverbot* em sete dimensões: palavras por mensagem, palavras por conversa, mensagens por conversa, exclusividade de palavras e uso de palavras, taquigrafia e emoticons.

O *Cleverbot* foi projetado para simular conversas naturais e aprender a partir da interação humana, contando com o *feedback* para se comunicar. Usando uma interface on-line síncrona individual que lembra um mensageiro instantâneo, um usuário humano digita e aguarda o *Cleverbot* enviar uma resposta. Assim, o agente "aprende" o que os humanos disseram nas conversas anteriores e usa esse conhecimento para determinar como responder em novas conversas.

Na referida pesquisa, os autores compararam 100 conversas aleatórias entre humanos contra 100 conversas humanas aleatórias com o *Cleverbot*. Esperava-se que os perfis linguísticos das conversas de pessoa a pessoa fossem consistentes. Hill, Ford e Farreras (2015) levantaram a hipótese de que as pessoas enviariam menos mensagens, escreveriam menos palavras por mensagem e exibiriam um vocabulário mais limitado ao se comunicarem com um *chatbot* em comparação com a comunicação com outra pessoa. Chegaram a essa hipótese considerando três características identificadas nas conversas humano-*chatbot* que não são encontradas nas interações humano-humano: a) as pessoas têm menos experiência em se comunicar com *chatbots* e não seria confiante e confortável a capacidade de comunicação de um agente conversacional; b) ainda que os *chatbots* sejam projetados para manter conversas, são limitados em sua capacidade de ter extensa discussão direcionada a objetivos e com

profundidade de informações sobre um tópico em particular; e c) não há elemento comum possível na história ou experiência compartilhada no contexto humano-*chatbot*.

Mesmo que usuários específicos tenham tido extensas conversas com um *chatbot*, o agente não tem conhecimento de tais interações e, portanto, é incapaz de fazer referência a elas. Como um resultado, esperava-se que as conversas entre humanos e *chatbot* fossem mais hesitantes, resultando em conversas mais curtas, mensagens mais curtas e um limitado vocabulário. Além disso, os autores também esperavam que a conversa anônima entre homem-máquina poderia levar a um maior uso de palavrões, se comparada a conversas humano-humano.

A partir dos resultados da pesquisa, eles concluíram que as pessoas se comunicavam com o *chatbot* por períodos mais longos (mais curtos também), quando comparados à conversação com outro ser humano. Além disso, a comunicação humano-*chatbot* carecia de grande parte da riqueza de vocabulário encontrada nas conversas entre as pessoas e apresentou um vocabulário envolvendo termos desrespeitosos. Esses resultados sugerem que, embora as habilidades de linguagem humana sejam transferidas facilmente para a comunicação humano-*chatbot*, há diferenças notáveis no conteúdo e na qualidade dessas conversas.

Assim, o referido estudo tem relação com a métrica de avaliação “Testes de conversação”, a qual pode ser utilizada para o desenvolvimento do *framework* proposto nesta tese.

2.3.5 Estudo 5 - Uma avaliação do sistema do agente conversacional

A crescente necessidade de métricas-padrão para avaliar e medir a qualidade das respostas produzidas pelos sistemas de agentes conversacionais, com base em diferentes abordagens e domínios, é proposta por Goh, Kumar e Choon (2016). Para demonstrar tal abordagem, os autores realizaram um estudo em que apresentaram os dados e os resultados obtidos, por meio de uma avaliação realizada em três sistemas de conversação diferentes, a saber: ELIZA (psicoterapeuta rogeriana), ALICE (entidade linguística artificial para computadores na Internet) e AINI (identidade artificial inteligente em linguagem natural).

Nesta pesquisa, utilizou-se o X-MINIPAR, o qual usa uma gramática própria antes de converter o resultado em uma árvore de dependência. Uma avaliação com o *corpus* SUSANNE analisa o texto de jornal com cerca de 500 palavras por segundo, o MINIPAR atinge cerca de 88% de precisão e 80%, no que se refere às relações de dependência.

No estudo, foi utilizada uma abordagem quantitativa para medir os progressos no domínio da análise de cobertura ampla e comparar e avaliar diferentes técnicas de análise. No desenvolvimento de um analisador de cobertura ampla, geralmente, é muito difícil prever as consequências de uma alteração no analisador ou na gramática. A eficiência e a cobertura são frequentemente objetivos conflitantes para um analisador. A avaliação quantitativa pode fornecer informações cruciais para determinar a adequação de um analisador em um domínio específico ou tarefa específica.

Nesta abordagem quantitativa para a avaliação de desempenho, um conjunto de 98 perguntas de estímulo foi usado. Esse conjunto de perguntas foi coletado nas conversas entre quatro juízes e ALICE no concurso de 2001, no Museu da Ciência, em Londres. As perguntas desses juízes foram usadas para simular a conversação com AINI, ALICE, *ALICE Silver Edition* e ELIZA remotamente via internet. O tempo de resposta para cada pergunta submetida a esses quatro agentes conversacionais é registrado, coletado e analisado para uma média e desvio padrão.

Neste contexto, os resultados da pesquisa de Goh, Kumar e Choon (2016) apontam que o tempo de resposta do AINI é realmente semelhante a outros sistemas que parecem incluir recursos e processamento menos exigentes. Os resultados também demonstraram que o tempo de resposta do AINI é consistente, apesar da incerteza no tipo de perguntas. Isso é ilustrado com o padrão relativamente baixo de desvio em comparação com o *ALICE Silver Edition*. O AINI foi construído para trabalhar com questões de domínio aberto, o que é diferente de outros agentes conversacionais que são restritos apenas a um domínio específico, como o terapeuta virtual em ELIZA ou o domínio irrestrito, como no ALICE. Dados do referido estudo mostram que o AINI é comparativamente melhor em termos de qualidade das respostas geradas. Um dos critérios que contribuíram para melhor pontuação de AINI é a capacidade de gerar respostas úteis dinamicamente, usando dois componentes avançados de raciocínio, nomeadamente explicações sobre falhas e respostas dinâmicas de geração para atender à condição, quando não houver respostas disponíveis.

Na abordagem quantitativa para avaliação de desempenho dos agentes conversacionais, o AINI foi melhor em comparação com ELIZA, ALICE e *ALICE Silver Edition*, e a classificação seguiu com a ALICE, *ALICE Silver Edition* e AINI. Embora no AINI o conhecimento das categorias estímulo-resposta tenha 12,9% a mais que o *ALICE Silver Edition*, o tempo de resposta mais lento da AINI ocorreu devido à introdução dos componentes NLUR que melhoram a qualidade das respostas.

Assim, a pesquisa de Goh, Kumar e Choon (2016) tem relação com as métricas de avaliação “Testes de conversação” e “Testes de Respostas”, as quais foram utilizadas para o desenvolvimento do *framework* da respectiva tese.

2.3.6 Estudo 6 - Métricas para o *Chatbot Analytics in 2019*

Segundo o site *AI Multiple*¹⁴, uma métrica é uma medida quantificável usada para rastrear e avaliar o status de um processo específico. De fato, definir as métricas é importante, pois muitos dos recursos do *chatbot* serão medidos a partir das respectivas métricas.

No caso de um *chatbot* recém-criado, as métricas podem variar bastante, afinal, após a implementação do agente conversacional, é preciso acompanhá-lo de perto. Como as expectativas são aumento da eficiência, resposta mais rápida e maior conversão, as empresas precisam definir as métricas corretas. Para tanto, o desempenho do *chatbot* pode ser monitorado e melhorar a eficiência. O blog *AI Multiple* sugere as seguintes métricas para um *chatbot*:

1 - Métricas de usuários

- a) Total de usuários: Captura o número de pessoas, usando seu *chatbot*. Isso é importante porque sua tendência mostra a mudança no número de usuários e, portanto, a quantidade de dados aos quais o *chatbot* foi exposto.
- b) Usuários ativos: Podem ser definidos como as pessoas que leem uma mensagem no *chatbot* em um período de tempo definido. O engajamento não é garantido, mas o conteúdo é visto pelos alunos.
- c) Usuários engajados: São esses usuários que se comunicam com o *chatbot*. Essa métrica é importante, pois o *chatbot* poderá fornecer as estatísticas da conversa. Não faz sentido se ele não puder iniciar a conversa com os usuários.
- d) Novos usuários: Novos usuários serão necessários para manter um número de usuários ativos, além de manter sua base de clientes forte.

2 - Métricas da mensagem

- a) Mensagens Iniciais da Conversa: Trata-se do número de mensagens em que se inicia a interação pelo *chatbot*. É possível obter uma resposta enviando mensagens aos usuários, no entanto, após a saudação inicial, é preciso que os usuários continuem enviando mensagens diretamente para o *chatbot*.

¹⁴ Disponível em: <https://blog.aimultiple.com/chatbot-analytics/>

- b) Mensagens de *chatbot*: Tratam do número total de mensagens enviadas pelo *chatbot* em cada interação. Isso mede a duração de uma conversa entre um usuário e o agente conversacional. Normalmente, deseja-se que o número de mensagens seja alto, mas há uma condição crítica: o *chatbot* precisa responder corretamente.
- c) Mensagens: Esta categoria mostra as mensagens enviadas pelo usuário. É necessário ver se ele se envolve com o *chatbot* ou não. Se essa categoria for significativamente baixa, não é preciso usar um *chatbot*.
- d) Mensagens perdidas (Misses): As mensagens perdidas são aquelas em que o *chatbot* não pode processar. Trata-se de uma métrica-chave no caso de conversas, envolvendo outros países onde o idioma é usado de maneira adaptada com o uso de palavras do idioma nativo (falsos cognatos).
- e) Total de conversas: Número de conversas iniciadas e concluídas com êxito em um determinado dia. Identifica os usuários engajados.
- f) Novas conversas: Número de novas conversas iniciadas. Captura os usuários inexperientes e as conversas iniciadas por aqueles que retornam sobre um assunto, problema ou dúvidas.

3 - Métricas de *chatbot*

- a) Taxa de Retenção: Esta é a porcentagem de usuários que retornam ao uso do *chatbot* em determinado período de tempo. Isso pode ser conseguido principalmente fornecendo um *chatbot* de alta qualidade que atenda às expectativas e necessidades dos alunos.
- b) Taxa de conclusão da meta: Captura a porcentagem de envolvimento bem-sucedido por meio do *chatbot*. Os usuários provavelmente tentarão acessar informações ou serviços diferentes. Assim, é possível mostrar o número de vezes que o *chatbot* processou, com êxito, a entrada e forneceu as informações solicitadas.
- c) Tempo de conclusão da meta / mensagens / toques: Os *chatbots* precisam fornecer uma experiência perfeita e eficiente e possuem substitutos, como aplicativos ou páginas da *web*. Minimizar o esforço para concluir uma meta pode melhorar a experiência do usuário.
- d) Taxa de *fallback*: Nenhum robô é perfeito, portanto, espera-se que os *chatbots* apresentem falhas. A taxa de *fallback* captura a quantidade de vezes em que o *chatbot* falhou ou passou por uma situação de quase falha. Uma alta taxa de *fallback* alerta para tentar encontrar novas fontes de dados ou conjuntos de treinamento para melhorar o desempenho.

- e) Satisfação do usuário: Os usuários ou as pessoas que se envolvem com o *chatbot* podem avaliar sua experiência para obter mais excelência no produto. Isso pode ser implementado como uma variável binária como “o *bot* teve um bom desempenho? - Sim ou Não”, ou podem ser criados formulários de avaliação mais complexos para classificar e fornecer pontos para cada categoria diferente. Essa métrica capturaria a eficácia geral do *chatbot* a partir do ponto de experiência do usuário fornecido diretamente pelo agente conversacional.

Essas são algumas das principais métricas de desempenho utilizadas regularmente, que foram apontadas no site *AI Multiple*. Salienta-se que possível alterar a métrica e criar métricas. Sendo assim, a filtragem baseada em conteúdo é uma maneira de obter os detalhes desejados da métrica.

Uma grande empresa focada no lado analítico dos *chatbots* é a *Botanalytics*. Ela está sediada em São Francisco e é a empresa líder em ferramentas de análise de conversação. Estabeleceu-se em 2016, e seu principal objetivo é melhorar as interações humano-*chatbot* e a interface do usuário conversacional por meio de dados. Ao oferecer análises para *chatbots*, capacita os desenvolvedores de agentes conversacionais, as marcas e as agências corporativas que visam melhorar seus *chatbots* com dados históricos e em tempo real.

O site *AI Multiple* aponta, ao menos, oito métricas para avaliação de agentes conversacionais. Tais métricas foram avaliadas e comparadas com as métricas elencadas nos trabalhos referidos, bem como com as métricas da NBR ISO 9126, e foram utilizadas para o desenvolvimento do *framework* proposto nesta tese.

2.3.7 Estudo 7 – *Chatbot Analytics*: as métricas essenciais que é preciso acompanhar

Para Phillips (2018), é importante ter informações sobre o andamento dos *chatbots*, para que se tenham dados sobre oportunidades de crescimento e de aprimoramento. Os indicadores-chave de desempenho podem variar, com base no caso de uso do *chatbot* e na demografia da base de usuários. No entanto, várias métricas importantes forneceram informações valiosas para o desenvolvimento e avaliação do *chatbot* para as mais diversas finalidades. São elas:

- Usuários Ativos: O monitoramento de usuários ativos é essencial para a maioria dos aplicativos de *software*, e com os *chatbots* não poderia ser diferente. Trata-se de métrica simples e poderosa para incluir em qualquer análise.

- Duração da sessão / etapas por conversa: Capacidade de manter uma conversa significativa com os usuários. Embora a duração ideal da sessão varie de acordo com seus casos de uso e no contexto da conversa, as sessões curtas são geralmente indicativas de alguma forma de falha, a menos que o *chatbot* possa resolver as perguntas dos usuários quase imediatamente. Idealmente, a maioria dos *chatbots* deve ter o objetivo de resolver a consulta de um usuário no menor número possível de etapas¹⁵.
- Classificações do *chatbot* do usuário: Permitir que os usuários classifiquem o *chatbot* é um método excepcional de fornecer a eles a oportunidade de expressar satisfação ou insatisfação com seus *chatbots*.
- Respostas de *fallback*: Os *chatbots* invocam respostas de *fallback* quando não conseguem encontrar uma resposta adequada à mensagem do usuário. Em vez de não dizer nada, é melhor que um *chatbot* responda informando ao usuário que não foi encontrada uma correspondência.

Neste trabalho, Phillips (2018) apresenta importantes métricas para avaliação e desenvolvimento de *chatbots* com suas respectivas descrições. Assim, as métricas apontadas pelo autor podem ser utilizadas para o desenvolvimento do *framework* desta tese.

2.3.8 Estudo 8 - Medindo a eficácia do *Chatbot*

Segundo o site *Corporate Living Actor*¹⁶, utilizar *Key Performance Indicator* (Indicador-Chave de Desempenho) pode ajudar a medir alguns comportamentos em relação ao agente conversacional, como, por exemplo, o que os usuários pensam sobre os seus *chatbots*, se estão satisfeitos com as respostas que eles fornecem, se estão aproveitando todas as potencialidades da ferramenta, se o agente conversacional tem um impacto positivo quando é contatado, entre outros indicadores que poderão ser definidos.

Assim, segundo informações do *Corporate Living Actor*, se não é possível avaliar os *chatbots*, provavelmente, é porque não existem os principais indicadores ou métricas de desempenho necessários. Essas medidas são indispensáveis para rastrear os resultados do *chatbot*, identificar possíveis obstáculos e melhorá-lo continuamente.

O *Corporate Living Actor* sugere, então, as seguintes métricas para a avaliação de *chatbots*:

¹⁵ De acordo com Phillips (2018), uma etapa de conversa é uma troca de mensagens entre um usuário e um *chatbot*.

¹⁶ Disponível em: <https://corporate.livingactor.com/en>

- Taxa de conversão: medida de um agente de bate-papo que transforma conversas on-line em novos negócios. Essa métrica tem uma perspectiva comercial, no entanto, no contexto educacional, poderá ser adaptada no sentido de o agente conversacional atender a dúvida do aluno.
- Taxa de autoatendimento: Métrica para verificar até que ponto um *chatbot* é capaz de resolver conversas sem criar uma situação que precisa ser resolvida por uma central de atendimento, por meio do envolvimento de humanos. Mas, pode-se adaptá-la para o contexto educacional, buscando entender até que ponto um *chatbot* consegue esclarecer a dúvida de um aluno sem que ele precise questionar o professor.
- Taxa de satisfação: Métrica para medir o grau de satisfação do usuário.
- Gatilhos de confusão: alguns *chatbots* ficam confusos ao tentar entender as perguntas dos usuários. Essa métrica deve medir o momento de quando a conversa entrou em um momento “estranho”.
- Taxas de Inteligência Artificial (IA) e Aprendizado de Máquina (AM): Métrica para verificar a capacidade de medir como os *chatbots* aprendem a ser mais eficientes.

Neste sentido, as métricas apontadas pelo *Corporate Living Actor* (Taxa de autoatendimento, Taxa de satisfação, Gatilhos de Confusão e Taxas de Inteligência Artificial e Aprendizado de Máquina) podem ser consideradas junto com as demais métricas dos demais trabalhos analisados e da ISO IEC 9126, para elaborar o *framework* proposto.

2.3.9 Estudo 9 - *ChatEval*: Uma ferramenta para avaliação do *chatbot*

A dificuldade em avaliar *chatbots* também foi percebida por Sedoc et al. (2019). Os autores reforçam que as práticas atuais de avaliação dos agentes acontecem a partir de avaliações por humanos. A saber, a falta de padronização nos procedimentos de avaliação e o fato de os parâmetros e o código do modelo raramente serem publicados dificultam experimentos sistemáticos para uma avaliação humana.

Os autores introduziram uma estrutura unificada para avaliação humana de *chatbots*, o *ChatEval*, que aprimora as ferramentas existentes e fornece relatórios para os pesquisadores compartilharem e compararem seus sistemas de diálogo. O *ChatEval* consiste em dois componentes principais: 1) uma base de código aberto para conduzir avaliação técnica e humana de *chatbots* de maneira padronizada; e 2) um portal da *web* para acessar código do modelo, parâmetros treinados e avaliação. As métricas de avaliação utilizadas no *ChatEval* são: diversidade léxica (conjunto de palavras pertencentes a determinada língua), semelhança média

(média da incorporação de palavras de uma resposta gerada e resposta verdadeira), tempo médio para formar uma resposta, complexidade da resposta, medida usando a probabilidade de o modelo prever a resposta correta e avaliação de expressões. O *ChatEval* é uma estrutura para avaliação sistemática de *chatbots* e fornece um repositório de código e parâmetros do modelo, conjuntos de comparação e comparações de modelos.

As métricas de avaliação utilizadas pelo *ChatEval* também serão analisadas com as demais métricas tanto dos demais trabalhos relacionados, como da ISO IEC 9126, a fim de elaborar o *framework* com métricas de avaliação e construção de agentes conversacionais.

2.3.10 Estudo 10 - Análise de expressões lexicais

Os parâmetros e métricas das expressões de emoção e das características corporais foram investigados por Hayashi (2016), que analisou como essas variáveis incorporadas nos agentes conversacionais educacionais podem impactar o desempenho dos estudantes.

No estudo, foi realizada também uma análise de rede lexical, com foco na ocorrência de palavras-chave na explicação do texto do participante como variáveis dependentes. Os resultados envolvendo a mineração de textos e o aprendizado de máquina mostram que, durante a atividade de explicação, as expressões de emoção, as características corporais e o gênero dos agentes conversacionais educacionais influenciam no desempenho dos alunos.

O estudo abrange informações de métricas sobre o comportamento humano e analisa tarefas executadas on-line, além de oferecer sugestões relacionadas ao *design* de sistemas eficientes de tutoria virtual. A referida pesquisa também foi desenvolvida com foco educacional, considerando, especificamente, parâmetros e métricas das expressões de emoção e características envolvendo textos.

2.3.11 Estudo 11- Métricas para o desenvolvimento de agentes conversacionais

Outro estudo sobre as métricas de avaliação envolvendo agentes conversacionais no contexto da engenharia de sistemas foi desenvolvido por O'Shea, Bandar e Crockett (2011). Para os autores, a tecnologia de agentes conversacionais tem dois pontos de interesse para os engenheiros de sistemas: o uso de técnicas de engenharia de sistemas na pesquisa de agentes conversacionais e a aplicação dos *chatbots* em desenvolvimento de projetos.

Os agentes conversacionais oferecem a oportunidade de automatizar aplicativos mais complexos do que é possível com as interfaces da *web* convencionais. Atualmente, esses

aplicativos exigem que um especialista humano no domínio faça a mediação entre o usuário e o aplicativo. Os autores dividem as métricas usadas para avaliar os agentes conversacionais em três categorias: medidas subjetivas, tentativas de medidas subjetivas e medidas objetivas, conforme descrito a seguir.

Medidas subjetivas:

- Usabilidade;
- Satisfação do usuário;
- Credibilidade do agente.

Tentativas de medidas subjetivas:

- Facilidade de uso / facilidade de tarefa;
- Facilidade de o usuário entender o agente;
- O entendimento do agente sobre a compreensão do usuário;
- Vários atributos cognitivos relacionados à compreensão e complexidade;
- Vários atributos relacionados à confiabilidade do agente;
- Vários atributos relacionados à experiência do usuário (do domínio ou uso do agente);
- A eficiência ou eficácia do agente;
- Vários atributos sobre comando e controle da conversa;
- O ritmo da interação;
- Se o agente se comportou conforme o esperado;
- Quão natural parecia o comportamento do agente;
- Vários atributos emocionais positivos (por exemplo, simpatia, prazer);
- Vários atributos emocionais negativos (por exemplo, tédio, confusão);
- Se o usuário usaria novamente.

Medidas objetivas

- Duração do diálogo / conversa;
- Contagem de voltas ao diálogo;
- Medidas de sucesso no nível da expressão ou conclusão da tarefa;
- Contagens de erros, correções ou taxas de erro percentual.

Entre todos os estudos envolvendo métricas de agentes conversacionais, o de O'Shea, Bandar e Crockett (2011) foi o que apontou o maior número de métricas para o desenvolvimento e avaliação de agentes conversacionais, e tais métricas poderão ser muito úteis para a avaliação e para a elaboração da proposta de *framework* apresentada na presente tese.

2.3.12 Estudo 12 - Conversar com agente humano

Interessados no problema de entender a estrutura permanente de conversação humana no contexto de agente humano e interação humano-humano, Griol e Molina (2015) apresentam uma metodologia estatística para detectar a estrutura dos diálogos falados, com base em um modelo usando árvores de decisão.

A avaliação do estudo contou com um *corpus* de diálogo coletado de usuários reais envolvidos em uma tarefa de solução de problemas. Os resultados da avaliação mostram que a segmentação automática de diálogo é muito eficaz nos modelos construídos, usando separadamente diálogos homem-agente ou diálogos homem-humano. Também é possível inferir a estrutura relacionada à tarefa dos diálogos humano-humano com um modelo aprendido utilizando apenas caixas de diálogo agente-humano.

Este estudo trouxe testes relacionados a testes de conversação e métricas de mensagens. Foi apresentada a importância de ampliar e subsidiar novas métricas para a avaliação de agentes conversacionais no contexto educativo.

2.3.13 Outros trabalhos de avaliação de *chatbot*

Os 12 estudos apontados nas subseções anteriores (da 2.3.1 à 2.3.12) foram coletados até o primeiro semestre de 2020 (período da qualificação da tese). Ressalta-se que as métricas referidas nessas pesquisas foram utilizadas para compor o *framework* para a avaliação de agentes conversacionais educacionais. Além destes trabalhos que aprofundaram estratégias de avaliação de *chatbot*, recentemente, outros estudos foram publicados e serão apresentados a seguir.

Desde dezembro de 2019, a Covid-19 vem causando uma enorme crise de saúde em todo o mundo, causando medo, dúvidas e incertezas. Especialistas clínicos informam que os pacientes com Covid-19 devem ser diagnosticados em estágio inicial para salvar suas vidas. Sabendo disso, Joseph Raj, Anton e Johnson Raj (2021) desenvolveram um *chatbot* para atender a sociedade que sofre com a Covid-19. O *chatbot* utilizou os algoritmos *Support Vector*

Machines (SVM), *Artificial Neural Network* (ANN), *Decision Tree*, *K- Nearest Neighbors* (KNN), *Random Forest* e *Logistic Regression*, os quais são utilizados para avaliar o desempenho da classificação. O *chatbot* é projetado usando a melhor técnica de aprendizado de máquina, a *Árvore de Decisão*, sendo considerada mais eficiente que as demais técnicas na resolução de problemas de classificação relacionados à Covid-19.

Desse modo, os administradores e médicos especialistas que lutam contra a Covid-19 podem usar o *chatbot* para avaliar a condição dos pacientes, a fim de tratá-los adequadamente, salvar suas vidas e evitar a classificação incorreta do tratamento. Esta pesquisa apresentou um estudo focado em vários desenvolvimentos recentes importantes em técnicas de aprendizado de máquina para problemas de classificação relacionados à Covid-19. Para os autores, a avaliação do *chatbot* contemplaria as métricas de Testes de Conversação e Qualidade das respostas, métricas já contempladas nesta tese.

Seguindo na linha de *chatbots* para auxiliar no combate à Covid-19, Almalki e Azeez (2020) realizaram um estudo de revisão de literatura sobre *chatbots* relacionados à referida doença. Os autores pesquisaram nas plataformas *PubMed/MED LINE*¹⁷ e *Google Scholar*, durante o período de janeiro a setembro de 2020, usando as palavras-chave Covid, *chatbot*, assistente virtual e “*AI enabled platform Covid*”. Dos 543 artigos inicialmente identificados, 9 foram elegíveis para inclusão no estudo. O total de seis pesquisas descrevem o desenvolvimento e a arquitetura de *chatbots* e apenas três tratam da experiência do usuário.

Registra-se que a revisão de escopo que compõe esta tese identificou cinco aplicações principais dos *chatbots* atuais de saúde, a saber: disseminação de informações e conhecimentos de saúde; autotriagem e avaliação de risco pessoal; monitoramento da exposição e notificações; rastreamento de sintomas e aspectos de saúde relativos à Covid-19; e combate à desinformação e notícias falsas.

A partir dos resultados de sua pesquisa, Almalki e Azeez (2020) apontam que o uso de *chatbots* de saúde apresentam muitos desafios, como a usabilidade e a aceitabilidade dos consumidores quanto ao sistema técnico. Trata-se de um tema altamente relevante, mas que ainda aponta fragilidades, principalmente no que concerne às técnicas e métricas para a avaliação dos *chatbots*. Cabe salientar, por fim, que as métricas Usabilidade e Confiabilidade, abordadas por esses autores, são utilizadas no *framework* proposto na presente tese.

Mokmin e Ibrahim (2021) também realizaram uma avaliação do *chatbot* como ferramenta de alfabetização em saúde entre estudantes de graduação. Nesta avaliação, os

¹⁷ Mais informações em: <https://pubmed.ncbi.nlm.nih.gov/>. Acesso em: março de 2022.

autores examinaram a utilidade, o desempenho e a aceitação da tecnologia de um *chatbot* desenvolvido para este fim. Um total de 75 estudantes de graduação estiveram envolvidos em uma entrevista semiestruturada durante um período total de dois meses. Cada entrevistado explorou o *chatbot* de saúde ativamente, por meio de dispositivos móveis. Os resultados da avaliação mostraram que 73,3% dos entrevistados acharam que o *chatbot* pode ajudar a entender diversos problemas de saúde e proporcionar uma boa conversa. A avaliação de desempenho também demonstrou que o *chatbot* contribuiu com um baixo percentual de saída, visto que menos de 37% dos usuários saíram do aplicativo. Em resumo, a avaliação geral apontou que o *chatbot* desenvolvido tem um potencial significativo para ser usado como agente conversacional para aumentar a alfabetização em saúde, principalmente entre estudantes e jovens adultos.

As métricas apontadas no estudo de Mokmin e Ibrahim (2021) também são contempladas no *framework* proposto nesta tese. Afinal, quando os autores examinam o quanto o *chatbot* estava proporcionando uma boa conversa, a métrica equivalente ao *framework* da tese é a Satisfação do *chatbot*; quando avaliam o desempenho, utilizam a métrica Performance; e quando avaliam o percentual de saída de usuários do aplicativo, referem-se à métrica Usuários Ativos. Assim, o estudo do *chatbot* realizado pelos autores segue um plano de elaborar um *chatbot* e de fazer sua análise. Constata-se que, das 19 métricas propostas no *framework* da presente tese, os referidos autores utilizaram apenas três para chegarem à conclusão do estudo.

Em uma pesquisa desenvolvida na Arábia Saudita, Almalki (2020) explorou as percepções de 166 usuários de *chatbots* na saúde. O autor argumenta que alguns usuários estavam mais dispostos a usar *chatbots* de saúde para buscar informações gerais sobre a Covid-19 (82,5%), ou seja, preferiam isso a pesquisar informações sobre tratamentos médicos referentes à doença (72,3%). Entre os resultados, Almalki (2020) conclui que os participantes que forneceram qualquer tipo de dados de serviços de saúde tiveram maior probabilidade de autoavaliar os sintomas da Covid-19 usando *chatbots*. Constata, também, que a conscientização dos participantes sobre esta ferramenta e o uso de *chatbots* de saúde foram baixos. No entanto, a maioria teve percepções positivas sobre essas tecnologias emergentes e demonstra vontade de utilizá-las.

O estudo de Almalki (2020) conclui, por fim, que novas pesquisas são necessárias para capturar a usabilidade do mundo real dessas novas tecnologias, empregando projetos metodológicos mais rígidos. Nesse sentido, destaca-se que novamente há uma possibilidade de uso do *framework* com métricas de avaliação de *chatbots* que pode contribuir com os projetos

metodológicos aos quais o autor se refere, afinal, diversas métricas podem - e muito - ajudar e colaborar para a avaliação e, até mesmo, a orientação de agentes conversacionais educacionais.

Neste âmbito, ao se lançar a pergunta: Conversar com um *chatbot* sofisticado é tão bom quanto conversar on-line com um estranho? Para respondê-la, Drouin et al. (2022) realizaram um experimento para estudar o processo de familiarização, em que foram atribuídos, aleatoriamente, 417 participantes em três condições: bate-papo face a face com um humano, bate-papo on-line com um humano e bate-papo on-line com um *chatbot* emocionalmente responsivo. Após um bate-papo de 20 minutos, os participantes relataram seu estado afetivo e avaliações relacionais do bate-papo. Além disso, todos os bate-papos foram gravados e os textos analisados por meio do programa *Linguistic Inquiry and Word Count*.

Em todas as condições, eles relataram níveis moderados de emoções positivas e baixos níveis de emoções negativas. No entanto, aqueles que conversaram com um humano também relataram gostar de seu parceiro de bate-papo e que seu parceiro foi mais responsivo. Enquanto isso, os participantes tiveram menos preocupações de conversação com o *chatbot*. Essas descobertas têm implicações para futuros estudos de interação mediada por computador, tendo em vista que as conversas com *chatbots* parecem ter diferentes efeitos no prazer de conversar e preocupações conversacionais em contextos de familiarização. De fato, esses resultados podem ajudar a melhorar a recepção e a comercialização de *chatbots*. Assim, o estudo realizado por Drouin et al. (2022) contempla as métricas já identificadas nos estudos anteriores: Confiabilidade, Ética e Comportamento, Humanidade e Afetividade.

2.4 COMPARAÇÃO DAS MÉTRICAS DOS TRABALHOS SELECIONADOS

A tabela 2.2 apresenta o mapeamento das métricas selecionadas e detalhadas na subseção 2.3 (Trabalhos relacionados). Salienta-se que a coluna “Estudos relacionados” se refere às pesquisas elencadas nas subseções 2.3.1 a 2.3.12.

Tabela 2.2 - Métricas dos trabalhos relacionados

Estudos relacionados	Ano de publicação	Usabilidade	Funcionalidade	Confiabilidade	Eficiência	Performance	Humanidade	Afetividade	Ética e Comportamento	Acessibilidade	Testes de conversação	Qualidade de respostas	Total de usuários	Usuários ativos	Usuários engajados	Métricas de mensagens	Retenção dos alunos	Manter a conversa	Satisfação do <i>Chatbot</i>	Respostas de <i>fallback</i>	Total de métricas
1	2014	X	X	X																	3
2	2017		X		X	X	X	X	X	X											7
3	2013				X						X										2
4	2015										X										1
5	2016										X	X									2
6	2019					X							X	X	X	X	X	X		X	8
7	2018													X		X		X	X	X	5
8	2019				X							X						X	X	X	5
9	2019										X	X				X					3
10	2016							X				X				X			X		4
11	2011	X	X	X	X	X	X	X				X				X		X	X	X	12
12	2015										X					X					2
Totais		2	3	2	4	3	2	3	1	1	5	5	1	2	1	6	1	4	4	4	

Fonte: Elaborado pelo autor (2021).

Na seção 2.1 desta tese, foi discutida a importância dos agentes conversacionais e de suas contribuições para o contexto educativo; no entanto, há uma lacuna a ser explorada, relacionada à necessidade de propor e identificar métricas para o desenvolvimento de agentes conversacionais neste.

Conforme Tabela 2.2, constata-se que os *chatbots* educacionais são desenvolvidos para contribuir em um determinado contexto educacional, em que geralmente é abordado um conhecimento específico (ensino de física, metodologias ativas, mídias na educação) e o estudo acaba avaliando e trabalhando contextos diversos, como, por exemplo, métricas de testes de conversação, qualidade da resposta e métricas envolvendo mensagens, as quais são as métricas que aparecem na maioria dos estudos relacionados.

Todas as métricas selecionadas e apresentadas na Tabela 2.2 têm origem em estudos de agentes conversacionais educacionais, os quais, de alguma forma, envolvem métricas de avaliação. De fato, muitos estudos não identificam uma determinada métrica, mas discutem, testam e apresentam resultados significativos, como é o caso do estudo relacionado 4, baseado especificamente em testes de conversação. Além disso, outros cinco estudos trataram de testes de conversação a partir de contextos e testes diferentes.

Nesse sentido, justifica-se a importância da escolha das métricas que envolvem a tabela 2.2, visto que todas elas têm origem de agentes conversacionais, especialmente para o uso educacional. Em muitas dessas pesquisas, não está sendo diretamente abordada uma métrica específica, como, a métrica de mensagens, por exemplo. Assim, o autor desta tese que identificou esta métrica realizou o cadastro na tabela 2.2.

Os estudos relacionados 11, 6 e 2 discutem e apresentam o maior número de métricas para o desenvolvimento de agentes conversacionais. No entanto as métricas que apontam aspectos, como Usabilidade e Funcionalidade, não aparecem em todas as pesquisas analisadas até o momento. Mas são encontradas em alguns deles, como, por exemplo, no estudo 2, o qual apresenta os requisitos e funcionalidades do agente (funcionalidade, eficiência, performance, acessibilidade). O estudo 6 aborda o item funcionalidade e aparecem com muita ênfase as métricas de usuários, usuários ativos, de mensagens e manter a conversa. Por outro lado, o estudo 3 apresenta ênfase na eficiência nos testes de conversação.

Ainda há algumas pesquisas envolvendo o desenvolvimento de *chatbots* que simplesmente aplicam questionários elaborados pelos autores, verificando-se que esses questionários não parecem ter uma referência de avaliação. A partir do exposto, constata-se que entregar sugestões de métricas de *software*, a fim de sugerir o que um *chatbot* educacional pode e deve ter para ser interessante e poder ser utilizado no contexto educacional é uma

possibilidade relevante para o desenvolvimento e avaliação dos futuros agentes conversacionais.

A tabela 2.2 trouxe os 12 estudos com o mapeamento das métricas selecionadas e detalhadas nas seções 2.3.1 até a seção 2.3.12. Esses estudos foram utilizados para compor parte das métricas de avaliação de agentes conversacionais educacionais, no entanto após a conclusão e validação das métricas durante o estudo da tese, por meio de pesquisas mais recentes apontadas na seção 2.3.13, identificou-se as seguintes métricas de avaliação registradas na tabela 2.3. Esta apresenta dados mais recentes referidos nas seções “2.1.4 Tendências em *chatbots*” e “2.3.13 Outros trabalhos de avaliação de *chatbot*”.

Tabela 2.3 - Métricas de outros trabalhos de avaliação de *chatbot*

Estudos relacionados	Ano de publicação	Usabilidade	Funcionalidade	Confiabilidade	Eficiência	Performance	Humanidade	Afetividade	Ética e Comportamento	Acessibilidade	Testes de conversação	Qualidade de respostas	Total de usuários	Usuários ativos	Usuários engajados	Métricas de mensagens	Retenção dos alunos	Manter a conversa	Satisfação do Chatbot	Respostas de <i>fallback</i>	Total de métricas	
Raj, Joseph; Anton; Raj, Johnson	2021										X	X										2
Almalki e Azeez	2020	X		X																		2
Mokmin e Ibrahim	2021					X								X					X			3
Almalki	2020	X																				1
Drouin et al.	2022			X			X	X	X													4
Worny et al.	2020		X		X										X							3
Folstad et al.	2021	X	X						X						X							4
Acheampong	2020						X	X														2
Casas et al.	2021						X	X											X			3
Totais		3	2	2	1	1	3	3	2		1	1		1	2				2			

Fonte: Elaborado pelo autor (2021).

Verificou-se que nas pesquisas realizadas no período de 2020 a 2022 não foram constatadas novas métricas além das já identificadas para o desenvolvimento do *framework* proposto nesta tese. Assim, embora tenham estudos recentes sobre a avaliação de agentes conversacionais, a presente pesquisa seguirá apontando as métricas dos estudos relacionados de avaliação e as métricas apresentadas na tabela 2.2, já que foram utilizadas para compor o referido *framework*.

3 PROCEDIMENTOS METODOLÓGICOS

Este capítulo detalha os métodos adotados para o desenvolvimento deste estudo, que consistem em uma pesquisa exploratória e descritiva, caracterizada por procedimentos bibliográficos, documentais e estudo de caso, com a finalidade de criar um *framework* com sugestões de métricas de avaliação de *software*. A ideia é elencar o que um agente conversacional educacional pode e deve ter para ser relevante, com vistas a ser utilizado como referência para a avaliação e construção de *chatbots* desenvolvidos para o contexto educacional.

Neste íterim, pesquisas exploratórias têm como “finalidade principal desenvolver, esclarecer e modificar conceitos e ideias, tendo em vista a formulação de problemas mais precisos” (GIL, 2008, p. 27). Para Gil (2008) e Prodanov (2013), esse tipo de pesquisa envolve levantamento bibliográfico, documental, bem como entrevistas não padronizadas. Tais mecanismos são desenvolvidos com o objetivo de proporcionar uma visão geral e aproximativa do assunto abordado e são realizados especialmente quando o tema escolhido é pouco explorado.

Já o procedimento de estudo de caso refere-se a um minucioso e profundo estudo de um ou mais objetos e, segundo Prodanov (2013), permite novas descobertas de aspectos que não foram previstos inicialmente, restringindo-se, assim, a um objeto, como uma empresa, um produto, uma família. Nesta tese, o estudo de caso foi utilizado para explorar situações reais com parte do público-alvo da pesquisa, que são os estudantes do curso Técnico em Informática para Internet, especificamente para entender se, em relação ao uso do computador e da internet, os estudantes têm um espaço para estudar, se têm conhecimentos em relação à lógica de programação, se já conhecem ou já desenvolveram algum site na linguagem HTML, CSS ou outra linguagem.

Parte-se da hipótese de que os alunos têm computadores ou *notebooks* com uma boa configuração de *hardware* e *software*, uma velocidade de Internet que atenda a videochamadas e que tenham acesso à plataforma *Teams* utilizada nas aulas; no entanto, acredita-se que eles não têm conhecimento de linguagem de programação, nunca desenvolveram um site e não conhecem agentes conversacionais, tampouco, os agentes ATENA, METIS e ALTEIA.

Já em relação à pesquisa bibliográfica, foi “desenvolvida a partir de material já elaborado, constituído principalmente de livros e artigos científicos” (GIL, 2008, p. 50). Quanto à análise documental, de acordo com Gil (2008), diferencia-se da pesquisa bibliográfica pela natureza das fontes. As pesquisas documentais se baseiam em materiais que não receberam um tratamento analítico ou que podem ser reelaborados de acordo com os objetivos da pesquisa.

No caso do presente estudo, os agentes conversacionais educacionais METIS, ATENA e ALTEIA são considerados como documentos, pois existem de forma independente em relação à pesquisa.

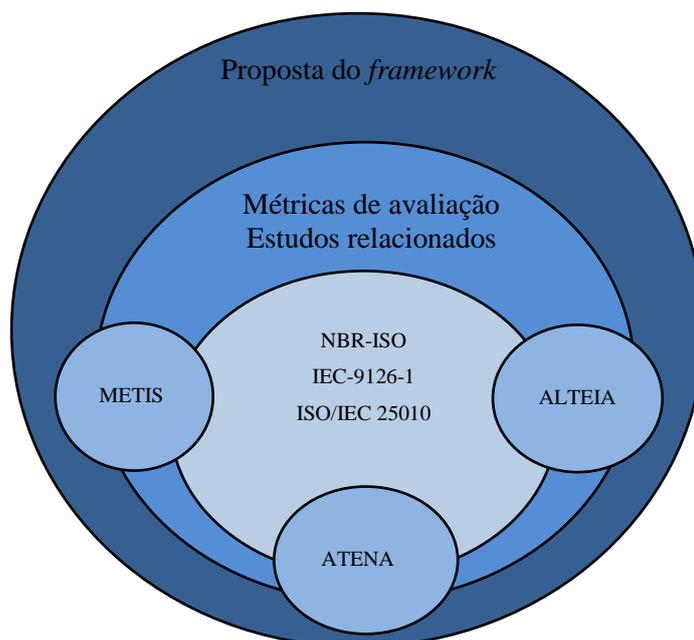
Quanto à pesquisa descritiva, trata-se de “quando o pesquisador apenas registra e descreve os fatos observados sem interferir neles. Visa descrever as características de determinada população ou fenômeno” (PRODANOV, 2013, p. 52). Para Gil (2008), são inúmeros os estudos que podem ser classificados sob este título e salientam-se aqueles que têm por objetivo estudar as características de um grupo, tais como: distribuição por idade, sexo, nível de renda, entre outros.

No caso da presente tese, a pesquisa descritiva foi utilizada para analisar as métricas do Modelo NBR-ISO IEC-25010-2011 e as métricas utilizadas nos trabalhos relacionados já apresentados na tabela 2.2, de forma a elaborar as novas métricas, a fim de identificar o que um agente conversacional pode e deve ter para ser relevante.

A escolha dos agentes conversacionais METIS, ATENA e ALTEIA, conforme apresentado na introdução desta pesquisa, resultou do fato de estes agentes terem sido desenvolvidos especificamente para o contexto educacional e no âmbito do Projeto AVATAR. Considerando que a proposta do presente estudo é desenvolver um *framework* com métricas de avaliação de *software*, a fim de elencar o que um agente conversacional educacional deve ter para ser utilizado como referência para a avaliação de *chatbots* de contexto educacional, após o uso, exploração e acesso às ferramentas, foi possível testar o *framework* proposto, bem como propor futuras melhorias dos respectivos agentes, com base na experiência dos alunos.

A figura 3.1 ilustra a relação dos temas-chave da respectiva tese.

Figura 3.1 – Relação dos temas-chave do estudo



Fonte: Elaborado pelo autor (2021).

Percebe-se que, conforme a figura 3.1, a proposta do *framework* foi aplicada na avaliação dos agentes conversacionais METIS, ATENA e ALTEIA. O *framework* será composto por métricas a partir das normas NBR-ISO IEC-9126-1, NBR ISO/IEC 25010 e das métricas de avaliação dos trabalhos relacionados já apresentados na tabela 2.2. Dessa forma, parte-se da hipótese que disponibilizar um *framework* com métricas de avaliação pode contribuir para o aprimoramento, avaliação e construção de agentes conversacionais educacionais.

3.1 CENÁRIO DO ESTUDO DE CASO

O cenário do presente estudo é o Sistema Fecomércio - Senac, na cidade de São Leopoldo-RS. Participaram da pesquisa 77 alunos, que estão no primeiro ano do Ensino Médio e cursam, paralelamente, o Curso Técnico em Informática para Internet. A idade dos alunos varia entre 15 e 17 anos.

O contexto de atuação das aulas ocorreu entre os meses de abril e junho de 2021, período no qual os alunos estavam assistindo às aulas de maneira híbrida (50% on-line e 50% presencial). Há alguns casos de estudantes que optaram por assistir às aulas somente na modalidade on-line, considerando que a legislação vigente contemplava essa possibilidade no período em que ocorreram as aulas.

As aulas sucederam de segunda a sexta-feira, das 7h30 às 12h30, e as aulas do Curso Técnico em Informática para Internet aconteceram de forma paralela aos demais componentes curriculares. Cada estudante tinha um *notebook* para seu uso exclusivo durante os encontros, uma conta no *Microsoft Teams* e acesso liberado à internet. Cada turma teve cinco horas semanais no curso de Técnico em Informática. Este curso, junto às novas áreas, é organizado em três trimestres anuais e, em cada trimestre, há previsão de trabalhar os conteúdos relatados na tabela a seguir.

Tabela 3.1 - Organização trimestres e conteúdos

Semestre	Data Início	Data Fim	Conteúdos	Carga horária
1º trim	22/02/2021	28/05/2021	* Fundamentação da Computação * Algoritmos * HTML * AIML	70 horas
2º trim	31/05/2021	06/09/2021	* HTML * CSS	70 horas
3º trim	08/09/2021	17/12/2021	* Sistemas operacionais * Redes de Computadores * Técnicas de Programação	70 horas

Fonte: Elaborado pelo autor (2021).

De acordo com a tabela 3.1, constata-se que os conteúdos trabalhados pelos alunos são conteúdos introdutórios de um curso Técnico em Informática para Internet, considerando que estão no primeiro ano do ensino médio. Assim, justificam-se conhecimentos sobre Fundamentos da Computação, algoritmos, linguagem HTML e uma introdução a AIML no primeiro trimestre. A partir do escopo desses conteúdos trabalhados no primeiro trimestre, os estudantes passam a ter condições de construir um site com as TAGS fundamentais para o desenvolvimento de um site na linguagem HTML e programar em AIML.

Registra-se, por fim, que esta pesquisa ocorreu durante o primeiro e segundo trimestre de estudo dos alunos, ou seja, eles tiveram algumas aulas de Fundamentos da Computação e o conhecimento de algumas *tags* mais básicas de HTML. O detalhamento dos dias, horários, carga horária e todas as atividades e fases da pesquisa será apresentado na próxima subseção.

3.2 PERFIL DEMOGRÁFICO

As aulas do Curso Técnico em Informática para Internet ocorreram junto ao Senac São Leopoldo, localizado no parque tecnológico da Tecnosinos. O Senac é uma instituição privada e não há estudantes com bolsas. Devido ao perfil dos alunos ser de iniciantes na escola e o

contexto da pesquisa ocorrer em meio a uma pandemia, na qual, no ano anterior, os alunos tiveram suas aulas no formato totalmente on-line; e, ainda, ao fato de a escola ser totalmente nova para todos os estudantes, entendeu-se que seria importante mapear o perfil e compreender como eles têm acesso à internet, como são seus computadores e o nível de informação que possuem.

Participaram desta pesquisa setenta e sete estudantes, sendo este o perfil demográfico dos participantes: 40 meninos e 37 meninas (todos os estudantes na faixa etária entre 15 até 17 anos). Para o desenvolvimento das aulas, procurou-se compreender quais os dispositivos de *hardwares*, *softwares* e meios de acesso à internet a que os estudantes tinham acesso. Para isso, realizou-se uma pesquisa inicial antes de quaisquer atividades.

3.3 FASES DA PESQUISA

Para uma melhor organização metodológica, a pesquisa divide-se em três fases, que, neste estudo, foram designadas como Fase 1, Fase 2 e Fase 3.

3.3.1 Fase 1 – Perfil dos alunos do Curso Técnico em Informática

Na fase 1, ocorreu o levantamento do perfil do público-alvo. Para isso, foi utilizado um questionário, com o propósito de identificar questões relacionadas ao perfil do aluno, tais como as experiências quanto ao uso dos agentes conversacionais, os meios de acesso à internet e como se dá o uso de cada ferramenta. Segundo Dias (2007), questionários e entrevistas permitem ao avaliador conhecer experiências, opiniões e preferências dos usuários ao utilizarem um determinado sistema.

Segundo Prodanov (2009), a entrevista é uma técnica de levantamento de dados primária e dá grande importância à descrição verbal dos entrevistados. O objetivo principal da entrevista é coletar informações de entrevistados sobre determinado assunto ou problema. Sendo assim, o questionário de entrevistas utilizado neste estudo teve o objetivo de conhecer o perfil do público em questão e entender qual é a aproximação dos estudantes com o objeto da pesquisa, questionando-os sobre suas experiências com o uso do computador, a forma e o tempo de acesso à internet, se fazem uso para lazer, estudo ou ambos e, também, o que cada usuário conhece sobre as ferramentas a serem exploradas neste estudo. Tais informações são fundamentais para a organização geral nos encontros do curso de informática.

O questionário foi aplicado antes de iniciar qualquer atividade ou prática com os alunos e teve como objetivo conhecer como se dá o acesso às tecnologias pelo público-alvo, analisando as ferramentas, tecnologias e conhecimentos técnicos em relação à programação ou linguagem de programação a que os participantes da pesquisa teriam acesso.

As questões aplicadas aos alunos podem ser conferidas no Apêndice B. Registra-se, por fim, que, para elaborar e aplicar o questionário, utilizou-se o *Google* Formulários.

3.3.2 Fase 2 – Conhecendo e avaliando os Agentes Conversacionais

A Fase 2 ocorreu em dois momentos, por meio das práticas de programação e da experiência dos alunos com os agentes conversacionais.

Fase 2.1 – Práticas de Programação: O primeiro momento da Fase 2 (Fase 2.1) foi mais longo, uma vez que ocorreu durante os encontros nas aulas do Curso Técnico em Informática para Internet, no Colégio Senac, em São Leopoldo, quando os alunos utilizaram os agentes conversacionais ATENA, METIS e ALTEIA e realizaram atividades práticas a partir das experiências identificadas na Fase 1.

Os conteúdos e propostas foram desenvolvidos ao longo da semana para ambas as turmas, e cada uma tinha cinco horas semanais de conteúdo do Curso Técnico em Informática. A turma A tinha aulas nas quintas-feiras, das 7h30 às 9h30, e nas sextas-feiras, das 9h30 às 12h30. A turma B tinha aulas nas segundas-feiras das 09h30 às 12h30 e quintas-feiras das 10h30 às 12h30. A tabela 3.2 apresenta os conteúdos, períodos semanais e carga horária desenvolvida na Fase 2.1:

Tabela 3.2 – Organização dos conteúdos desenvolvidos na Fase 2.1

Aula	Período	Conteúdo	Carga horária
1	05/04/21 à 09/04/21	Introdução ao HTML: <ul style="list-style-type: none"> ● MS <i>Visual Studio Code</i>; ● Tags (<i>html, head, title, body, charset, div, h1, h2</i> e <i>a</i>); ● O que é CSS; ● Criar o arquivo externo <i>.css</i>; ● Link entre HTML e CSS; ● Seletores; 	5h
2	12/04/21 à 16/04/21	Introdução ao HTML: <ul style="list-style-type: none"> ● <i>Box Model</i>; ● <i>Width, Height</i>; ● <i>Font, font-size, font-family</i>; 	5h

		<ul style="list-style-type: none"> ● <i>Background</i>; ● <i>Text-align</i>; ● Comentários; 	
3	19/04/21 à 23/04/21	<p>Introdução ao HTML:</p> <ul style="list-style-type: none"> ● Instalar o <i>Visual Studio Code</i>; ● Ajustar imagens no código html com as <i>tags height</i> e <i>width</i>; ● Revisar cada <i>TAG</i> no código html; ● Verificação do suporte de digitação dos códigos no <i>Visual Studio Code</i>; <p>Analisar alguns comandos e códigos no <i>VSCoDe</i>;</p>	5h
4	26/04/21 à 30/04/21	<p>Introdução a agentes conversacionais;</p> <ul style="list-style-type: none"> ● Apresentação e uso dos agentes conversacionais: ATENA, METIS e ALTEIA; ● Práticas de conversação com os agentes do estudo; ● Relacionar algumas TAGS do código para um agente conversacional na linguagem <i>Artificial Intelligence Markup Language (AIML)</i> com os códigos de construção de página <i>web</i> (HTML); ● Desenvolver como tais linguagens e conhecimentos se complementam. 	5h
5	03/05/21 à 07/05/21	<p>HTML;</p> <ul style="list-style-type: none"> ● <i>Tags</i> de vídeo; ● <i>Background</i>; ● <i>Color</i>; ● Listas ordenadas; ● Listas não ordenadas; <p>Agentes conversacionais:</p> <ul style="list-style-type: none"> ● Construir arquivos no formato <i>.aiml</i>; ● Fazer <i>upload</i> e <i>download</i> no Servidor da UFRGS; ● Inserir o código de conversação do AIML em sua página pessoal HTML; <p>Acesso ao <i>Program-O</i> do servidor UFRGS disponibilizado para a pesquisa:</p> <p><i>Links</i> utilizados para a prática: Acesso ao <i>Program-O</i>: http://avatar.cinted.ufrgs.br/higia.old/admin/</p> <p>Conversação: http://avatar.cinted.ufrgs.br/higia.old/gui/jquery/</p>	5h
6	10/05/21 à 14/05/21	<p>Semana de práticas envolvendo os Conteúdos:</p> <ul style="list-style-type: none"> ● Criação de <i>Tags</i> em AIML, ● Inserção do <i>link</i> de conversação do <i>chatbot</i> no site dos alunos. 	5h
		Hospedagem de <i>site</i> no <i>Git Hub</i> :	

7	17/05/21 à 21/05/21	<ul style="list-style-type: none"> ● Introdução ao <i>Github</i>; ● Cada aluno criou uma conta no <i>Github</i>; ● Fazer upload dos arquivos .html no <i>Github</i>; ● Testar os sites desenvolvidos na plataforma <i>Github</i>; ● Diferenciar os arquivos do <i>site</i> (.html) dos arquivos do servidor que hospedou os agentes conversacionais. 	5h
Carga horária total de práticas			35h

Fonte: Elaborado pelo autor (2021).

Conforme apresentado na tabela 3.2, pode-se verificar que, na Fase 2.1, houve a oportunidade de integrar os conteúdos das aulas do Curso Técnico em Informática para Internet ao experimento dos agentes, para que os alunos pudessem conhecê-los, criar códigos de conversação, adicionar tais códigos no agente conversacional e, ainda, integrar a conversação junto ao *site* desenvolvido pelos estudantes.

Destaca-se que a Fase 2.1 ocorreu dois meses após o início das aulas do Curso Técnico em Informática para a Internet, ou seja, os alunos já tinham conhecimento sobre os Fundamentos da Computação, e, posteriormente, foram introduzidos na linguagem de Programação HTML, CSS, AIML, em que tiveram a oportunidade de não apenas testar e interagir com o agente conversacional, mas também de criar códigos com diálogos elaborados por eles, fazer *download* e *upload* dos arquivos .aiml. Houve a disponibilidade de um servidor da UFRGS para a pesquisa, o que permitiu a possibilidade de adicionar o *link* de conversação a partir deste servidor nos sites em que os alunos estavam elaborando os repositórios.

Fase 2.2 – Experiências dos alunos com os agentes conversacionais: O segundo momento da Fase 2 (Fase 2.2) ocorreu ao final das aulas do Curso Técnico em Informática para Internet no Senac, em São Leopoldo, quando foi possível coletar informações a partir das experiências dos alunos com o uso dos agentes conversacionais.

Após o período de construção dos *sites* e de conversação da agente conversacional de teste da Fase 2.1, na semana seguinte, discentes novamente acessaram os *links* dos agentes conversacionais para diálogos e conversação. Nesta Fase, eles já tinham conhecimentos sobre conceitos de agentes conversacionais e sobre como desenvolver códigos no formato .aiml. Além disso, também já tinham integrado os *links* de conversação junto ao primeiro site desenvolvido.

O acesso aos agentes conversacionais da pesquisa, pelos alunos do Técnico, teve como objetivo avaliar os agentes. Para realizar a avaliação, cada aluno acessou um agente conversacional de cada vez e realizaram conversações. Reservou-se o tempo de uma hora para

atuarem na conversação com cada agente conversacional, , para que pudessem responder ao questionário de avaliação a partir das experiências de conversação com as agentes.

A seguir, apresenta-se data, horário e carga horária destinada para a avaliação dos agentes conversacionais do estudo METIS, ATENA, ALTEIA para cada uma das turmas.

Tabela 3.3 - Organização dos dias e horários desenvolvidos na Fase 2.2

Turma	Data	Horário	Agente	Carga horária
A	04/06/2021	09:30 às 10:30	METIS	1:00
A	04/06/2021	10:30 às 11:30	ALTEIA	1:00
A	04/06/2021	11:30 às 12:30	ATENA	1:00
B	31/05/2021	09:30 às 10:30	METIS	1:00
B	31/05/2021	10:30 às 11:30	ALTEIA	1:00
B	31/05/2021	11:30 às 12:30	ATENA	1:00

Fonte: Elaborado pelo autor (2021).

A fase 2.2 envolveu a aplicação de um questionário (Apêndice C), com perguntas baseadas nas métricas da NBR ISO/IEC 9126. Tal questionário foi um estudo piloto para modelagem do *framework* proposto para avaliação dos agentes conversacionais. Nesta fase, optou-se por não utilizar as métricas dos trabalhos relacionados, pois elas ainda não haviam passado pela análise dos especialistas.

O questionário registrou informações cujo peso de cada questão pode variar de 5 a 1, sendo peso 5 quando o aluno concorda totalmente, e peso 1 quando discorda totalmente. Caso o estudante não tenha condições de opinar, também fica registrada essa informação. O objetivo principal deste questionário é identificar as principais métricas de avaliação de agentes conversacionais e, conseqüentemente, analisar quais métricas avaliadas são mais impactantes positivamente para serem melhoradas a partir de tais indicadores nos agentes conversacionais METIS, ATENA e ALTEIA.

3.3.3 Fase 3- Participação dos Especialistas

A fase 3 envolveu a participação de especialistas, com vistas a validar o *framework* para avaliação de agentes conversacionais educacionais desenvolvido no presente trabalho. Neste estudo, entende-se especialistas como profissionais que atuam como desenvolvedores,

professores de Tecnologia da Informação, pesquisadores e autores de artigos de agentes conversacionais ou *chatbot*.

Para essa fase, foram enviados e-mails aos participantes, solicitando respostas em relação às métricas. Em arquivos anexos, estavam o formulário com todas as métricas de avaliação e o Termo de Consentimento Livre Esclarecido. O respectivo formulário pode ser conferido no Apêndice H. Os e-mails foram enviados aos seguintes especialistas:

- Lista de e-mail do PPGIE. O e-mail foi enviado pela Secretaria do Programa;
- Grupo no total de 40 professores da área de Tecnologia da Informação;
- Autores de artigos com publicações sobre o tema Agentes Conversacionais ou *Chatbot*, totalizando 42 autores;
- Profissionais de Tecnologia da Informação.

O formulário desenvolvido para a aplicação na Fase 3 (disponível no Apêndice H) envolve métricas da ISO/IEC 9126-1, ISO/IEC 25010:2011 e métricas de trabalhos relacionados. Neste contexto, a análise dos especialistas é importante para validar as métricas, analisar se faz sentido o nome da métrica e, também, a relação das métricas em suas respectivas categorias, subcategorias e possibilidades de questionamento para avaliação.

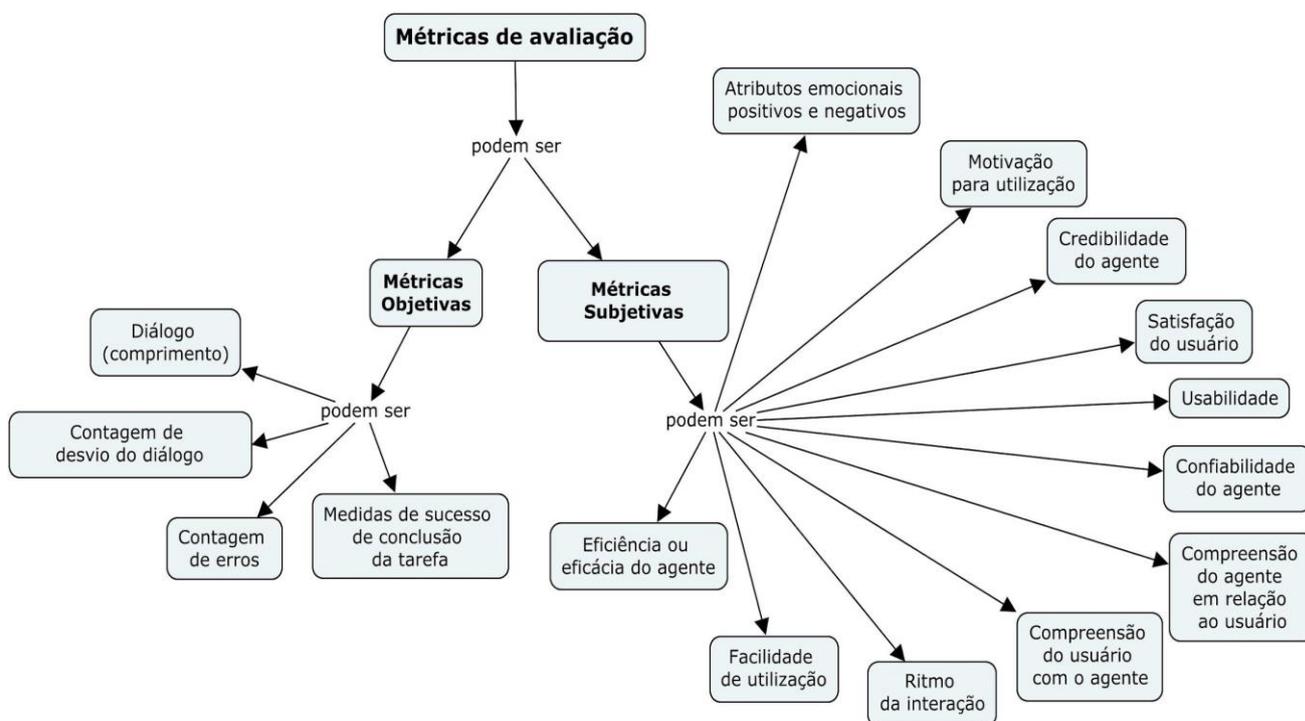
O detalhamento do percurso de construção do *framework* pode ser conferido na próxima seção, onde será apresentado detalhadamente e sistematicamente como o instrumento que será utilizado na fase 3.

4 PERCURSO DE CONSTRUÇÃO DO *FRAMEWORK*

Esta seção tem o objetivo de apresentar a proposta do desenvolvimento do *framework* para avaliação de *chatbots* voltados ao uso educacional. Será demonstrado o percurso da construção e o levantamento das métricas para a avaliação de agentes conversacionais até as métricas escolhidas e desenvolvidas.

A tabela 2.2, já apresentada na subseção 2.4 da presente tese, trouxe um mapeamento das métricas utilizadas em 12 trabalhos elencados neste estudo, as quais são representadas no mapa conceitual, conforme está demonstrado a seguir.

Figura 4.1 – Métricas de avaliação para agentes conversacionais



Fonte: Silveira et al. (2017)¹⁸.

Conforme o mapa conceitual elaborado pelo autor da presente tese, as métricas de avaliação podem ser objetivas e subjetivas. Nas objetivas, estão alguns itens de avaliação, como comprimento do diálogo, contagem de desvio, contagem de erros e medidas de sucesso na conclusão de uma tarefa. Phillips (2018) e O'Shea, Bandar e Crockett (2011) também

¹⁸ As métricas de avaliação para agentes conversacionais foram construídas em conjunto com a Profa. Dra. Liane Margarida Rockenbach Tarouco e com os colegas de doutorado, na disciplina Projetos - Projeto e Desenvolvimento de Agentes Conversacionais II, em 2017/2.

apresentam as respectivas métricas objetivas conforme já identificadas nos trabalhos relacionados. Já as métricas envolvem métricas como eficiência, eficácia, facilidade de uso, ritmo de interação, compreensão do usuário com o agente, compreensão do agente em relação ao usuário, confiabilidade, usabilidade, satisfação do usuário, credibilidade, motivação e afetividade. As respectivas métricas também aparecem nos estudos de Niculescu et al. (2014), Radziwill e Benton (2017), Brandão, Reis e Rocha (2013), Hayashi (2016) e O'Shea, Bandar e Crockett (2011), as quais também foram apontadas nos trabalhos relacionados.

Ao se comparar as métricas sinalizadas no mapa mental identificado na figura 4.1, observa-se que algumas também são identificadas na NBR-ISO IEC-9126-1 e ISO 25010, conforme se pode identificar na tabela 4.1.

Tabela 4.1 – Comparação de métricas de avaliação para agentes conversacionais x métricas da NBR-ISO IEC-9126-1 e NBR-ISO IEC-25010

Métricas	Contempla as NBR-ISO IEC 9126-1	Contempla as NBR-ISO IEC 25010
Subjetivas		
Atributos emocionais positivos e negativos		
Motivação para utilização		X
Credibilidade do Agente	X	X
Satisfação do usuário	X	X
Usabilidade	X	X
Confiabilidade do agente	X	X
Compreensão do agente em relação ao usuário		X
Compreensão do usuário com o agente	X	X
Ritmo de interação	X	X
Facilidade de interação		
Facilidade de utilização	X	X
Eficiência ou eficácia do agente	X	X
Objetivas		
Diálogo de comprimento		
Contagem de desvio de diálogo		
Contagem de erros		
Medidas de sucesso de conclusão da tarefa		

Fonte: Elaborado pelo autor (2021).

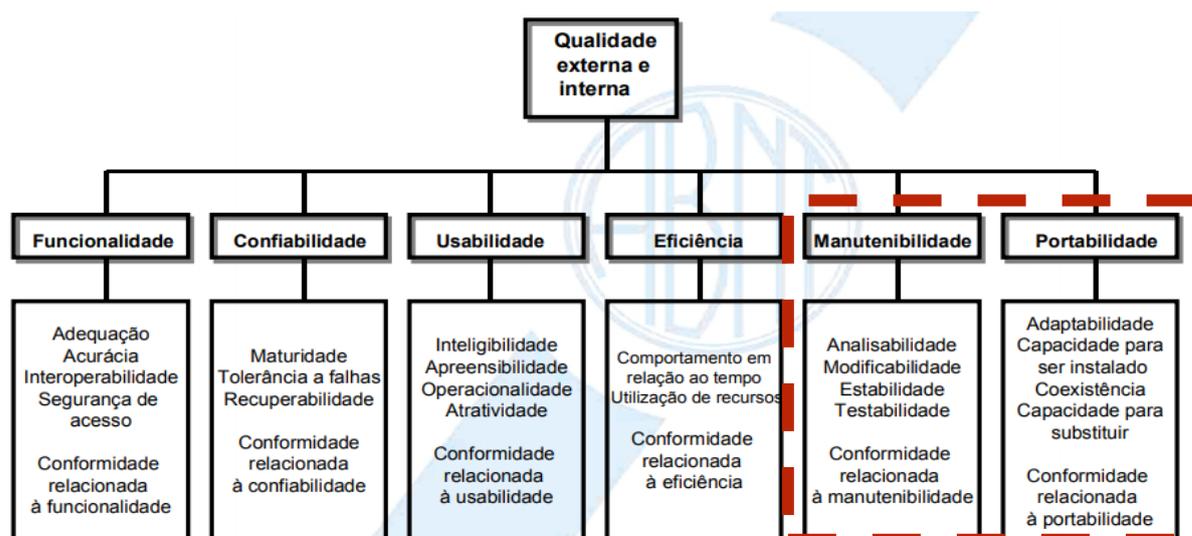
Conforme a tabela 4.1, observa-se que a NBR-ISO IEC-9126-1 contempla as métricas subjetivas Eficiência ou eficácia do agente, Facilidade de utilização, Ritmo da interação, Compreensão do usuário com o agente, Compreensão do agente em relação ao usuário, Confiabilidade do agente, Usabilidade, Satisfação do usuário e Credibilidade do agente. Já a ISO/IEC 25010 contempla as mesmas métricas da ISO/IEC 9126 acrescidas das métricas Compreensão do agente em relação ao usuário e Motivação para utilização. Por outro lado, há

diversas métricas da ISO/IEC-9126-1 e ISO/IEC 25010 que não estão contempladas nas métricas de avaliação apontadas no mapa mental (Figura 4.1).

Assim, ainda conforme a tabela 4.1, constata-se que não há um consenso entre as métricas da ISO/IEC-9126-1 e ISO 25010 e as métricas já identificadas para o uso de agentes conversacionais (Figura 4.1), cuja falta de padronização para o desenvolvimento de agentes conversacionais educacionais reforça a oportunidade em desenvolver um *framework* com recursos e métricas para a construção e avaliação especificamente de *softwares* de agentes conversacionais educacionais. Ressalta-se, ainda, que as métricas objetivas não foram contempladas pela ISO/IEC-9126-1 e pela ISO/IEC 25010.

As próximas métricas apresentadas na figura 4.2 referem-se à manutenibilidade e à portabilidade. Cabe enfatizar que elas não foram identificadas nos estudos relacionados.

Figura 4.2 – Manutenibilidade e Portabilidade: Características da NBR-ISO IEC-9126-1 que não aparecem nos estudos relacionados

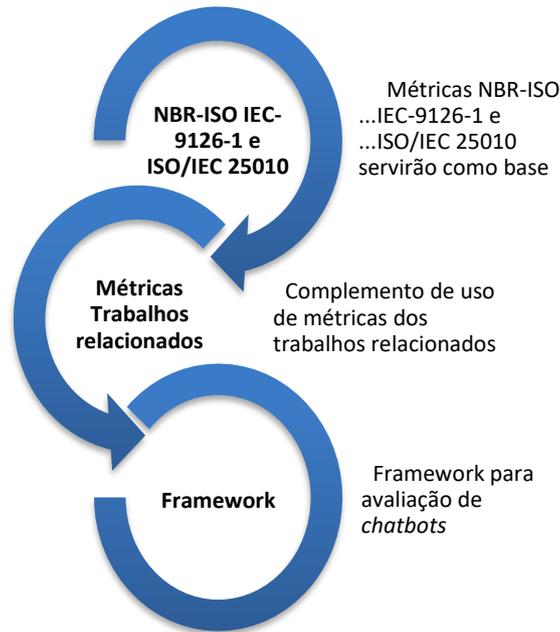


Fonte: Elaborado pelo autor, com base na NBR-ISO IEC-9126.

Nesse sentido, compreende-se que é importante utilizar as métricas dos trabalhos relacionados apresentados na Tabela 2.2 e adicionar as métricas não contempladas pela NBR-ISO IEC-9126-1 apontadas na figura 4.2 (Manutenibilidade e Portabilidade). Também serão contempladas as métricas da ISO/IEC 25010.

Na sequência, apresenta-se o percurso de construção das métricas para a avaliação de agente conversacional.

Figura 4.3 - Percurso das métricas para desenvolvimento do *framework*



Fonte: Elaborado pelo autor (2021).

A tabela 4.2 apresenta a compilação das métricas a partir da ISO/IEC 25010, dos Trabalhos Relacionados e já apresentados na subseção 2.3 e, também, da figura 4.1, que faz a comparação das métricas de avaliação para agentes conversacionais, a fim de se fazer um levantamento das que foram utilizadas para a construção do *framework*.

A compilação das respectivas métricas apresentadas na tabela 4.2 dão origem à estrutura do *framework*. Dessa forma, nesta tabela, a coluna Categoria apresenta o nome da categoria sugerida pelo autor da tese. Já a coluna Subcategoria relaciona as métricas que pertencem às respectivas Categorias apresentadas na primeira coluna; e, por fim, a coluna Origem traz a relação de autores que constam na subseção 2.3 dos trabalhos relacionados, bem como as métricas que tiveram origem da figura 4.1 ou na ISO / IEC 25010.

Tabela 4.2 – Compilação das métricas ISO/IEC 9126, 25010, Figura 4.1 e trabalhos relacionados

Categoria	Subcategoria	Origem
Funcionalidade	Funcionalidade completa	Qualidade do produto - ISO / IEC 25010 O’Shea, Bandar e Crockett (2011) Figura 4.1
	Correção funcional	Qualidade do produto - ISO / IEC 25010
	Adequação funcional	Qualidade do produto - ISO / IEC 25010
	Qualidade das respostas	Goh, Kumar e Choon(2016) Corporate <i>Living Actor</i> (2019) Sedoc et al. (2019) Hayashi (2016) O’Shea, Bandar e Crockett (2011)
	Respostas de <i>fallback</i>	AI Multiple (2019)

		Phillips (2018) <i>Corporate Living Actor</i> (2019) O'Shea, Bandar e Crockett (2011)
	Métricas de mensagens	Phillips (2018) Sedoc et al. (2019) Hayashi (2016) O'Shea, Bandar e Crockett (2011) Griol e Molina (2015) Figura 4.1
Eficiência	Comportamento em relação ao tempo	Qualidade do produto - ISO / IEC 25010 Radziwill e Benton (2017) Brandão, Reis e Rocha (2013) <i>Corporate Living Actor</i> (2019) O'Shea, Bandar e Crockett (2011) Figura 4.1
	Utilização de recursos	Qualidade do produto - ISO / IEC 25010
	Eficácia	ISO 25010 – Qualidade em Uso Qualidade do produto - ISO / IEC 25010 Estudos relacionados: 2, 6, 11 Radziwill e Benton (2017) AI Multiple (2019) O'Shea, Bandar e Crockett (2011)
	Capacidade	Qualidade do produto - ISO / IEC 25010
Compatibilidade	Coexistência	Qualidade do produto - ISO / IEC 25010
	Interoperabilidade	Qualidade do produto - ISO / IEC 25010
Usabilidade	Reconhecimento de adequação	Qualidade do produto - ISO / IEC 25010 Niculescu et al. (2014) O'Shea, Bandar e Crockett (2011) Figura 4.1
	Aprendizagem	Qualidade do produto - ISO / IEC 25010
	Operabilidade	Qualidade do produto - ISO / IEC 25010
	Proteção contra erros de usuário	Qualidade do produto - ISO / IEC 25010 Figura 4.1
	Estética da interface do usuário	Qualidade do produto - ISO / IEC 25010
Confiabilidade	Acessibilidade	Qualidade do produto - ISO / IEC 25010 Estudo Relacionado: 2
	Maturidade	Qualidade do produto - ISO / IEC 25010 Niculescu et al. (2014) O'Shea, Bandar e Crockett (2011) Figura 4.1
	Disponibilidade	Qualidade do produto - ISO / IEC 25010
	Tolerância ao erro	Qualidade do produto - ISO / IEC 25010 Figura 4.1
	Recuperabilidade	Qualidade do produto - ISO / IEC 25010
Segurança	Confidencialidade	Qualidade do produto - ISO / IEC 25010
	Integridade	Qualidade do produto - ISO / IEC 25010
	Não Repúdio	Qualidade do produto - ISO / IEC 25010
	Prestação de contas	Qualidade do produto - ISO / IEC 25010
	Autenticidade	Qualidade do produto - ISO / IEC 25010
Manutenção	Modularidade	Qualidade do produto - ISO / IEC 25010
	Reutilização	Qualidade do produto - ISO / IEC 25010
	Analisabilidade	Qualidade do produto - ISO / IEC 25010
	Modificabilidade	Qualidade do produto - ISO / IEC 25010
	Testabilidade	ISO 25010 Brandão, Reis e Rocha (2013) Satu, Parvez e Mamun (2015) Goh, Kumar e Choon (2016)

		Sedoc et al. (2019)
Portabilidade	Adaptabilidade	ISO 25010
	Instalabilidade	ISO 25010
	Substituidade	ISO 25010
Satisfação	Utilidade / Motivação para utilização	ISO 25010 – Qualidade em Uso Figura 4.1
	Confiança	ISO 25010 – Qualidade em Uso
	Prazer/satisfação	ISO 25010 – Qualidade em Uso Estudos relacionados: 7, 8, 10 Phillips (2018) Corporate Living Actor (2019) Hayashi (2016) Figura 4.1
	Conforto	ISO 25010 – Qualidade em Uso
	Afetividade	Estudos relacionados: 2, 10, 11 Radziwill e Benton (2017) Hayashi (2016) O’Shea, Bandar e Crockett (2011) Figura 4.1
	Humanidade	Radziwill e Benton (2017) O’Shea, Bandar e Crockett (2011)
	Ética e Comportamento	Radziwill e Benton (2017) Tabela 4.1
Livre de riscos	Riscos de saúde e segurança	ISO 25010 – Qualidade em Uso
	Risco ambiental	ISO 25010 – Qualidade em Uso
Cobertura de contexto	Compleitude do contexto	ISO 25010 – Qualidade em Uso
	Flexibilidade	ISO 25010 – Qualidade em Uso Figura 4.1
	Manter a conversa	Estudos relacionados: 6, 7, 8, 11, 12 AI Multiple (2019) Phillips (2018) Corporate Living Actor (2019) O’Shea, Bandar e Crockett (2011). Griol e Molina (2015) Tabela 4.1
Controle de usuários	Usuários ativos	AI Multiple (2019) Phillips (2018)
	Total de usuários	Estudos relacionados: 6

Fonte: Elaborado pelo autor (2021).

Assim, conforme o levantamento das métricas da tabela 4.2, constata-se que a categoria Funcionalidade aparece tanto nas normas ISO/IEC 25010, ISO/IEC 9126, como nos trabalhos relacionados. A categoria Funcionalidade aparece principalmente em relação às subcategorias Qualidade das respostas e métricas de Respostas de *fallback* (captura a quantidade de vezes que o *chatbot* falhou ou passou por uma situação de quase falha). Isso demonstra o quanto a categoria funcionalidade é importante, segundo os estudos realizados e sinaliza que os *chatbots* já elaborados têm a categoria Funcionalidade como uma relação a responder corretamente aos estudantes.

A categoria Eficiência aparece bastante nos estudos relacionados e também na norma ISO/IEC 25010, em que houve preocupação e discussão na subcategoria de Comportamento em relação ao tempo e, também, em relação à eficácia, aparecendo tais subcategorias seis vezes.

A Categoria Compatibilidade apareceu apenas na norma ISO/IEC 25010 e em nenhum dos doze estudos relacionados abordou ou tratou da categoria Compatibilidade.

Já a Categoria Usabilidade e suas respectivas subcategorias apareceram apenas na ISO/IEC 25010, destacando-se apenas a subcategoria Reconhecimento de adequação nos Estudos Relacionados de Niculescu et al. (2014) e O'Shea, Bandar e Crockett (2011), bem como no levantamento de requisitos apresentados pelo autor na figura 4.1. Assim, percebe-se que os trabalhos relacionados não avaliaram a métrica Usabilidade.

A Categoria Confiabilidade aparece na ISO/IEC 25010 e nos trabalhos relacionados, no entanto sem detalhamento de qual a subcategoria. A Categoria Segurança é abordada apenas na ISO/IEC 25010, não sendo contemplada nos Trabalhos Relacionados pesquisados. Por outro lado, a subcategoria Testabilidade da Categoria Manutenção aparece em 5 Trabalhos Relacionados e, também, na ISO/IEC 25010. A Categoria Portabilidade também aparece apenas na ISO/IEC 25010.

A Categoria Satisfação registrou apenas as subcategorias Prazer/satisfação junto aos Trabalhos Relacionados de Phillips (2018), site *Living Actor*, Hayashi (2016) e a figura 4.1. Por outro lado, destaca-se que aparecem em três trabalhos relacionados à métrica Afetividade: pesquisas de Radziwill e Benton (2017), Hayashi (2016) e O'Shea, Bandar e Crockett (2011) e na figura 4.1. A métrica Humanidade foi identificada nos trabalhos de Radziwill e Benton (2017) e O'Shea, Bandar e Crockett (2011). Registra-se, por fim, que, na categoria Satisfação, acrescentaram-se as subcategorias Afetividade, Humanidade e Ética e Comportamento não registrados na ISO/IEC 25010.

A Categoria Livre de Riscos é contemplada apenas pela ISO/IEC 25010. A Categoria Cobertura de Contexto também é contemplada apenas pela ISO/IEC 25010, no entanto, inseriu-se nesta categoria a subcategoria "Manter a conversa" que estava como métrica nos Trabalhos Relacionados do site *AI Multiple*, Phillips (2018), site *Os agentes*, O'Shea, Bandar e Crockett (2011) e Griol e Molina (2015). Na nova Categoria Controle de usuário, foi adicionada as subcategorias Usuários ativos e Total de usuários.

Assim, o *framework* proposto é importante e inovador por mapear, identificar e validar métricas para a avaliação e construção de agentes conversacionais, pois, de acordo com a tabela 4.2, constata-se que não há um consenso entre a definição de métricas para a avaliação de *softwares*. Afinal, diferentes métricas foram levantadas na ISO/IEC 25010, nos trabalhos

relacionados e nas métricas compiladas, as quais foram apresentadas na figura 4.1. Ressalta-se que diversas métricas aparecem em diferentes contextos; nesses casos, geralmente, é levado em consideração o contexto do trabalho.

Além disso, na tentativa de encontrar e conciliar as métricas mencionadas anteriormente, percebe-se que há muitas categorias e subcategorias para testar o agente conversacional e outras destinadas ao desenvolvimento do agente conversacional. Assim, a partir do levantamento das métricas descritas na tabela 4.2, acredita-se que o *framework* para avaliação de agentes conversacionais poderá atender a diferentes grupos de usuários, como Usuários, Testadores e Desenvolvedores.

Pensando em um público potencial para o uso do *framework* – e, também, em classificar esse público que poderá usufruí-lo, entende-se, neste estudo, que os Usuários são os estudantes que desejam utilizar as métricas de avaliação dos agentes conversacionais. Os Testadores são usuários avançados (alunos e professores) que tem como objetivo avaliar métricas de funcionamento dos agentes. Já os Desenvolvedores são os programadores, analistas de sistemas e profissionais de tecnologia da informação que atuarão no desenvolvimento dos *chatbots* educacionais e precisam de métricas como base para desenvolver os parâmetros desses agentes conversacionais.

Acredita-se que, muitas vezes, as métricas dos usuários são diferentes das métricas dos especialistas, pois parte-se da hipótese de que um especialista pode se preocupar com questões relacionadas à funcionalidade, à eficiência, à portabilidade ou à manutenção. A métrica de Manutenção e Portabilidade pode estar mais direcionada para o tipo de usuários desenvolvedores de *chatbots*, enquanto o Usuário pode tentar entender simplesmente se o *chatbot* é capaz de fornecer respostas adequadas durante os diálogos. Assim, sugerem-se as categorias dos três perfis de usuários que poderão utilizar as métricas:

- a) Métricas para o Perfil Usuários: Funcionalidade, Usabilidade, Satisfação e Cobertura de contexto;
- b) Métricas para o Perfil de Testadores: Funcionalidade, Eficiência, Usabilidade, Confiabilidade, Satisfação, Cobertura de contexto, Controle de usuários;
- c) Métricas para o Perfil de Desenvolvedores: Compatibilidade, Usabilidade, Confiabilidade, Segurança, Manutenção, Portabilidade, Livre de riscos.

Conforme as categorias sugeridas, constata-se que a categoria Funcionalidade pode ser mais adequada ao Perfil de Usuários e Testadores. Já as Categorias Manutenção e Portabilidade estão mais relacionadas ao tipo de usuário Testadores e Desenvolvedores. Vale salientar que a proposta não é criar ou dividir as métricas selecionadas para o *framework* para diferentes perfis

de usuários, mas, sim, demonstrar que ele pode ser utilizado e personalizado para diferentes finalidades a partir de distintos interesses, afinal, parte-se do princípio de que os Perfis de Usuários e de Desenvolvedores tenham diferentes olhares e interesses para avaliar e construir agentes conversacionais.

Uma vez que houve o levantamento das métricas de avaliação de agentes conversacionais a partir da ISO/IEC 9126-1, da ISO/IEC 25010 e dos trabalhos relacionados, elaborou-se uma compilação que pode ser conferida na tabela 4.3. Esta tabela descreve as Categorias, as Subcategorias e um exemplo de pergunta para avaliação das métricas. Nela, ainda há espaço para os campos Aceitar, Rejeitar e Ajuste Sugerido. Foi utilizada para a avaliação da Fase 3, período em que os especialistas sinalizaram se aceitariam ou rejeitariam as métricas apresentadas para composição do *framework*.

A seguir, pode-se observar os campos descritos, conforme na tabela 4.3:

- A Categoria é uma grande área de atuação, como Funcionalidade e Usabilidade, por exemplo.
- As métricas utilizadas nas Subcategorias estão distribuídas de acordo com a relação ou aproximação com as Categorias.
- A origem (autores, Métricas ISO/IEC ou figura 4.1) de cada categoria e das subcategorias foi identificada na tabela 4.2.
- A coluna Pergunta trata de uma questão que foi utilizada para avaliar a métrica de avaliação do agente conversacional. Esta pergunta foi utilizada como exemplo, mas pode ser configurada e personalizada no *framework*.
- O campo Aceitar e Rejeitar deve ser preenchido pelos especialistas, caso concordem ou não com a métrica identificada em suas Subcategorias.
- A coluna Ajustes sugeridos trata de um espaço no qual o especialista pode registrar suas observações sobre a métrica avaliada.

Tabela 4.3 – Categorias, subcategorias e perguntas

Categoria	Subcategoria	Pergunta	Aceitar	Rejeitar	Ajuste sugerido
Funcionalidade	Funcionalidade completa	O <i>chatbot</i> apresenta funções que atendam a sua funcionalidade completa?			
	Correção funcional	Não será utilizado.			
	Adequação funcional	Não será utilizado.			
	Qualidade das respostas	O <i>chatbot</i> apresenta respostas adequadas à conversação?			
	Respostas de <i>fallback</i>	O <i>chatbot</i> é capaz de apresentar a quantidade de vezes em que falhou ou passou por uma situação de falha?			
	Métricas de mensagens	O <i>chatbot</i> apresenta mensagens iniciais de conversação ou saudações, a fim de iniciar um diálogo?			
Eficiência	Comportamento em relação ao tempo	O <i>chatbot</i> apresenta um tempo rápido de respostas?			
	Utilização de recursos	Não será utilizado.			
	Eficácia	O <i>chatbot</i> atende, com precisão e integridade, quanto ao objetivo do diálogo?			
	Capacidade	Não será utilizado.			
Compatibilidade	Coexistência	Não será utilizado.			
	Interoperabilidade	O <i>chatbot</i> é capaz de interagir com outro sistema ou plataforma específica.			
Usabilidade	Reconhecimento de adequação	Você reconhece que o <i>chatbot</i> é apropriado para suas necessidades?			
	Aprendizagem	O <i>chatbot</i> permite aprender como usá-lo em situações de emergência?			
	Operabilidade	O <i>chatbot</i> é fácil de operar e controlar?			
	Proteção contra erros de usuário	Em caso de algum possível erro na resposta, o <i>chatbot</i> segue funcionando para novos diálogos.			
	Estética da interface do usuário	Não será utilizado.			
	Acessibilidade	O <i>chatbot</i> pode ser utilizado por pessoas com diferentes características físicas?			

Confiabilidade	Maturidade	O <i>chatbot</i> atende às necessidades de confiabilidade em operação normal?			
	Disponibilidade	Quando você precisou, o <i>chatbot</i> estava disponível ou vai estar no momento em que você precisar?			
	Tolerância ao erro	Caso algum <i>software</i> do seu notebook, computador ou telefone apresente um possível erro de atualização em outro aplicativo, o <i>chatbot</i> segue funcionando normalmente?			
	Recuperabilidade	Não será utilizado.			
Segurança	Confidencialidade	O <i>chatbot</i> tem algum tipo de usuário e senha, ou ainda, um <i>link</i> específico para que seja possível utilizá-lo?			
	Integridade	O <i>chatbot</i> garante que o usuário não irá alterar determinadas configurações, como programações, botões, informações na base de conhecimento?			
	Não Repúdio	Não será utilizado.			
	Prestação de contas	Não será utilizado.			
	Autenticidade	O <i>chatbot</i> permite sua identificação, a fim de garantir o sigilo ou diálogos já estabelecidos?			
Manutenção	Modularidade	Não será utilizado.			
	Reutilização	Informações da base de conhecimentos poderão ser aproveitadas para outros futuros <i>chatbots</i> ?			
	Analisabilidade	Não será utilizado.			
	Modificabilidade				
	Testabilidade	O <i>chatbot</i> possibilita estabelecer critérios de testes a serem realizados, a fim de determinar se os critérios estabelecidos foram atendidos?			
Portabilidade	Adaptabilidade	O <i>chatbot</i> funciona em diferentes sistemas operacionais e em diferentes configurações de <i>hardware</i> ?			
	Instalabilidade	Não será utilizado.			
	Substituidade	Não será utilizado.			
Satisfação	Utilidade / Motivação para utilização	Após usar o <i>chatbot</i> , sinto-me satisfeito em relação ao uso de meus objetivos, tais como os resultados do uso?			
	Confiança	Confio que o <i>chatbot</i> vai comportar-se conforme o pretendido?			
	Prazer/satisfação	O <i>chatbot</i> atendeu as suas necessidades pessoais?			
	Conforto	Não será utilizado.			

	Afetividade	O <i>chatbot</i> apresenta diálogos com saudações, cordialidade e é capaz de responder, com cordialidade, as interações humoradas dos participantes?			
	Humanidade	O <i>chatbot</i> é capaz de interagir de maneira convincente, satisfatória e natural?			
	Ética e Comportamento	O <i>chatbot</i> apresenta respeito, inclusão e preservação da dignidade durante os diálogos?			
Livre de riscos	Riscos de saúde e segurança	Não será utilizado.			
	Risco ambiental	Não será utilizado.			
Cobertura de contexto	Compleitude do contexto	Não será utilizado.			
	Flexibilidade	Não será utilizado.			
	Manter a conversa	O <i>chatbot</i> é capaz de manter o diálogo?			
Controle de usuários	Usuários ativos	O <i>chatbot</i> permite identificar os usuários ativos ou on-line?			
	Total de usuários	O <i>chatbot</i> permite identificar o total de usuários que conversaram com os usuários?			

Fonte: Elaborado pelo autor (2021).

Uma vez apresentado o percurso da construção das métricas para o *Framework*, a próxima seção aborda um sistema desenvolvido especificamente para automatizar a relação de métricas (categorias, subcategorias, questionamentos para avaliação, definição e escolha das métricas) ou o gerenciamento automático delas, dos questionários, e da automatização das respostas.

No presente capítulo, abordou-se a construção e percurso das métricas para a construção e avaliação de *chatbots*. A próxima subseção apresenta o sistema desenvolvido especificamente para a presente tese, com a finalidade de automatizar o uso das métricas, as avaliações e questionários avaliativos.

4.1 CHATAVAL: APLICATIVO WEB PARA AUTOMATIZAR O FRAMEWORK

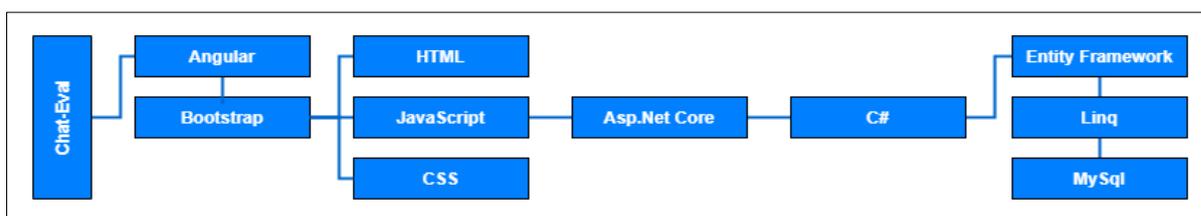
Com o intuito de sistematizar o *framework*, desenvolveu-se um aplicativo com o objetivo de automatizar o processo de inserção de categorias, subcategorias, escolha das métricas de avaliação, aplicação dos questionários e entrega automatizada dos resultados avaliados a partir das métricas escolhidas. O *ChatAval* é uma contribuição para a automatização do *framework* proposto nesta tese. Ele foi desenvolvido pelo ex-orientando do autor desta tese, aluno Thiagus Ferreira, como produto de seu Trabalho de Conclusão de Curso apresentado à Faculdade CNEC Gravataí¹⁹. O período de desenvolvimento e testes do *ChatAval* envolveu o segundo semestre de 2020 e o primeiro semestre de 2021.

Desenvolvido como uma aplicação *web*, cada funcionalidade do *ChatAval* foi alinhada minuciosamente para automatizar o *framework* proposto neste estudo, o que permitiu a configuração de métricas de categorias e de subcategorias, a criação dos formulários avaliativos, a disponibilização de preenchimento automatizado por parte dos usuários desses formulários e a visualização de um quadro comparativo dos resultados das pesquisas, para o caso de haver mais um agente conversacional avaliado. Atualmente, o *ChatAval* está hospedado nos servidores do Programa de Pós-Graduação em Informática da Educação, da Universidade Federal do Rio Grande do Sul.

A figura 4.4 apresenta as tecnologias utilizadas na construção da ferramenta *ChatAval* para automatizar o *framework* proposto na presente tese.

¹⁹ Mais informações sobre a Faculdade CNEC Gravataí em: <https://educacaosuperior.cneccbr.com.br/gravatai>

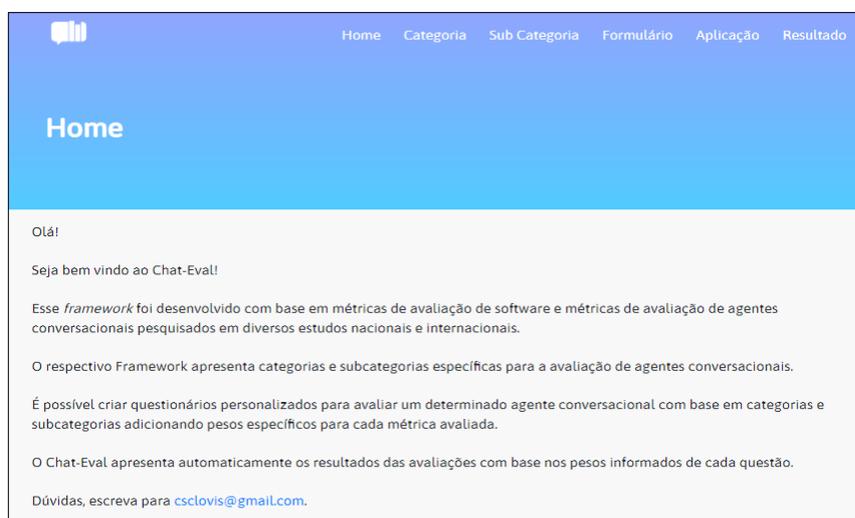
Figura 4.4 – Organograma das tecnologias envolvidas no protótipo



Fonte: Ferreira (2021).

A proposta do *ChatAval* envolve o desenvolvimento da página *web* e os serviços que a ela respondem. Para o desenvolvimento da interface gráfica, o que corresponde ao *front-end*, utilizou-se o Angular²⁰ em conjunto com o *bootstrap*²¹, os quais produzem todo o HTML, *JavaScript* e *CSS*²² necessário. Para o serviço *back-end*, isto é, para o que está por trás da aplicação do *ChatAval*, utilizou-se o Asp.Net Core²³ com a linguagem C#, acessando-se o banco de dados criado em MySQL²⁴, por meio do Linq²⁵, como parte do *Entity Framework*²⁶.

Apresentam-se, a seguir, as telas e funcionalidades do *ChatAval*, começando pela tela inicial do aplicativo, conforme demonstra a figura 4.5.

Figura 4.5 - Tela inicial do *ChatAval*

Fonte: Ferreira (2021).

²⁰ Informações sobre o Angular disponível em: https://www.hostgator.com.br/blog/o-que-e-o-angular/?gclid=CjwKCAiA1JGRBhBSEiwAxXblwVNnS0RTeISkx75WmeTgNeBvKRVNwkt3HAMjHXBL1xBRIBtECo9csBoCiuQQA_vD_BwE. Acesso em: março de 2022.

²¹ Mais informações sobre o Bootstrap no site: <https://bityli.com/SFhjw>. Acesso em: março de 2022.

²² Site para mais informações sobre HTML, CSS e JS: <https://bityli.com/OVvqZ>. Acesso em: março de 2022.

²³ Site para mais informações sobre o Asp.Net Core: <https://bityli.com/XiCUa>. Acesso em: março de 2022.

²⁴ Site para mais informações sobre o MySQL: <https://bityli.com/uXRMQ>. Acesso em: março de 2022.

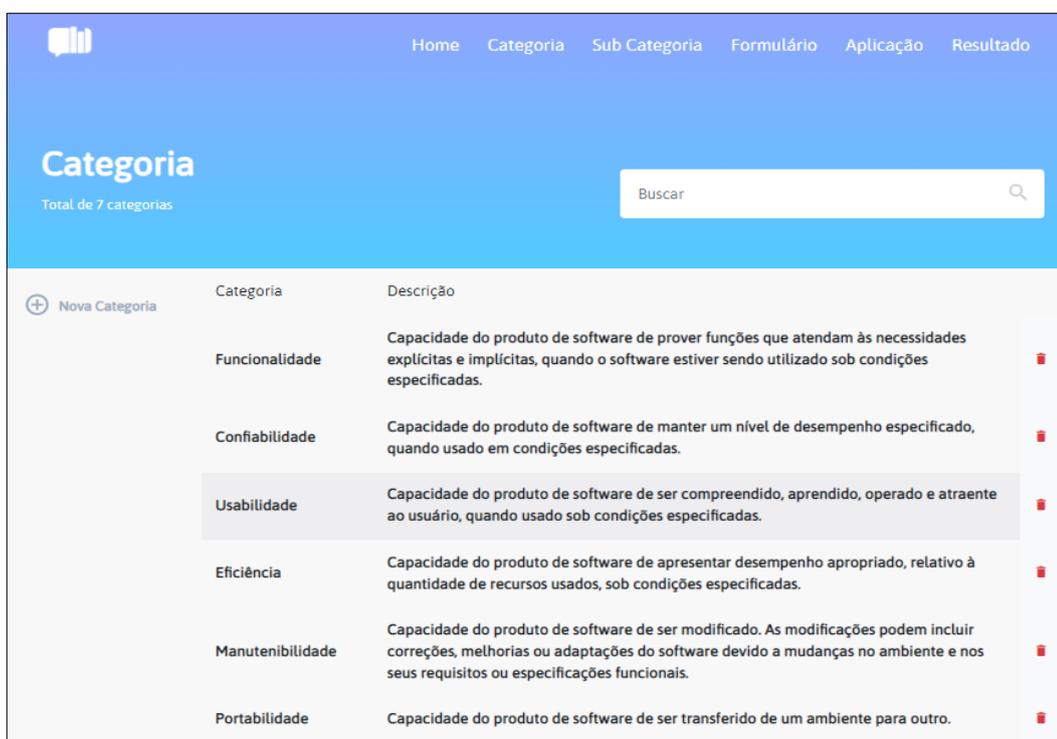
²⁵ Site para mais informações sobre o LINK: <https://bityli.com/WlxiW>. Acesso em: março de 2022.

²⁶ Mais informações sobre o *Entity Framework* no site: <https://bityli.com/txnRe>. Acesso em: março de 2022.

Conforme a figura 4.5, a tela de boas-vindas apresenta os menus Categoria, Subcategoria, Formulário, Aplicação e Resultado, que são funcionalidades específicas para o funcionamento do *ChatAval* alinhado estrategicamente com a proposta do *framework* desenvolvida neste estudo. Nesta tela, observa-se um texto de boas-vindas com algumas informações técnicas passadas para os autores.

Já na tela de categorias, são reveladas todas as categorias desenvolvidas no *framework*, como demonstrado na figura 4.6.

Figura 4.6 - Tela com os dados de categorias



+ Nova Categoria	Categoria	Descrição	
	Funcionalidade	Capacidade do produto de software de prover funções que atendam às necessidades explícitas e implícitas, quando o software estiver sendo utilizado sob condições especificadas.	■
	Confiabilidade	Capacidade do produto de software de manter um nível de desempenho especificado, quando usado em condições especificadas.	■
	Usabilidade	Capacidade do produto de software de ser compreendido, aprendido, operado e atraente ao usuário, quando usado sob condições especificadas.	■
	Eficiência	Capacidade do produto de software de apresentar desempenho apropriado, relativo à quantidade de recursos usados, sob condições especificadas.	■
	Manutenibilidade	Capacidade do produto de software de ser modificado. As modificações podem incluir correções, melhorias ou adaptações do software devido a mudanças no ambiente e nos seus requisitos ou especificações funcionais.	■
	Portabilidade	Capacidade do produto de software de ser transferido de um ambiente para outro.	■

Fonte: Ferreira (2021).

Ressalta-se que é possível cadastrar uma nova categoria e alterar ou excluir uma categoria existente. Ao cadastrar-se uma nova categoria, é importante informar um texto com sua descrição para auxiliar a documentação. Todas as demais subcategorias, que são as métricas, devem ser cadastradas, obrigatoriamente, dentro de uma categoria.

Caso o usuário deseje cadastrar uma nova categoria, uma outra tela de inclusão é apresentada, conforme se verifica na figura 4.7.

Figura 4.7 - Tela para inclusão de uma nova categoria

Fonte: Ferreira (2021).

Além do *link* para inclusão de uma nova categoria, é possível clicar sobre uma categoria contida na lista para editá-la. Já a exclusão de uma categoria é simbolizada por um botão vermelho com um ícone de lixeira, conforme ilustrado na figura 4.8.

Figura 4.8 - Tela com os dados de subcategorias

	Categoria	Subcategoria	Descrição	Questão
Funcionalidade	5			
Confiabilidade	4			
Usabilidade	6			
Eficiência	3			
Manutenibilidade	5			
Portabilidade	5			
Nova Categoria	0			
Profa Liane	0			

Categoria	Subcategoria	Descrição	Questão
Funcionalidade	Adequação	Capacidade do produto de software de prover um conjunto apropriado de funções para tarefas e objetivos do usuário especificados.	O agente conversacional possibilita funções apropriadas para o usuário, tal como botões, caixas de diálogos?
Funcionalidade	Acurácia	Capacidade do produto de software de prover, com o grau de precisão necessário, resultados ou efeitos corretos ou conforme acordados.	O agente conversacional retorna respostas de acordo com os diálogos de forma correta pesquisados ?
Funcionalidade	Interoperabilidade	Capacidade do produto de software de interagir com um ou mais sistemas especificados.	O agente conversacional pode ser integrado com outros sistemas, links ou apps a partir de seu link?
Funcionalidade	Segurança de acesso	Capacidade do produto de software de estar de acordo com normas, convenções ou regulamentações previstas em leis e prescrições similares relacionadas à funcionalidade.	Para cadastrar novas tags na base de conhecimento do seu agente conversacional, você entende que há segurança para acessar e cadastrar novas bases de conhecimentos?

Fonte: Ferreira (2021).

No menu Aplicações, encontra-se o nome dos agentes conversacionais cadastrados junto a uma descrição comentando sobre o *chatbot* em avaliação. Neste menu, é possível que

novos agentes conversacionais sejam cadastrados no *ChatAval* e que os já cadastrados possam ser alterados ou excluídos, como demonstrado na figura 4.9.

Figura 4.9 - Tela com os dados de agentes conversacionais

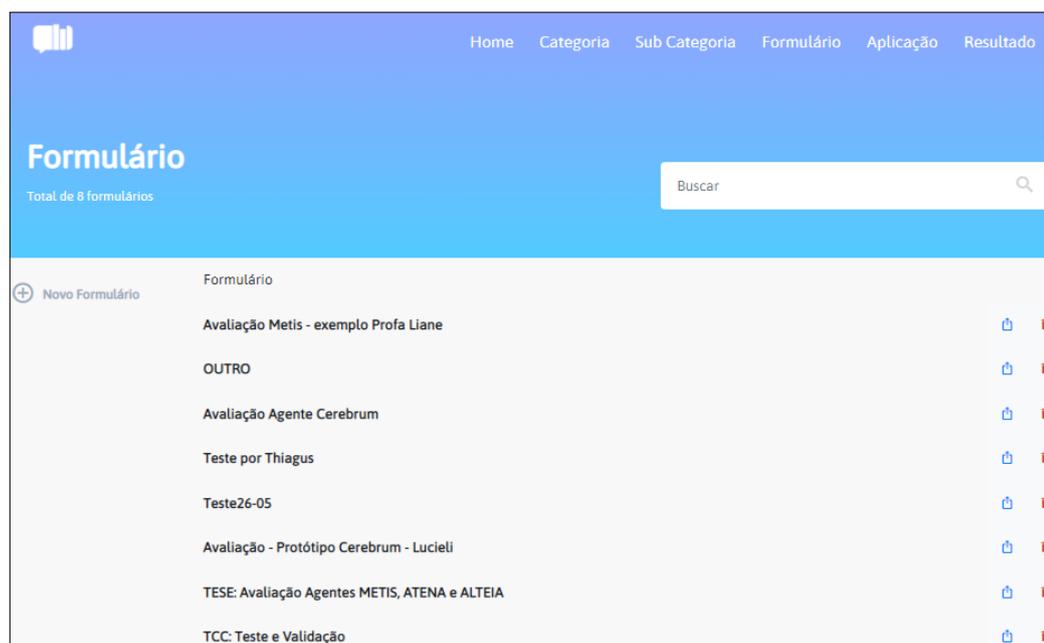
Aplicação	Descrição
+ Nova Aplicação METIS	Agente conversacional Metis (Mediadora de Educação em Tecnologia Informática e Socializadora) foi desenvolvida com a finalidade de dialogar com os alunos por meio de uma interface que simula uma conversação online (chat) entre pessoas. A base de dados conversacionais da Metis foi construída a partir de perguntas e respostas elaboradas por especialistas na área da educação, mídias e tecnologias.
Watson	Avaliação do IBM Watson
BotCin - Kátia	Tutor que apresenta conteúdos de cinemática.
Protótipo Cérebro - Lucieli	Agente Conversacional que aborda questões algébricas de primeiro grau.
ATENA	Agente Tutor para Ensino e Navegação no Ambiente – ATENA. Desenvolvido para conversação na área de física e também possui capacidade de diálogo e acompanha o estudante em seu percurso no cenário imersivo virtual. Disponível em http://avatar.cinted.ufrgs.br/atenaweb/gui/?query/
ALTEIA	Agente Lexical de apoio a Tarefas Educacionais – ALTEIA. Chatbot especialista em Metodologias Ativas, Sala de Aula invertida e ferramentas de autoria. Possui capacidade de entregar respostas aos estudantes em relação a conceitos e responder a diversas dúvidas relacionadas a sua especialidade. Disponível em: http://avatar.cinted.ufrgs.br/alteia/gui/plain/index.php#end

Fonte: Ferreira (2021).

Nesta figura, observa-se que as aplicações cadastradas são os agentes conversacionais METIS, Watson, Botclin, Protótipo Cérebro, ATENA e ALTEIA, os quais foram incluídos pelo autor da tese para testes e experimentos durante o desenvolvimento, construção e validação do protótipo *ChatAval*.

Para que se possa cadastrar as questões avaliativas de um determinado agente conversacional educacional, são necessárias as categorias e as subcategorias cadastradas no *ChatAval*. A partir dessas informações, é possível incluir um formulário avaliativo e alterar ou excluir um existente, como se verifica na figura 4.10.

Figura 4.10 - Tela com os dados de formulários



Fonte: Ferreira (2021).

Para criar um formulário avaliativo, basta clicar sobre o *link* Novo Formulário ou sobre um já existente da lista. A figura 4.11, a seguir, demonstra esta janela apresentada para inclusão e alteração de formulários avaliativos.

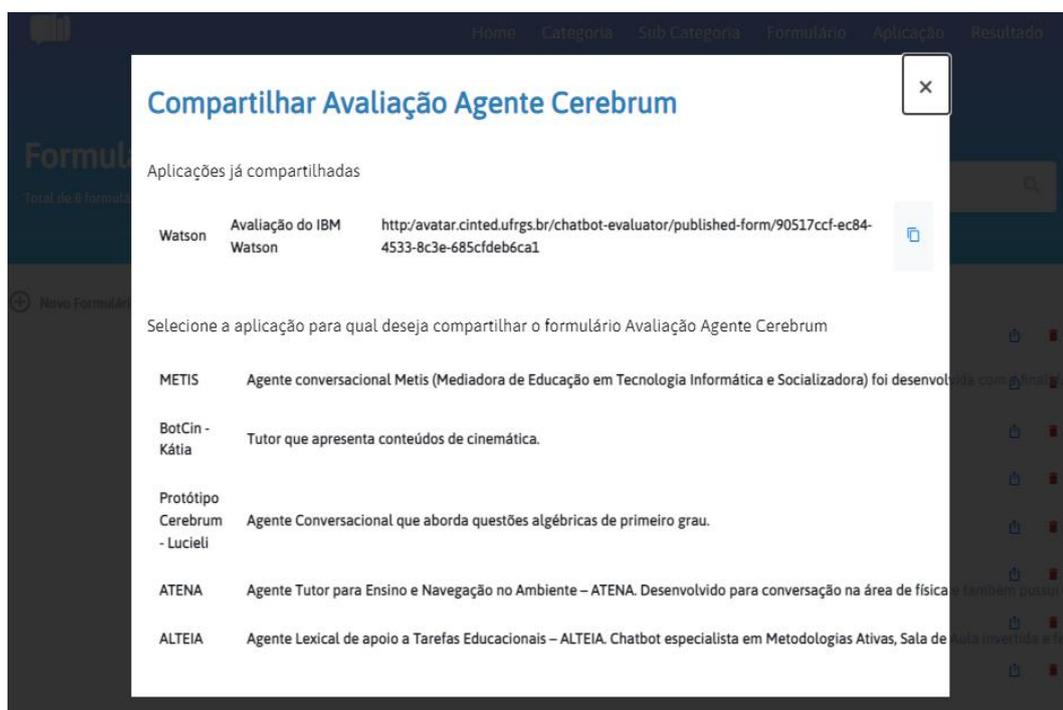
Figura 4.11 - Tela para inclusão de um novo formulário

Fonte: Ferreira (2021).

Na tela de inclusão, pode-se informar o nome do novo formulário, alterar as informações do texto inicial, que aparecerá durante as avaliações, e selecionar as categorias e subcategorias que serão avaliadas. Ao escolher-se uma subcategoria, a categoria a qual ela pertence é igualmente selecionada e a questão de avaliação se torna um campo texto, permitindo fazer ajustes na pergunta padrão. Também é apresentado um campo para o peso da pergunta ou da subcategoria dentro daquela categoria. Desse modo, cada métrica deve receber um peso e a soma dos pesos das métricas precisa ser igual a 10,0, assim como a soma do peso de cada categoria também precisa ser igual a 10,0.

Após a finalização do formulário avaliativo, é necessário vincular o questionário ao agente conversacional educacional que se deseja examinar, conforme se observa na figura 4.12. Enfatiza-se que é possível avaliar mais de um agente conversacional educacional, considerando as mesmas métricas, questões e pesos. Isso permite estabelecer uma comparação a partir dessas mesmas métricas de avaliação.

Figura 4.12 - Tela para vincular o formulário a um agente conversacional



Fonte: Ferreira (2021).

A tela de vincular o formulário a um *chatbot* oferece duas seções: a primeira apresenta a relação dos agentes conversacionais já cadastrados ao questionário; enquanto a segunda mostra a lista dos não cadastrados, para que possam ser vinculados ao questionário elaborado. No momento em que um questionário é vinculado a um agente conversacional cadastrado, o

ChatAval cria automaticamente um *link*, que direciona o usuário para um formulário avaliativo, conforme se pode constatar na figura 4.13, que ilustra algumas questões e opções de resposta.

Figura 4.5 - Tela para responder um formulário sobre um agente conversacional

Avaliação
Referente à aplicação Protóti

O agente conversacional METIS pode ser integrado com outros sistemas, links ou apps a partir de seu link?

1 2 3 4 5

Para cadastrar novas tags na base de conhecimento no agente conversacional METIS você entende que há segurança para acessar e cadastrar novas bases de conhecimentos?

1 2 3 4 5

Caso teu sistema operacional sobra alguma atualização de software e você não possa atualizar, o o agente conversacional segue funcionando?

1 2 3 4 5

Fonte: Ferreira (2021).

Quando o questionário é preenchido, pode-se visualizar os resultados já calculado pelo sistema *ChatAval*. Para se ter acesso a eles, deve-se clicar no último item do menu, como registrado na figura 4.14 a seguir.

Figura 4.6 - Tela com os dados de resultado de formulários

Home Categoria Sub Categoria Formulário Aplicação Resultado

Resultado
Total de 7 formulários com resposta

Buscar

Formulário

- Avaliação Metis - exemplo Profa Liane
- Avaliação Agente Cerebrum
- Teste por Thiagus
- Teste26-05
- Avaliação - Protótipo Cerebrum - Lucieli
- TESE: Avaliação Agentes METIS, ATENA e ALTEIA
- TCC: Teste e Validação

Fonte: Ferreira (2021).

O menu Resultado indica a relação dos formulários que contêm respostas ou, ao menos, os que receberam uma avaliação. Para obter-se detalhes desses resultados, basta acessar o formulário desejado. Ao fazer isso, uma nova janela é apresentada, conforme demonstra a figura 4.15, que ilustra dados fictícios, com o intuito de apresentar a respectiva funcionalidade.

Figura 4.7 - Tela com mais detalhes do resultado de um formulário

Resultados do formulário Avaliação Metis - exemplo Profª Liane		
Categoria	Peso	NOTAS METIS
Confiabilidade	3.0	
Maturidade	5.0	3.8
Recuperabilidade	5.0	3.8
	SUBTOTAL	- 2.3
Funcionalidade	4.0	
Acurácia	3.3	2.5
Interoperabilidade	3.3	3.3
Segurança de acesso	3.4	3.4
	SUBTOTAL	- 3.7
Usabilidade	3.0	
Apreensibilidade	3.0	3.0
Inteligibilidade	3.0	3.0
Operacionalidade	4.0	4.0
	SUBTOTAL	- 3.0
	TOTAL	- 8.9

Fonte: Elaborado pelo autor (2022).

No caso deste exemplo da figura, as categorias utilizadas foram Confiabilidade, Funcionalidade e Usabilidade. Para a categoria Confiabilidade, atribuiu-se o peso de 3,0; para a Categoria Funcionalidade 4,0; e para a Categoria Usabilidade o peso 3,0. A soma dos pesos

das três categorias totaliza 10,0. Essa é uma regra que deve ser atribuída. Caso a soma das notas das métricas seja diferente de 10,0, o sistema *ChatAval* vai informar o usuário, por meio de mensagem, e não permitirá a continuidade das demais funcionalidades.

Ainda na figura 4.15, escolheram-se as subcategorias Maturidade e Recuperabilidade, que estão dentro da categoria Confiabilidade. O peso atribuído para ambas é de 5,0. Este valor também foi configurado pelo administrador do questionário e tais notas das subcategorias poderiam ser diferentes, desde que a soma seja igual a 10,0. Caso a soma das notas atribuídas contabilize um valor diferente de 10,0, o sistema sinaliza visualmente, selecionando os campos que precisam ser reajustados com uma borda na cor vermelha, e não permitirá a continuidade do processo até que a soma dos valores das subcategorias seja igual a 10,00.

Seguindo a linha de raciocínio na figura 4.15, as subcaracterísticas Acurácia, Interoperabilidade e Segurança de acesso receberam, respectivamente, os pesos 3,3, 3,3 e 3,4. Como resultado pelos usuários, foram obtidas as notas 2,5 para Acurácia, 1,5 para Interoperabilidade e 1,5 para a métrica Segurança de acesso. Registra-se que esse exemplo fictício de avaliação apresentado na figura 4.15 teve unicamente o objetivo de demonstrar as funcionalidades e o resultado da avaliação de um ou mais agentes conversacionais a partir das métricas estabelecidas.

Pelo exposto, constata-se que o sistema está programado para calcular as médias após cada avaliação do *chatbot* avaliado. A tabela 4.4 mostra a relação entre o número escolhido na tabela *Likert* e o peso da nota atribuída a ele.

Tabela 4.4 – Pesos da Tabela *Likert*

Item atribuído na tabela <i>Likert</i>	Valor do peso atribuído
1	0,0
2	2,5
3	5
4	7,5
5	10,0

Fonte: Elaborado pelo autor (2021).

Assim, conforme a tabela 4.4, se para uma determinada métrica avaliada o avaliador escolher a opção 4 da tabela *Likert*, por exemplo, o peso ou a nota será de 7,5; se escolher a opção 2, o peso da nota será de 2,5 e assim sucessivamente.

A fórmula utilizada para essa conversão pelo *ChatAval* é apresentada na figura 4.16.

Figura 4.16 – Fórmula para o cálculo das notas

$$\frac{\text{Nota (de 1 a 5)} \times \text{peso}}{\text{nota}} =$$

Fonte: Elaborado pelo autor (2021).

De acordo com a fórmula apresentada nesta figura, caso o avaliador clique na opção 5 da tabela *Likert*, o sistema *ChatAval* vai fazer o seguinte cálculo: $((5 \times 10) / 5)$; logo, o resultado da avaliação da métrica será 10,0.

5 RESULTADOS

Este capítulo tem como propósito apresentar os resultados obtidos em cada uma das fases do presente estudo.

5.1 RESULTADOS DA FASE 1 – PERFIL DOS ALUNOS DO CURSO TÉCNICO EM INFORMÁTICA

A Fase 1 teve como objetivo fazer o levantamento do perfil dos alunos do curso Técnico em Informática para Internet. Procurou-se identificar questões relacionadas às experiências deles quanto ao uso dos agentes conversacionais e quanto os meios de acesso à internet, analisando como se dá o uso de cada ferramenta, a fim de conduzir a atividade de qualificar o público-alvo para a avaliação, para examinar, de forma preliminar, os *chatbots* avaliados neste estudo e testar as funcionalidades da ferramenta *ChatAval*. A identificação do perfil dos estudantes foi relevante para esta pesquisa pela dificuldade de encontrar estudantes ou profissionais que atendam aos critérios relacionados à experiência em uma linguagem de programação e à experiência com *chatbots*. Assim, conhecer o perfil contribuiu, de forma significativa, para qualificar o grupo de participantes.

Os resultados desta fase apontam que os alunos têm uma pequena base de conhecimento em relação à tecnologia da informação, tendo em vista que estão no primeiro ano do ensino médio, cursando, paralelamente, o Curso de Técnico em Informática para Internet. Constatou-se que o público-alvo possui computadores adequados quanto ao *hardware* e *software*, com boa velocidade de internet para navegar. A grande maioria deles conta com local propício para os estudos. Apurou-se também que, embora os alunos tenham bons recursos de *hardware*, *software* e internet, a maioria não tem base tecnológica para programação, desenvolvimento de códigos ou experiência de criação de *sites*, o que sinalizou a necessidade de aprofundar os conteúdos relatados na Fase 2.1.

A coleta de dados durante a Fase 1 possibilitou a qualificação dos alunos, relacionando a proposta de avaliação dos agentes e do *ChatAval* com o conteúdo programático proposto no conteúdo do curso. Os detalhes das características levantadas sobre o perfil dos alunos podem ser consultados no Apêndice I.

5.2 RESULTADOS DA FASE 2 – CONHECENDO E AVALIANDO OS AGENTES CONVERSACIONAIS

A Fase 2 ocorreu em dois momentos: práticas de programação (Fase 2.1) e experiências dos alunos com os agentes conversacionais (Fase 2.2).

5.2.1 Resultados Fase 2.1 – Práticas de programação

O primeiro momento da Fase foi a mais longo, uma vez que ocorreu durante os encontros nas aulas do Curso Técnico em Informática para Internet, no Colégio Senac, em São Leopoldo, e envolveu práticas de construção de *sites* e base de conhecimento em *chatbots*.

Nesta fase, constatou-se a necessidade de investir, de qualificar os alunos, para que pudessem realizar uma avaliação específica dos agentes conversacionais. Isso foi possível porque, conforme dados da Fase 1, os estudantes apresentaram uma demanda de conhecimentos técnicos em relação à linguagem de programação e aos agentes conversacionais, permitindo que ficassem preparados para se dedicarem à referida avaliação.

Nesta fase, ocorreu a apresentação dos conteúdos propostos, conforme já apresentados na Tabela 3.2. As três primeiras aulas, que contabilizaram 15 horas para cada uma das turmas, trataram de conteúdos relacionados à Introdução ao HTML, em que foram abordadas as principais *tags* de HTML e a instalação do *Visual Studio Code* para a manipulação dos respectivos códigos em HTML e AIML. A figura 5.1 registra uma das aulas híbridas, formato adotado durante a pandemia da Covid-19.

Figura 5.1 – Distribuição da turma presencial x on-line



Fonte: Elaborado pelo autor (2021).

Além das aulas e explicações necessárias sobre os conceitos para avaliação dos *chatbots*, os alunos também tinham momentos para a realização de exercícios e atividades em outros ambientes e espaços do Senac, conforme ilustrado na figura 5.2.

Figura 5.2 – Aulas em diferentes espaços do Senac



Fonte: Elaborado pelo autor (2021).

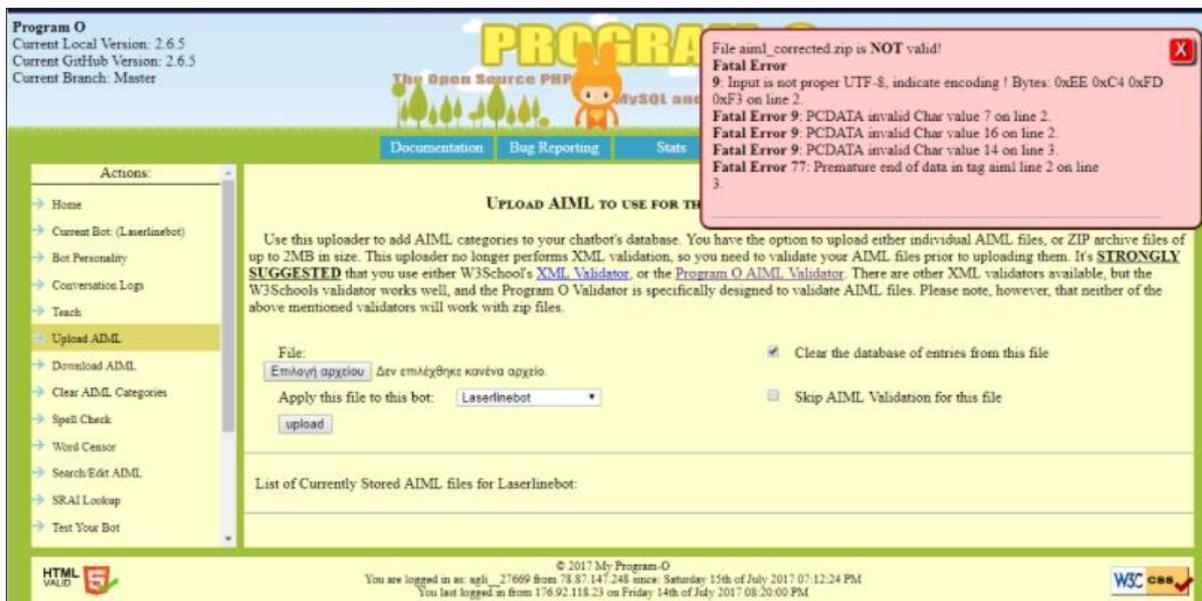
Nas aulas 1, 2 e 3 (datas e conteúdo revelados na tabela 3.2), os estudantes tiveram a oportunidade de ter conhecimentos sobre HTML e sobre as diversas *tags* que contemplam a informática, além de produzir algum conteúdo e ter um *site* desenvolvido.

A aula 4 foi dedicada especificamente para apresentar os agentes conversacionais: ATENA, METIS e ALTEIA. A atividade proposta nesta aula teve como objetivo propiciar a interação dos discentes com cada um dos agentes, para que pudessem descobrir algumas possibilidades e termos para conversação, como, por exemplo, inserir diálogos com duas ou três palavras, e não uma frase com cinco ou mais vocábulos. Durante a conversação, eles percorriam as habilidades das agentes, realizavam suas avaliações e observavam como as respostas estavam sendo retomadas pelos *chatbots*.

Ainda na aula 4, o professor e, também, autor da presente tese trouxe a estrutura básica dos códigos em AIML, com o intuito de relacionar algumas *tags* de AIML com o HTML. Na ocasião, foram desenvolvidos exercícios e trechos de conversação.

Na aula 5, introduziu-se o ambiente *Program-O* disponibilizado para as conversações especificamente deste estudo. A figura 5.3 apresenta a tela e o *layout* utilizado no experimento com os alunos.

Figura 5.3 – Ambiente que os alunos tinham acesso para gerenciar as classes



Fonte: <https://github.com/Program-O/Program-O/issues/306>.

Nesta aula, os participantes da pesquisa fizeram *uploads* e *download* dos arquivos, utilizando a base de conhecimentos já construída. A proposta da aula 5 foi mostrar a eles que era possível construir perguntas e adicionar respostas, o que aumentaria a base de conhecimento do agente conversacional sobre um determinado assunto. A figura 5.4 demonstra parte do código na linguagem AIML desenvolvida por um aluno participante da pesquisa. Ele faz a pergunta: “O que é a tag ‘body’?” e, em seguida, deixa a resposta abaixo, na linha 79.

Figura 5.4 – Trecho de código elaborado por aluno

```

77
78 <pattern> O que é a tag "body"? </pattern>
79 <template> Tudo o que vemos em um site, estará dentro da tag body. Como o próprio nome sugere, a tag body define o corpo de um site. É
   nela que <br/>
80 estarão as informações visíveis para o usuário da internet. Tudo que você está lendo nesse site, os links, imagens e vídeos que estão
   vendo, <br/>
81 estão dentro da tag "body".
82 </template>
83 </category>
84
85 <category>
86
87 <pattern> O que você sabe sobre o Senac-RS? </pattern>
88 <template>
89     DESCOBRIR MAIS SOBRE O SENAC-RS: <br/>
90
91     <a href="https://www.senacrs.com.br/institucional_missao.asp" target="_blank"> - Missão, Visão e Princípios; </a> <br/>
92     <a href="https://www.senacrs.com.br/sustentabilidade.asp" target="_blank"> - Senac Sustentabilidade; </a> <br/>
93     <a href="https://www.senacrs.com.br/pdf/codigo_etica.pdf" target="_blank"> - Código de ética; </a> <br/>
94     <a href="https://www.senacrs.com.br/institucional_premiacoes.asp" target="_blank"> - Premiações; </a> <br/>
95     <a href="https://www.senacrs.com.br/" target="_blank"> - Geral do Site. </a> <br/>
96 </template>
97
98 </category>
99
100 </aiml>

```

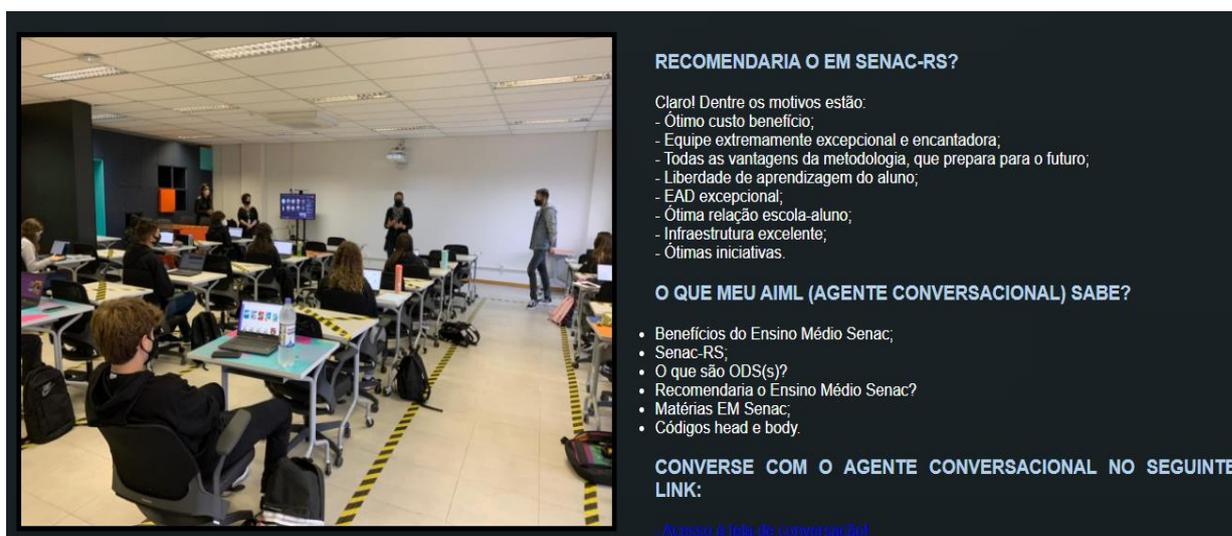
Fonte: Código desenvolvido por aluno(a) participante da pesquisa.

Outro detalhe na figura 5.4 trata da pergunta disponível na linha 87 da respectiva figura: “O que você sabe sobre o Senac-RS?”. A partir da linha 89, verificam-se diversos *hyperlinks* na resposta do agente conversacional, o que demonstra que o aluno conseguiu mesclar os conhecimentos da linhagem HTML e relacionar na linguagem AIML. Percebe-se, ainda, que somente neste arquivo, o qual pode ser chamado de classe, há 100 linhas de códigos misturando o aprendizado do curso com os conhecimentos de programação na linguagem AIML.

A partir da aula 6, focou-se na elaboração de *tags* em AIML e inserção do *link* de conversação do *chatbot* nos *sites* desenvolvidos pelos alunos.

A próxima tela mostra a continuação do mesmo site apresentado na figura 5.4, em que se apresenta parte do site que contém o espaço reservado para o uso e conversação do agente conversacional, conforme figura 5.5. Esta revela outra disposição de *layout* criado por outro(a) aluno(a). Neste exemplo, há informações do *site* e um espaço para conversação e diálogos.

Figura 5.5 – Exemplo de inserção do agente conversacional no *site* desenvolvido

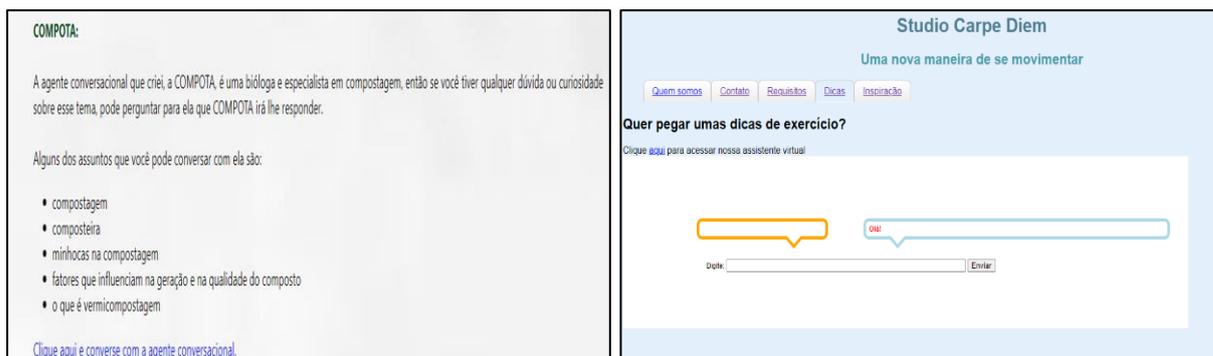


Fonte: *Site* desenvolvido por aluno(a) participante da pesquisa.

Ainda conforme a figura 5.5, observa-se que o(a) aluno(a) se preocupou em informar os assuntos que o agente conversacional tem conhecimento e foi disponibilizado um *link* para acesso e conversação logo abaixo. Esse é o resultado do estudo de 30 horas, envolvendo conhecimentos de AIML, HTML e agentes conversacionais.

Até aqui, foram evidenciados alguns exemplos de sites, mas houve produção de *sites* por 77 estudantes ao longo das 30 horas de desenvolvimento. A figura 5.6 mostra outros exemplos de *layouts* de *sites* desenvolvidos por outros alunos participantes da pesquisa.

Figura 5.6 – Compilação de diferentes sites com o agente conversacional integrado



Fonte: Site desenvolvido por alunos(as) participantes da pesquisa.

Por fim, na aula 6, abordou-se a hospedagem dos *sites* construídos pelos alunos no *GitHub*. O *GitHub* é uma plataforma gratuita que permite hospedar o código fonte dos *sites* desenvolvidos, permitindo o acesso de qualquer lugar do mundo, desde que se tenha acesso à Internet. Durante esta aula, cada aluno criou sua conta gratuita no *GitHub* e praticou fazer o *upload* dos arquivos que compunham os *sites* criados. Além disso, os estudantes testaram a funcionalidade do *site* e corrigiram os erros até o correto funcionamento de todos os *sites*.

A figura 5.7 apresenta um exemplo de *site* desenvolvido pelos participantes da pesquisa e hospedado na plataforma *GitHub*.

Figura 5.7 – Exemplo de *site* hospedado no *GitHub* pelo aluno



Fonte: Site desenvolvido por alunos(as) participantes da pesquisa.

Assim, na etapa 2.1 da presente pesquisa de tese: a) procurou-se introduzir conhecimentos da linguagem HTML e o conceito de agentes conversacionais; b) possibilitou-se conversas e diálogos com os agentes conversacionais METIS, ATENA e ALTEIA; c)

explorou-se o *Program-O*; e d) utilizou-se o *GitHub* para hospedar virtualmente os arquivos que compõem o *site*, o que possibilitou abrir o *site* criado pelos alunos a partir de qualquer navegador de qualquer lugar, desde que se tenha acesso à internet.

Os *sites* desenvolvidos tiveram o objetivo tornar possível a prática dos conteúdos e *tags* aprendidas. Neste primeiro exemplo, cada estudante teve a opção de criar um *site* sobre um assunto de sua escolha. No *site* desenvolvido por eles, cada um criou um espaço com informações de agentes conversacionais, forneceu informações sobre o conhecimento desses agentes e, a partir disso, apresentou-se o *link* de conversação da agente de teste.

As práticas de programação com os alunos proporcionaram conhecimentos técnicos e oportunidade de experiência de conversação com as agentes. Assim, acredita-se que os participantes da pesquisa estariam mais preparados para a avaliação preliminar dos agentes conversacionais, usando o questionário baseado em métricas trabalhadas nesta tese. Paralelamente, realizou-se o teste do *ChatAval*, envolvendo todas as suas funcionalidades.

Nesta etapa, o objetivo foi de mesclar os conhecimentos de HTML, a linguagem de marcação AIML, proporcionar uma experiência de agentes, de construção de códigos em AIML e de disponibilizar os arquivos do *site* no *GitHub*.

A prática de conversação realizada pelos alunos junto aos agentes conversacionais METIS, ATENA e ALTEIA ofereceu o primeiro contato dos alunos com estes agentes e permitiu que compreendessem algumas limitações de conversação, como, por exemplo, digitar frases longas e o agente retornar que não tem conhecimento sobre aquele assunto e informar os assuntos sobre os quais tem conhecimento. Além disso, os alunos conseguiram perceber que, além de entregar a resposta à pergunta, é possível adicionar *links* para outros *sites* ou *links* de vídeos, de modo a complementar as informações solicitadas. Por conseguinte, essa prática de conversação dos agentes conversacionais foi muito importante para que os estudantes compreendessem as possibilidades e limitações dos diálogos.

A prática com os agentes avaliados envolveu, ainda, a demonstração de como é importante informar ao usuário alguns assuntos que os agentes conversacionais estão programados para responder, pois entende-se que existe a possibilidade de os *chatbots* não entregarem todas as respostas do seu domínio de conhecimento.

5.2.2 Resultados Fase 2.2 – Experiências dos alunos com os agentes conversacionais

A fase 2.2 envolveu um segundo momento de conversação entre os estudantes do Curso Técnico em Informática com os agentes conversacionais por, aproximadamente, uma hora para

cada agente conversacional avaliado neste estudo. Cada aluno participante da pesquisa conversou individualmente com cada uma das três agentes conversacionais utilizadas no estudo. A organização de datas, horários e carga horária reservados para cada turma e agente conversacional avaliada ocorreu conforme o previsto e apresentado na Tabela 3.3.

Esta fase envolveu a aplicação de um questionário baseado nas métricas da NBR-ISO IEC 9126-1. Tal questionário (Apêndice C) contribuiu para o estudo e testes pilotos para modelagem do *framework* proposto para avaliação dos agentes conversacionais. Embora a NBR-ISO IEC 9126-1 tenha sido substituída pela NBR-ISO IEC 25010, optou-se por seguir essa normativa, porque as características utilizadas estão contempladas em ambas as normativas. A estrutura do escopo desenvolvido nesta fase tem o objetivo de servir como base para o protótipo para o *framework*.

A tabela 5.1 apresenta o percurso para construção do instrumento de avaliação que foi aplicado aos alunos na fase 2.2. A primeira coluna apresenta a característica e, logo abaixo da primeira coluna, são listadas as subcaracterísticas. Cada uma delas é considerada, neste estudo, como uma métrica examinada no instrumento de avaliação dos alunos.

A segunda coluna da tabela 5.1, tem o objetivo de registrar a descrição das métricas, de acordo com o ISO/IEC 9126-1. A partir da descrição das métricas, fez-se a descrição da terceira coluna “Pergunta elaborada pelo autor, com base na métrica e descrição da mesma”, ou seja, a apresentação de cada métrica e de seu respectivo conceito, elaborou-se uma pergunta que está no questionário do aluno, para que pudesse avaliar os agentes conversacionais. Cabe ressaltar que os alunos não têm conhecimento sobre o percurso e sobre o que motivou a elaboração da pergunta.

Tabela 5.1 – Percurso para a construção e avaliação dos agentes

Característica e subcaracterística	Descrição da métrica: Referência ISO/IEC 9126-1	Pergunta elaborada pelo autor com base na métrica e descrição da mesma
Característica 1. USABILIDADE		
1.1 Atratividade	Capacidade do produto de <i>software</i> de ser atraente ao usuário.	Enquanto você utiliza o agente conversacional, você o considera atrativo?
1.2 Operacionalidade	Capacidade do produto de <i>software</i> de possibilitar ao usuário operá-lo e controlá-lo.	O agente conversacional possibilita que você consiga operá-lo intuitivamente?
1.3 Apreensibilidade	Capacidade do produto de <i>software</i> de possibilitar ao usuário aprender sua aplicação.	O agente conversacional apresenta características de fácil compreensão e pode ser aprendido e operado facilmente?

Característica 2. FUNCIONALIDADE		
2.1 Acurácia	Capacidade do produto de <i>software</i> de prover, com o grau de precisão necessário, resultados ou efeitos corretos ou conforme acordados.	O agente conversacional retorna as respostas de acordo com os diálogos, de forma correta, e de acordo com o conteúdo digitado?
2.2 Interoperabilidade	Capacidade do produto de <i>software</i> de interagir com um ou mais sistemas especificados.	O agente conversacional continua funcionando enquanto algumas alterações na base de conhecimento estão sendo implementadas?
Característica 3. CONFIABILIDADE		
3.1 Recuperabilidade	Capacidade do produto de <i>software</i> de restabelecer seu nível de desempenho especificado e recuperar os dados diretamente afetados no caso de uma falha.	Caso o servidor ou computador apresente problemas de travamento, os dados permanecem quando o serviço é restabelecido?
Característica 4. EFICIÊNCIA		
4.1 Comportamento em relação ao tempo	Capacidade do produto de <i>software</i> de fornecer tempos de resposta e de processamento, além de taxas de transferência, apropriados, quando o <i>software</i> executa suas funções, sob condições estabelecidas.	O agente conversacional apresenta um bom acesso em relação ao tempo para trazer a resposta durante os diálogos?
4.2 Conformidade relacionada à eficiência	Capacidade do produto de <i>software</i> de estar de acordo com normas e convenções relacionadas à eficiência.	O agente conversacional possui eficiência para abrir, fechar e carregar as telas de diálogos e de respostas, tendo em vista a configuração do seu computador e a velocidade de sua internet?
Característica 5. MANUTENIBILIDADE		
5.1 Analisabilidade	Capacidade do produto de <i>software</i> de permitir o diagnóstico de deficiências, ou causas de falhas no <i>software</i> , ou a identificação de partes a ser modificadas.	O agente conversacional informa diagnóstico de falhas para o usuário?
5.2 Testabilidade	Capacidade do produto de <i>software</i> de permitir que o <i>software</i> , quando modificado, seja validado.	É possível fazer testes e validações no agente conversacional antes do uso da versão final ou na versão final?

Fonte: Elaborado pelo autor com base na ISO/IEC 9126.

Observa-se que não foram utilizadas todas as características e subcaracterísticas na tabela, visto que, segundo a normativa ISO / IEC 9126-1, não é obrigatório utilizar as métricas distribuídas em todas as características e subcaracterísticas. Isso possibilita selecionar aquelas que contemplem o objetivo da avaliação do *software* desejado. Apenas na característica na norma ISO/IEC 9126 são 6 características e 25 subcaracterísticas.

Neste estudo piloto de avaliação das métricas, examinou-se, também, a funcionalidade do protótipo *ChatAval*, que foi apresentado na Seção 4 da presente tese. Esta etapa abrangeu a avaliação, cadastro das perguntas elaboradas, configuração de pesos das características e subcaracterísticas. A figura 5.8 apresenta parte das configurações do questionário.

Figura 5.8 – Configuração do cadastro das perguntas no *ChatAval*

Alterar Formulário

Pesquisa Avaliação Tese - Alunos Senac

Estimados alunos do Senac,

Chegou o momento de fazermos um teste como avaliadores dos agentes conversacionais testados. Trata-se de uma avaliação com base em métricas de avaliação de software.

A partir de sua avaliação, será possível fazer sugestões de melhorias para versões futuras do agente conversacional avaliado.

Escolha as questões para o formulário

Obs: A soma dos pesos das categorias utilizadas precisa ser 10. Assim como a soma dos pesos das sub categorias selecionadas para uma mesma categoria também precisa ser 10.

Funcionalidade

2

<input type="checkbox"/>	Adequação	O agente conversacional possibilita funções apropriadas para o usuário, tal como botões, caixas de diálogos?	
<input checked="" type="checkbox"/>	Acurácia	O agente conversacional retorna respostas de acordo com os diálogos de forma correta de acordo com o conteúdo digitado?	5
<input checked="" type="checkbox"/>	Interoperabilidade	O agente conversacional continua funcionando enquanto algumas alterações na base de conhecimento estão sendo implementadas?	5
<input type="checkbox"/>	Segurança de acesso	Para cadastrar novas tags na base de conhecimento do seu agente conversacional, você entende que há segurança para acessar e cadastrar novas bases de conhecimentos?	
<input type="checkbox"/>	Conformidade relacionada à funcionalidade	O agente conversacional apresenta informações se é um software livre, proprietário ou informações para seu uso?	

Fonte: Elaborado pelo autor (2021).

Conforme a figura 5.8, cadastrou-se o formulário de perguntas com informações relacionadas ao título da pesquisa, um texto de apresentação e as métricas de características e subcaracterísticas. Esta figura apresenta apenas um recorte ou parte das perguntas. Consta-se que o peso atribuído à característica Funcionalidade foi 2,0; já as Subcaracterísticas Acurácia e Interoperabilidade estão com peso 5,0. A pergunta das respectivas subcaracterísticas também foi ajustada para avaliar especificamente os agentes conversacionais abordados no presente estudo. Ainda conforme a figura 5.8, observa-se que as subcaracterísticas Adequação, Segurança de Acesso e Conformidade relacionadas à funcionalidade não foram utilizadas nesta avaliação. Poder contar com a escolha das categorias e das subcategorias e atribuir peso a elas é uma funcionalidade desenvolvida no *ChatAval* para tratar especificamente das métricas de avaliação para os agentes conversacionais.

Após o cadastro e configurações de pesos das características e subcaracterísticas, um *link* de pesquisa é gerado automaticamente, para que os alunos possam responder ao questionário de avaliação. A figura 5.9, a seguir, apresenta o questionário gerado automaticamente após as configurações de avaliação no *ChatAval*.

Figura 5.9 – Questionário gerado automaticamente pelo *ChatAval*



Pesquisa Avaliação Tese - Alunos Senac

Referente à aplicação METIS
Estimados alunos do Senac,

Chegou o momento de fazermos um teste como avaliadores dos agentes conversacionais testados. Trata-se de uma avaliação com base em métricas de avaliação de software.

A partir de sua avaliação, será possível fazer sugestões de melhorias para versões futuras do agente conversacional avaliado.

As métricas de avaliação estão nas questões as quais você deve optar por uma pontuação entre 1 e 5, sendo:

- 1) Discordo totalmente;
- 2) Discordo;
- 3) Indiferente;
- 4) Concordo;
- 5) Concordo Plenamente.

Desde já agradecemos por sua avaliação.
Equipe de Pesquisa

É possível fazer testes e validações no agente conversacional antes do uso da versão final ou versão final?

1 2 3 4 5

O agente conversacional possui eficiência para abrir, fechar e carregar as telas de diálogos e de respostas, tendo em vista a configuração do seu computador e velocidade de sua internet?

1 2 3 4 5

Fonte: Elaborado pelo autor (2021).

Esta figura ilustra apenas duas questões de avaliação, no entanto salienta-se que as demais questões seguem abaixo até o término dos dados. Uma vez que todos os alunos realizaram suas avaliações no questionário, os resultados também são apresentados automaticamente, como se pode consultar na figura 5.10.

Figura 5.10 - Avaliação das métricas para a avaliação do *software*

Resultados do formulário Pesquisa Avaliação Tese - Alunos Senac - final				
Categoria	Peso	NOTAS		S
		ALTEIA	ATENADOIS	
Confiabilidade	1.0			
Recuperabilidade	10.0	6.7	6.4	6.4
	SUBTOTAL	-	0.7	0.6
Eficiência	3.0			
Comportamento em relação ao tempo	4.0	3.6	3.2	3.6
Conformidade relacionada à eficiência	6.0	4.7	4.7	4.7
	SUBTOTAL	-	2.5	2.5
Funcionalidade	2.0			
Acurácia	5.0	4.2	3.5	3.9
Interoperabilidade	5.0	3.2	3.3	3.5
	SUBTOTAL	-	1.5	1.5
Manutenibilidade	1.0			
Analisabilidade	5.0	1.9	2.1	2.1
Testabilidade	5.0	4.0	3.9	4.0
	SUBTOTAL	-	0.6	0.6
Usabilidade	3.0			
Apreensibilidade	2.0	1.7	1.7	1.7
Atratividade	4.0	1.8	2.1	2.2
Operacionalidade	4.0	3.0	2.8	2.9
	SUBTOTAL	-	2.0	2.0
	TOTAL	-	7.2	7.2

Fonte: Elaborado pelo autor (2022). Adaptado de acordo com a NBR-ISO IEC 9126-1.

A figura 5.10 apresenta o resultado calculado automaticamente pelo *ChatAval*. Apresentam-se as categorias e suas respectivas subcategorias com os pesos atribuídos para cada um dos agentes conversacionais ALTEIA, ATENA e METIS avaliados no estudo.

Na categoria Confiabilidade, o peso da avaliação foi de 1,0, considerando uma escala de 10,0, e utilizou-se apenas a subcategoria Recuperabilidade, que ficou com peso de 10,0 por ser a única subcategoria atribuída nesta avaliação. O resultado da avaliação dos alunos registrou para a categoria Recuperabilidade, que teve como pergunta “*Caso o servidor ou computador apresente problemas de travamento, os dados permanecem quando o serviço é restabelecido?*”, o peso de 6,7 para a ALTEIA e 6,4 para os agentes ATENA e METIS. No subtotal, constata-se que a ALTEIA ficou com 0,7 e ATENA e METIS ficaram com 0,60, considerando que a nota para essa categoria é de no máximo 1,0.

Já a categoria Eficiência teve o peso 3,0. As subcategorias Comportamento em relação ao tempo e Conformidade relacionada à eficiência tiveram seus pesos, respectivamente, de 4,0 e 6,0. No subtotal, a ALTEIA e METIS ficaram com 2,5 e a ATENA com 2,4 tendo em vista o peso máximo de 3,0.

Em relação à categoria Funcionalidade, configurou-se o peso de 2,0. As subcategorias utilizadas foram Acurácia e Interoperabilidade, ambas com peso de 5,0. Observa-se que, no subtotal, a ALTEIA e METIS ficaram com uma avaliação de 1,5 ponto, enquanto a ATENA ficou com 1,4, considerando a nota máxima de 2,0 da categoria.

A categoria de Manutenibilidade teve apenas 1,0 ponto no total da funcionalidade. Utilizaram-se as subcategorias Analisabilidade e Testabilidade, ambas com 5,0 pontos cada. No subtotal, as três agentes conversacionais avaliadas no estudo ficaram com 0,6 pontos.

Por fim, a categoria Usabilidade teve o peso de 3,0 pontos. As subcategorias utilizadas foram Apreensibilidade com 2,0, Atratividade com 4,0 e Operacionalidade com 4,0 pontos. Acredita-se que os melhores resultados nas avaliações dos agentes se deram porque uma agente conversacional apresentou uma resposta mais elaborada, trazendo um vídeo de apoio, por exemplo, ou ainda porque houve uma resposta e funcionalidade mais assertiva para a pergunta realizada pelos alunos em relação aos demais agentes. A subcaracterística Apreensibilidade questionou sobre as características de fácil compreensão e se pode ser aprendido e operado facilmente e todas os agentes conversacionais tiveram excelente avaliação. Isso demonstra que, além dos desafios de implementação e alimentação da base de conhecimento para as dúvidas elaboradas é fundamental apresentar respostas claras, objetivas, bem elaboradas e com materiais e leituras de apoio.

Novamente os três agentes conversacionais avaliados no estudo ficaram com um total de 2,0 pontos, considerando um total de 3,0 pontos. O resultado dos agentes conversacionais foi de 7,2 para os agentes ALTEIA e METIS; e 6,9 para o ATENA, em uma escala de 10,0, considerando as métricas avaliadas e apresentadas.

Vale destacar que, entre os três agentes conversacionais, o ATENA recebeu bem menos investimento em termos de enriquecimento da base de conhecimento, no entanto foi avaliado de forma próxima dos demais agentes. O resultado obtido a partir do ATENA demonstra a importância da avaliação a partir do *framework* proposto na presente tese, pois avaliou-se não apenas uma métrica, como a base de conhecimento, por exemplo, mas um conjunto de métricas escolhidas pelo autor da tese: Atratividade, Operacionalidade, Apreensibilidade. Além disso, acredita-se que esse resultado se dá também em razão de os três agentes estarem envolvidos em um mesmo contexto tecnológico e terem suas funcionalidades e *layouts* muito parecidos.

Sem dúvida, caso outras métricas (categorias e subcategorias) de avaliação de *software* estivessem sido escolhidas para a avaliação, o resultado, possivelmente, seria diferente, considerando tanto uma média superior como uma média inferior, quando comparadas ao resultado.

Acredita-se que, embora os agentes conversacionais tenham diferentes conhecimentos para a conversação, as funcionalidades, percepções e conhecimentos apresentam características muito semelhantes, o que fez as notas de avaliação atribuídas ficarem próximas por poucos décimos de diferenças entre cada um dos agentes conversacionais avaliados.

Constata-se que, com exceção da métrica Atratividade do agente conversacional ALTEIA, que ficou com 1,8, considerando o peso 4,0; e a métrica Analisabilidade em que o ALTEIA obteve 1,9, enquanto o ATENA e o METIS contabilizaram 2,1, considerando o peso de 5,0, as demais avaliações ficaram com notas superiores à metade da nota atribuída, tendo como parâmetro de avaliação as métricas de avaliação de *software* sob a visão e experiência de uso dos alunos do Curso Técnico em Informática.

Vale salientar que o escopo apresentado na figura 5.10, além de avaliar a percepção dos alunos em relação ao uso dos agentes conversacionais explorados neste estudo, teve como finalidade testar e mapear as métricas como um estudo piloto para a construção do *framework* sistematizado e apresentado na fase 3.

Além disso, os resultados apresentados na figura 5.10 também procuram demonstrar as funcionalidades de cadastro de perguntas, com base nas subcategorias, atribuição de notas para as características e subcaracterísticas, apresentação do formulário com as questões elaboradas

para a avaliação dos agentes conversacionais e a apresentação dos resultados, tendo em vista o ponto de vista dos alunos do curso Técnico em Informática para Internet.

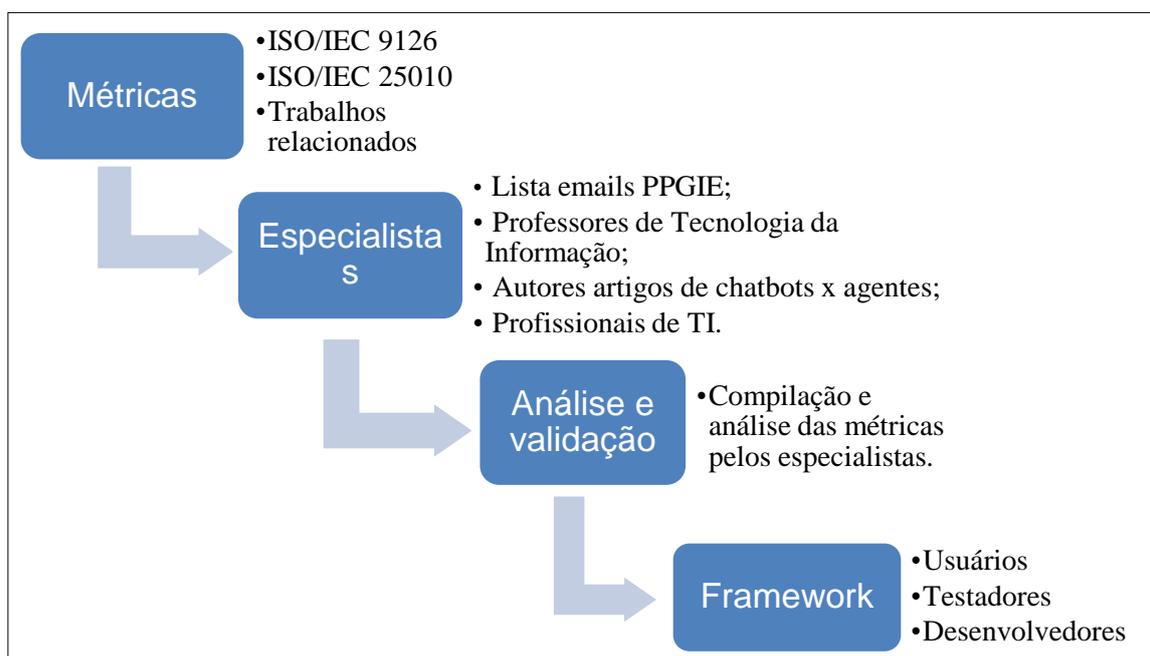
A opinião dos estudantes foi importante, pois tiveram a oportunidade de conhecer conceitos de agentes conversacionais, construir *sites* em HTML, construir base de conhecimentos em AIML e, ainda, criar um espaço no *site* para falar sobre seu *chatbot*. Eles também gerenciaram os arquivos de conversação no *Program-O*. Antes de realizar a avaliação, o professor e pesquisador da presente tese também comentou as questões, buscando esclarecer dúvidas sobre o que avaliar. Assim, após tal experiência dos alunos com os códigos em AIML e HTML, entende-se que eles teriam condições de fazer avaliação sobre os agentes para testar as funcionalidades do *ChatAval*.

Por fim, nesta fase, procurou-se construir um pré-teste para as métricas de avaliação de agentes conversacionais, tendo como base inicial a ISO/IEC 9126-1, a fim de elaborar um instrumento para a construção do *framework* com as métricas de avaliação de agentes conversacionais. Nesse sentido, os alunos foram envolvidos na avaliação preliminar de agentes conversacionais, usando um questionário baseado em métricas trabalhadas na tese. Neste estudo, elaborou-se o percurso de cada característica, subcaracterística, o conceito da subcaracterística e como essa característica pode ser reavaliada em forma de pergunta.

5.3 RESULTADOS DA FASE 3 – PARTICIPAÇÃO DOS ESPECIALISTAS

A fase 3 contou com a participação de profissionais que atuam como desenvolvedores, professores de Tecnologia da Informação, pesquisadores, autores de artigos sobre agentes conversacionais ou *chatbots*, os quais, neste estudo, são denominados especialistas. O objetivo desta fase foi validar as métricas desenvolvidas a partir do olhar dos especialistas para a composição das métricas do *framework*.

A figura 5.11 apresenta o percurso para a construção das métricas, o qual envolve a compilação das métricas com origem na ISO/IEC 9126, ISO/IEC 25010 e trabalhos relacionados; o grupo de especialistas que validou as métricas; a análise e validação das métricas; e, por fim, a composição final do *framework*, que pode ser utilizado por usuários diversos, os quais classificam-se, neste estudo, como usuários, testadores e desenvolvedores.

Figura 5.11 – Percurso para a composição das métricas do *framework*

Fonte: Elaborado pelo autor (2021).

Para uma melhor organização da apresentação do percurso para a composição do *framework* (Fase 3), as próximas subseções detalham essas fases, obedecendo à ordem demonstrada na figura 5.11.

5.3.1 Apresentação das métricas analisadas pelos especialistas

O primeiro momento para a construção do *framework* foi o desenvolvimento de um instrumento contendo as métricas da ISO/IEC 9126-1, ISO/IEC 25010:2011 e algumas métricas propostas nos trabalhos relacionados, as quais foram organizadas em categorias e subcategorias (métricas), de acordo com suas especificidades, conforme já apresentado na seção 4 da presente tese, e podem ser consultadas na tabela 4.2, no entanto reapresentam-se as métricas de maneira resumida na tabela 5.2.

Tabela 5.2 – Compilação das métricas utilizadas e analisadas pelos especialistas

Categorias	Métricas
Funcionalidade	Funcionalidade completa, Qualidade das respostas, Respostas de <i>fallback</i> , Métricas de mensagens
Eficiência	Comportamento em relação ao tempo, Eficácia
Compatibilidade	Interoperabilidade
Usabilidade	Reconhecimento de adequação, Aprendizagem, Operabilidade, Proteção contra erros de usuário, Acessibilidade

Confiabilidade	Maturidade, Disponibilidade, Tolerância ao erro
Segurança	Confidencialidade, Integridade, Autenticidade
Manutenção	Reutilização, Testabilidade
Portabilidade	Adaptabilidade
Satisfação	Utilidade / Motivação para utilização, Confiança, Prazer/satisfação, Afetividade, Humanidade, Ética e Comportamento
Cobertura de contexto	Completude do contexto, manter a conversa
Controle de usuários	Usuários ativos, total de usuários

Fonte: Elaborado pelo autor (2022).

Todas as métricas elencadas foram distribuídas por categorias para serem validadas. Como o objetivo é a validação de um agente conversacional para cada uma das métricas, foram elaboradas questões para avaliar um determinado agente conversacional. Uma vez construídas as categorias, métricas e exemplos de questões para cada uma das métricas, enviou-se um instrumento, criado pelo autor da tese, para os especialistas validarem as métricas propostas. Este instrumento pode ser conferido no Apêndice H denominado Instrumento Modelo de Avaliação das Métricas por Especialistas. Ele apresenta a contextualização do estudo, as diretrizes de avaliação, as categorias, as subcategorias, o texto exemplo para avaliação das subcategorias, as opções de aceitar ou rejeitar as métricas e uma opção para os especialistas reescreverem o item com o ajuste sugerido.

5.3.2 Apresentação do perfil dos especialistas

O processo de avaliação e validação das métricas por especialistas (Apêndice H) iniciou com o pedido de participação na pesquisa, mediante envio de e-mails à lista do PPGIE e de disparo de mensagens via *WhatsApp*, para autores de artigos relacionados a *chatbots* ou agentes conversacionais. Os especialistas retornaram o e-mail com o documento preenchido (Apêndice H) junto ao Termo de Consentimento Livre Esclarecido (Apêndice I).

Por conta de não haver conhecimento do número de especialistas que retornariam as mensagens com as análises preenchidas, optou-se por mapear posteriormente o perfil dos participantes. A figura 5.12 apresenta a atuação dos especialistas.

Figura 5.12 – Grupo de especialistas que validaram o *framework*



Fonte: Elaborado pelo autor, com base nas informações do TCLE enviado dos especialistas.

Conforme a figura 5.12, constata-se que houve um público de avaliadores que estão inseridos nos campos da Educação, Computação e Informática na Educação. Isso demonstra que o *framework* foi validado por diversos profissionais de diferentes campos de atuação profissional. Os avaliadores da área de Computação foram classificados, neste estudo, como os profissionais que atuam com Tecnologia da Informação e não têm vivências ou experiências na docência. Os avaliadores que estão na categoria de Educação são professores de diversas áreas da educação, mas que usam as tecnologias no contexto de suas disciplinas. Já os especialistas classificados como Informática na Educação são os profissionais formados nas áreas de Computação, Sistemas de Informação, Licenciatura em Computação e outras formações, os quais atuam como docentes na área de Tecnologia da Informação.

A formação dos respectivos especialistas que validaram o *framework* compõe-se de 15% de especialistas (formação *Lato sensu*), 55% mestres e 30% doutores.

5.3.3 Resultados da análise e validação das métricas

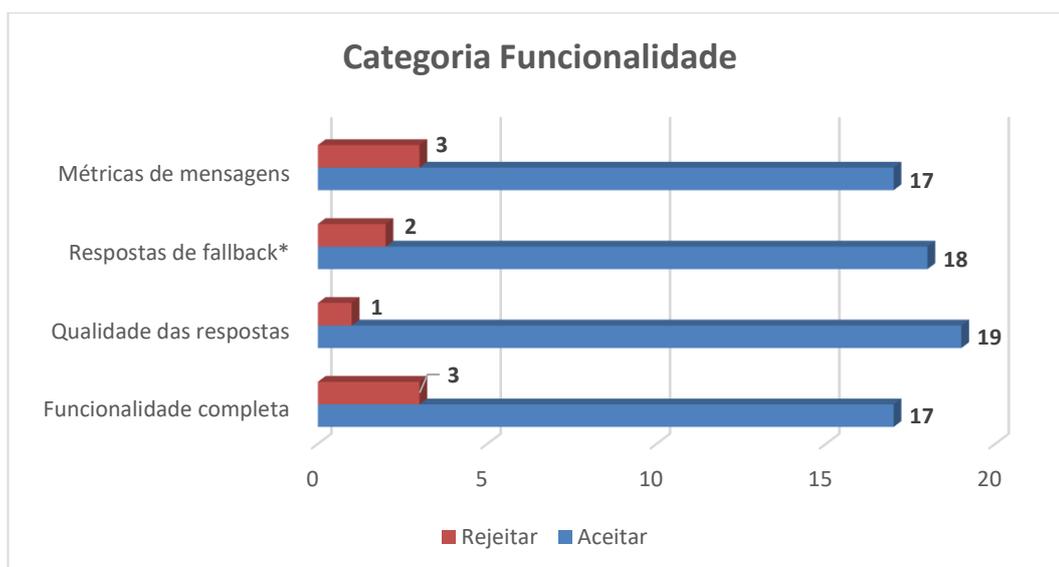
Esta subseção apresenta o resultado da avaliação das métricas realizadas pelos especialistas (Apêndice H). Os resultados serão apresentados por categorias, considerando que são 31 métricas distribuídas em 11 categorias. No instrumento de avaliação por parte dos especialistas, cada categoria tem uma ou mais métricas de avaliação. Cada métrica tem uma afirmação relacionada à métrica específica.

As diretrizes envolveram se a questão é relevante para a avaliação de *chatbots*, e, caso o especialista entendesse como não relevante, poderia rejeitá-la, havendo a opção para poderem escrever uma justificativa no campo "Ajustes ao item". No respectivo campo, os avaliadores tiveram como critérios analisar se a questão condiz com a categoria/métrica indicada. Em caso negativo, solicitou-se que o especialista sugerisse qual a categoria seria mais apropriada, ou que fizesse a sugestão para o ajuste do texto para a avaliação da respectiva métrica.

Para organizar a apresentação dos resultados alcançados a partir das respostas dos especialistas, as considerações de cada métrica são apresentadas de acordo com as categorias nas foram distribuídas.

A primeira categoria a ser apresentada é a **Funcionalidade** com suas respectivas métricas. A Figura 5.13 apresenta o resultado de cada uma das métricas.

Figura 5.13 – Resultados métricas categoria Funcionalidade



Fonte: Elaborado pelo autor (2021).

De acordo com a figura 5.13, constata-se que a categoria Funcionalidade obteve a maioria dos resultados de “Aceitar as métricas”. Já a questão do *framework*, de acordo com a avaliação realizada pelos especialistas, três deles rejeitaram a métrica Funcionalidade Completa e Métricas de Mensagens. A seguir, apresenta-se a questão e sugestão dos especialistas para as métricas que foram Aceitas ou Rejeitadas (tabela 5.3).

Tabela 5.3 – Transcrição das avaliações dos especialistas - categoria Funcionalidade

Métrica: Funcionalidade completa	
ID Especialista	Transcrição das respostas dos Especialistas
ID Esp_02	“Aqui é uma linha tênue, poderia aceitar, porém, assim como o humano precisa estar em constante aprendizado, <i>chatbot</i> é um dos temas da inteligência artificial e precisa, a cada novo assunto, estar aprendendo. Então, falar funcionalidade completa eu discordo, mesmo sabendo que ele pode ser aprendido futuramente.”
ID Esp_03	“Questão muito genérica. O que seria funcionalidade completa?”
ID Esp_06	“Entendo que a descrição do item deveria substituir o trecho ‘a sua funcionalidade completa’ por ‘a funcionalidade completa descrita pelo <i>chatbot</i> em análise.”
ID Esp_07	“Seria bom descrever o que é uma funcionalidade completa.”
ID Esp_10	“Cuidar que a ‘funcionalidade’ pode ser conflituosa sem um escopo bem definido.”
Métrica: Qualidade das respostas	
ID Esp_10	“Cuidar o contexto/realidade do usuário pra ser ‘adequada’, <i>chatbots</i> falando sozinho criarem um idioma e era adequado pra eles, mas ilegível pra humanos.”
ID Esp_13	“As opções pré-definidas, muitas vezes, não atendem a expectativa do usuário, encerrando a conversa.”
Métrica: Respostas de fallback	
ID Esp_05	“O <i>chatbot</i> é capaz de identificar a quantidade de vezes que falhou ou passou por uma situação de falha.”
ID Esp_13	“O perfil de <i>chatbot</i> utilizado por mim, como bancários, serviços públicos, não retornam <i>fallback</i> . Seria importante um canal para demonstrar as expectativas do cliente.”
ID Esp_18	“Esta questão é deveras importante para o programador e desenvolvedor, porém, para o usuário, não sei se teria muita relevância se não estiver apresentada em um contexto em que o <i>chatbot</i> justifica sua falha em responder à pergunta realizada.”
Métrica: Métrica de mensagens	
ID Esp_05	“O <i>chatbot</i> apresenta mensagens iniciais/finais de conversação ou saudações, a fim de iniciar ou encerrar um diálogo.”
ID Esp_15	“O <i>chatbot</i> disponibiliza apresentar informações sobre tempo de atendimento.”
ID Esp_19	“Eu acredito que precisa complementar. O <i>chatbot</i> precisa ir além disso, suportar conversa fiada, encerrar uma conversa.”

Legenda: Identificador do participante – ID.

Fonte: Elaborado pelo autor (2021).

Em relação à métrica **Funcionalidade completa**, contata-se que, conforme a tabela 5.3, nenhum dos especialistas orientaram para tirar ou excluir a respectiva métrica. O *Esp_02* discute a métrica de funcionalidade completa devido ao *chatbot* precisar estar sempre aprendendo, mas o objetivo da respectiva métrica é validar a funcionalidade no momento da interação. Já o *Esp_03* menciona como questão genérica, mas não invalida a métrica, o que indica que pode ser ajustado o texto para trabalhar no TC. O texto para a respectiva métrica foi substituído conforme a indicação do *Esp_06*, que sugere alterar o texto para “a funcionalidade completa descrita pelo *chatbot* em análise”.

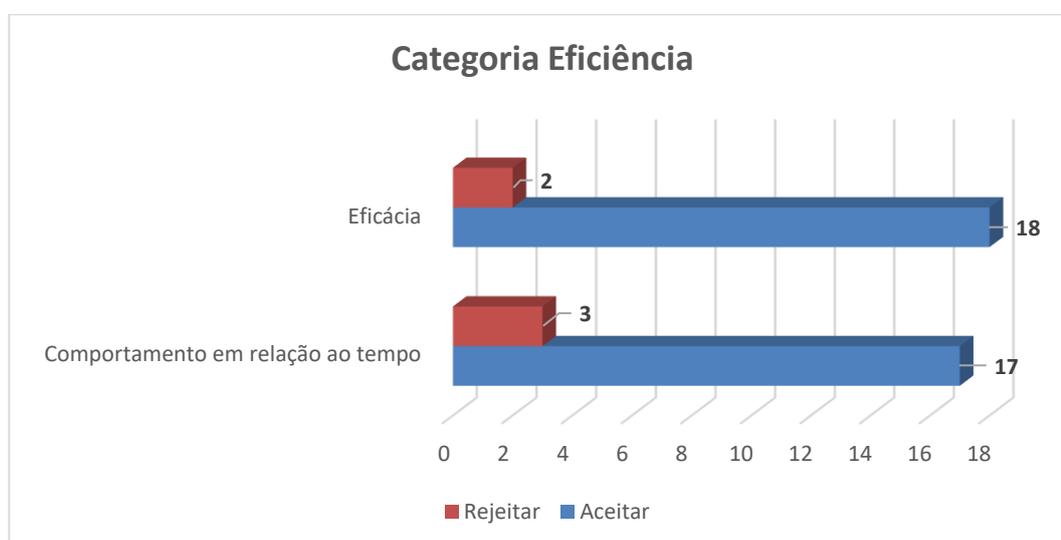
Já em relação à métrica **Qualidade das respostas**, os especialistas não orientaram excluí-la do *framework*, mas atentar para as opções pré-definidas, conforme orienta o *Esp_13*. Assim, essa métrica será mantida.

A respeito de **Respostas de fallback**, assim como as métricas discutidas até aqui, não foi sugerida sua exclusão da relação de métricas do *framework*. O *Esp_05* sugere que o texto de avaliação seja alterado para “O *chatbot* é capaz de identificar a quantidade de vezes que falhou ou passou por uma situação de falha.”, e tal sugestão será acolhida e irá compor o texto de validação da métrica no *framework*. A métrica é importante para programadores e desenvolvedores (*Esp_18*), o que demonstra que o *framework* proposto pode atender a diferentes níveis de usuários, conforme já abordado anteriormente neste estudo.

Por fim, na Categoria Funcionalidade, na **Métricas de mensagens**, o especialista *Esp_05* sugere que o texto seja alterado para “O *chatbot* apresenta mensagens iniciais/finais de conversação ou saudações, a fim de iniciar ou encerrar um diálogo.”, e este texto substituirá o anterior. Novamente, não aparece, em nenhum momento, que essa métrica não é válida, assim será mantida no *framework*.

A segunda Categoria a ser apresentada é a **Eficiência** e seus resultados podem ser conferidos na figura 5.14.

Figura 5.14 - Resultados métricas categoria Eficiência



Fonte: Elaborado pelo autor (2021).

A Categoria Eficiência teve suas métricas aceitas pela grande maioria dos especialistas. Destaca-se que a métrica que está contabilizada no gráfico como “Rejeitar” não significa que essa não deverá manter-se no *framework*, mas que há ajustes e considerações a fazer em relação ao texto original. Apresentam-se as constatações dos especialistas na tabela 5.4.

Tabela 5.4 - Transcrição das avaliações dos especialistas - categoria Eficiência

Especialista	Transcrição das respostas dos Especialistas
Métrica: Comportamento em relação ao tempo	
ID Esp_05	“O <i>chatbot</i> apresenta um tempo adequado de respostas.”
ID Esp_16	“O <i>chatbot</i> apresenta um tempo adequado de respostas. Rápido é relativo, pois depende do usuário. Acredito que poderia haver alguma pergunta do tipo: O <i>chatbot</i> permite algum ajuste no tempo de resposta? (assim o usuário pode adequar o tempo de resposta ao seu perfil).”
ID Esp_17	“O <i>chatbot</i> apresenta rapidamente as respostas. ‘um tempo rápido’ me pareceu pouco adequado.”
Métrica: Eficácia	
ID Esp_10	“Somar uma métrica do retorno com sugestões de temas quando não entender algo para além de uma mensagem que não entendeu”
ID Esp_11	“O <i>chatbot</i> contempla com precisão e integridade o objetivo do diálogo.”
ID Esp_19	“Esse item não está sobrepondo o item ‘qualidade das respostas?’ Comento isso assumindo que, se ele falha ao apresentar respostas adequadas, ele não estará atendendo com precisão o objetivo do diálogo. Não sei o que poderia alterar.”

Legenda: Identificador do participante – ID.

Fonte: Elaborado pelo autor (2021).

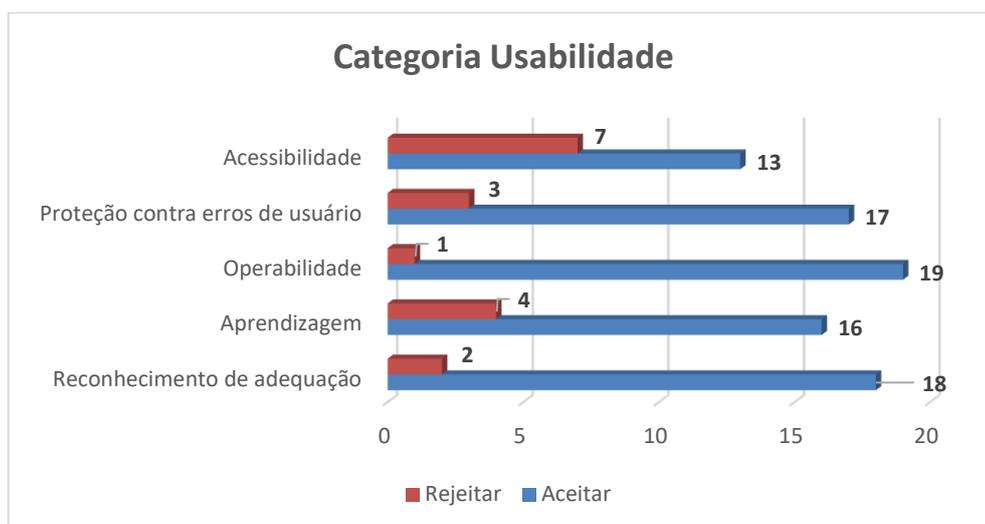
De acordo com a tabela 5.4, a métrica **Comportamento em relação ao tempo** continuará a compor o *framework*, sendo que a descrição do texto do *Esp_05* será acolhida e passará a ser: “O *chatbot* apresenta um tempo adequado de respostas”. Os demais especialistas (*Esp_16* e *Esp_17*) que apontaram sugestões serão igualmente contemplados pelo ajuste.

Já a métrica **Eficácia** passará a ter o texto sugerido pelo *Esp_11*: “O *chatbot* contempla com precisão e integridade o objetivo do diálogo”. Os demais comentários, como, por exemplo, do *Esp_19*, que tem dúvidas sobre o item “qualidade das respostas” serão esclarecidos quando ele estiver com o *framework* em mãos com os conceitos de cada métrica relacionados às questões a serem avaliadas. Assim, neste momento, a indicação do respectivo especialista não será acolhida pelo autor da tese.

A próxima categoria é a **Compatibilidade**. Essa categoria teve uma aceitação de 100% dos especialistas. Dessa forma, não há sugestões a serem demonstradas na forma de gráficos, assim como acontece nas categorias anteriores.

A Categoria seguinte que apresenta resultados é a **Usabilidade**. Suas métricas e dados contabilizados para a opção “Aceitar e Rejeitar” são apresentados na figura 5.15.

Figura 5.15 - Resultados métricas categoria Usabilidade



Fonte: Elaborado pelo autor (2021).

Esta categoria obteve grande aceitação de suas respectivas métricas a partir da avaliação dos especialistas. Para um maior detalhamento, apresentam-se, na tabela 5.5, as considerações dos especialistas.

Tabela 5.5 - Transcrição das avaliações dos especialistas - categoria Usabilidade

Especialista	Transcrição das respostas dos Especialistas
Métrica: Reconhecimento de adequação	
ID Esp_16	“Necessidades de quem? Não entendi. Talvez: É possível reconhecer que o <i>chatbot</i> é uma solução adequada ao(s) problema(s) que se quer resolver.”
ID Esp_17	“O <i>chatbot</i> é apropriado para suas necessidades.”
Métrica: Aprendizagem	
ID Esp_05	“O <i>chatbot</i> oferece mecanismos para que o usuário possa aprender a usá-lo em situações de emergência ou situação de uso imediato.”
ID Esp_14	“(1) as demais categorias parecem se referir ao <i>bot</i> , mas esta parece se referir a ‘aprendizagem’ do usuário. Sugiro alguma outra categoria como ‘facilidade’. (2) Acredito que esta questão poderia abordar a capacidade do <i>bot</i> de compreender exclamações emergenciais do usuário, como: ajuda! socorro! de forma simples, direta e descontextualizada, com a finalidade de atender a requisitos de ‘emergência e situações de uso imediato’, ao invés de abordar o quão rápido o usuário aprenderá a usar o <i>bot</i> . Isto é, tirar a carga do usuário aprender o <i>bot</i> , mas, sim, o <i>bot</i> aprender o que o usuário quer dizer.”
ID Esp_16	“O <i>chatbot</i> possui uma interface amigável, com linguagem clara e navegação intuitiva, o que permite seu uso imediato.”
ID Esp_19	“Aprendizagem de uso (cuidado com o termo aprendizagem).”
Métrica: Operabilidade	
ID Esp_16	“O <i>chatbot</i> é fácil de usar.”
Métrica: Proteção contra erros de usuário	
ID Esp_09	“Erro de resposta do usuário ou agente conversacional?”
ID Esp_10	“Ambíguo. O ‘erro na resposta’, por parte do <i>chatbot</i> ou do usuário? Talvez, 1 métrica desta pra cada.”
ID Esp_13	“Em alguns dados, entra em <i>looping</i> .”

ID Esp_19	“Esse item está na categoria correta? Eu acho que ele deveria estar no item funcionalidade.”
Métrica: Acessibilidade	
ID Esp_05	“O <i>chatbot</i> pode ser utilizado por pessoas com diferentes características físicas e mentais.”
ID Esp_07	“Melhor seria descrever quais são essas características, usar exemplos.”
ID Esp_09	“Ser mais específico. Quais necessidades especiais, por exemplo, um cego? Síndrome de Down?”
ID Esp_10	“Considerar ao uso de categoria acessibilidade, ao invés de ‘características físicas’, usar ‘deficiências físicas e cognitivas’, mas aí tu abres um escopo muito grande. Fecha em 1, ao exemplo, o ‘ <i>chatbot</i> pode ser utilizado por pessoas com deficiência visual’, e outra ‘... com deficiência motora’, e outra ‘deficiência auditiva.’”
ID Esp_11	“O <i>chatbot</i> pode ser utilizado por pessoas com diferentes características físicas e intelectuais?”
ID Esp_13	“Não ficou claro. Aqui se refere a deficientes? Se sim, acredito que não atendem a totalidade dos casos.”
ID Esp_16	“Nem toda deficiência é física, pode ser intelectual. Sugestão: O <i>chatbot</i> pode ser utilizado por pessoas com diferentes tipos de deficiência. E também é importante lembrar que ele pode ser acessível para um tipo de deficiência e não para outro. Por exemplo, pode ser acessível a pessoas cegas, mas não a pessoas surdas (caso não tenha tradução para Libras e/ou seja integrado a algum avatar intérprete de Libras).”
ID Esp_19	“Esse item precisa ser opcional na minha opinião. Eu entendo que se o <i>chatbot</i> ser acessível, ele conseguirá atingir a usabilidade universal (se não estou enganado, é o termo adotado pela turma de IHC). Mas nem todo <i>chatbot</i> precisa ser acessível ou precisa? Acho que não há problema com o item, apenas sugiro que seja um item opcional.”

Legenda: Identificador do participante – ID.

Fonte: Elaborado pelo autor (2021).

Em relação à métrica **Reconhecimento de adequação**, verifica-se que não foi invalidada, no entanto sua descrição para a avaliação será a partir da sugestão do Esp_17: “O *chatbot* é apropriado para suas necessidades.”

A métrica **Aprendizagem** teve 4 rejeições pelos especialistas. Constata-se que não houve rejeições em relação a excluir a métrica da categoria, apenas sugestão de alteração do texto para a avaliação. Para esta métrica, será acolhida a sugestão do Esp_5, e será realizada, portanto, a alteração no texto original, que passará a ser: “O *chatbot* oferece mecanismos para que o usuário possa aprender a usá-lo em situações de emergência ou situação de uso imediato”.

Já a métrica **Operabilidade** teve apenas uma rejeição que neste caso não foi de invalidar a métrica, mas de alterar o texto. O texto será ajustado de acordo com sugestão do especialista Esp_16: “O *chatbot* é fácil de usar”.

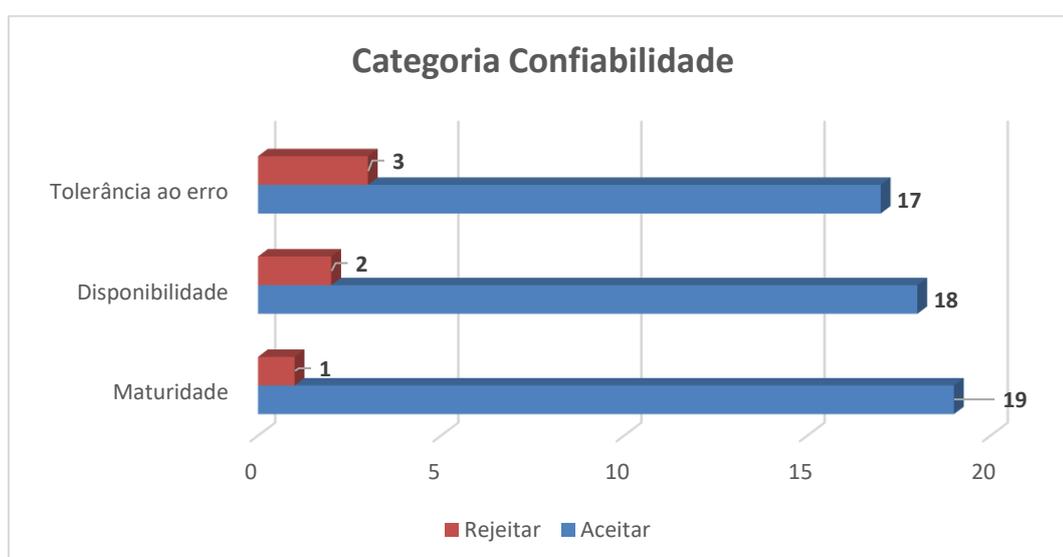
A métrica **Proteção contra erros de usuário** contabilizou 3 rejeições dos especialistas. Salienta-se aqui que as rejeições não foram para invalidar a métrica, mas para realizar pequenas alterações nos textos para a avaliação. Destaca-se a mesma dúvida dos Esp_09 e Esp_10, por questionarem o termo “erro”, que aparece na questão de pesquisa, gerando dúvidas se o erro está relacionado ao *chatbot* ou ao usuário. Esse texto será ajustado

no *framework*. O especialista Esp_19 sugere que a métrica em questão seja realocada para a categoria **Funcionalidade**, no entanto entende-se que a métrica Proteção contra erros de usuários está relacionada ao erro de usuários do que a erros de funcionamento técnico; assim a a métrica seguirá na categoria Usabilidade.

Para concluir as métricas da categoria Usabilidade, a de **Acessibilidade** foi a que obteve o total de 7 rejeições por parte dos especialistas, considerando-se todas as métricas propostas para compor o *framework*. Vale destacar que, embora tenham ocorrido rejeições para esta métrica, nenhuma foi no sentido de excluí-la do *framework*, mas de realizar ajustes ou questionamentos em relação ao público da acessibilidade sobre um tipo de deficiência específico, como, por exemplo, deficiência visual. Salienta-se que essa métrica tem o objetivo de sinalizar se o *chatbot* pode ser utilizado por pessoas com algum tipo de limitação física ou intelectual e tal texto pode ser ajustado para cada público em específico, como pessoas com deficiência auditiva, visual, física, entre tantas outras possibilidades que são tratadas pela acessibilidade. O Esp_11 sugeriu alterar o texto da métrica para “O *chatbot* pode ser utilizado por pessoas com diferentes características físicas e intelectuais?”. Após a análise dos demais avaliadores, este texto será acolhido para compor o *framework*, pois acredita-se que ele contempla grande parte dos questionamentos realizados.

A próxima categoria apresentada é a **Confiabilidade**, a qual pode ser conferida na figura 5.16.

Figura 5.16 - Resultados métricas categoria Confiabilidade



Fonte: Elaborado pelo autor (2021).

Conforme a figura 5.16, a categoria Confiabilidade teve uma nota muito boa por parte dos especialistas. A tabela 5.6 apresenta a transcrição dos especialistas em relação às métricas da categoria Confiabilidade:

Tabela 5.6 - Transcrição das avaliações dos especialistas - categoria Confiabilidade

Especialista	Transcrição das respostas dos Especialistas
Métrica: Tolerância ao erro	
ID Esp_10	“Cuidar ou especificar melhor qual <i>software</i> ou ‘algum <i>software</i> não necessário pro <i>chatbot</i> ’, pois, por exemplo, se falhar navegador e o <i>chatbot</i> funciona em navegador, já era.”
ID Esp_13	“Em alguns casos, é necessário fazer a atualização primeiro.”
ID Esp_19	“Isso aqui não está muito claro. Está assumindo que o <i>chatbot</i> faça parte de um sistema de <i>software</i> ? Se for isso, deve ser um item opcional. Como trabalho com teste de <i>software</i> , acredito que o termo correto seria ‘tolerância à falha’.”
Métrica: Disponibilidade	
ID Esp_11	“Quando você precisou, o <i>chatbot</i> estava disponível? Se não estava, as informações necessárias estavam?”
Métrica: Maturidade	
ID Esp_19	“O problema aqui é ‘o que quer dizer com confiabilidade em operação’. Seria apresentar informações confiáveis? Se for isso, não sei se a categoria deveria se chamar ‘maturidade’. Isso me preocupa.”

Legenda: Identificador do participante – ID.

Fonte: Elaborado pelo autor (2021).

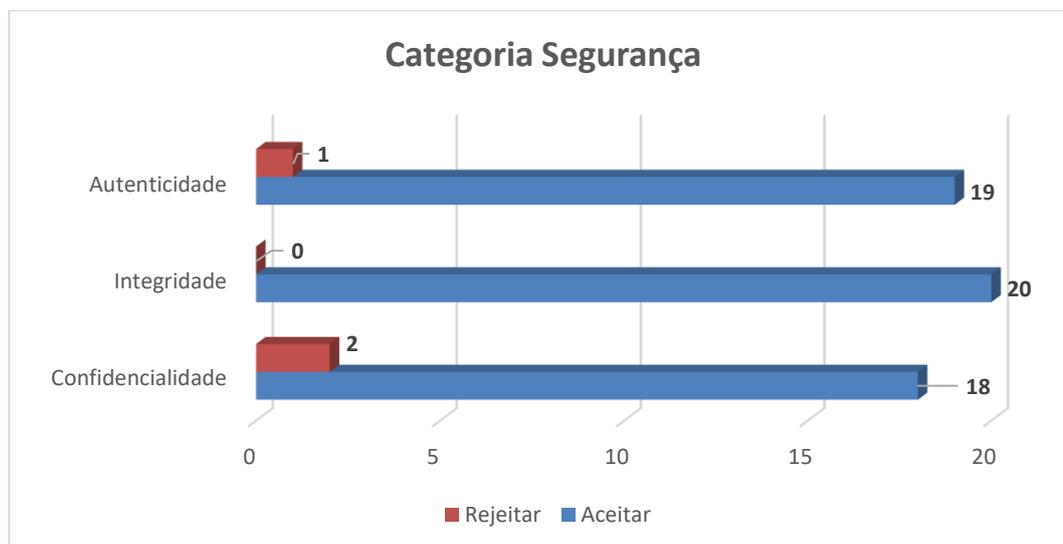
Na métrica **Tolerância ao erro**, houve 3 rejeições, mas muito relacionadas ao texto da métrica, ou sugestões de ordem de atualização de *softwares*. O Esp_19 declara-se como testador de *software* e sugeriu alterar o nome de métrica para “Tolerância à Falha”. Tal sugestão será acatada e a métrica passará a ter esta nova nomenclatura.

A métrica **Disponibilidade** teve uma única sugestão do Esp_11 para ajustar o texto para “Quando você precisou, o *chatbot* estava disponível? Se não estava, as informações necessárias estavam?”. No entanto, observa-se que há duas questões na sugestão do respectivo especialista, assim, analisando o texto original enviado para o especialista e a sugestão recebida, a métrica passará a ter o texto de avaliação “Quando você precisou, o *chatbot* estava disponível”?

A métrica **Maturidade** teve apenas uma rejeição, no entanto discutiu-se mais o nome da métrica e não sobre o que deve ser avaliado. Entende-se que não houve uma boa sugestão e argumentação por parte do especialista. Assim, o nome da métrica e o seu respectivo texto serão, portanto, mantidos de acordo com a proposta inicial do autor da tese.

A próxima categoria a ser avaliada é a Segurança e os resultados podem ser conferidos na figura 5.17.

Figura 5.17 - Resultados métricas categoria Segurança



Fonte: Elaborado pelo autor (2021).

Assim como as demais, a Categoria Segurança não teve nenhuma métrica rejeitada. Destaca-se a métrica **Integridade** aceita por todos os especialistas. A tabela 5.7 registra as avaliações dos especialistas da categoria referida.

Tabela 5.6 - Transcrição das avaliações dos especialistas - categoria Segurança

Especialista	Transcrição das respostas dos Especialistas
Métrica: Confidencialidade	
ID Esp_05	“O <i>chatbot</i> garante a confidencialidade de dados do usuário por algum tipo <i>login</i> com senha ou, ainda, redireciona para um link específico de acesso.”
ID Esp_18	“Acredito que esta questão caiba melhor em acesso ou acessibilidade.”
ID Esp_19	“Esse item me preocupa. Cuidado para não criar algo que irá engessar o <i>chatbot</i> . Talvez, o mais correto para esse item seja que o <i>chatbot</i> garanta que os diálogos produzidos durante a conversação sejam armazenados em local seguro e que eles não serão usados para fins indevidos (pensar em algo na linha da LGPD).”
Métrica: Integridade	
ID Esp_18	“Outras possibilidades devem ser consideradas, como perguntas sem nexos (sentido) que possam vir a confundir e presumir a funcionalidade de resposta do <i>chatbot</i> .”
Métrica: Autenticidade	
ID Esp_06	“Acredito que a melhor subcategoria seja Disponibilidade, pois a preocupação é a legitimidade dos dados ou diálogos que serão disponibilizados.”

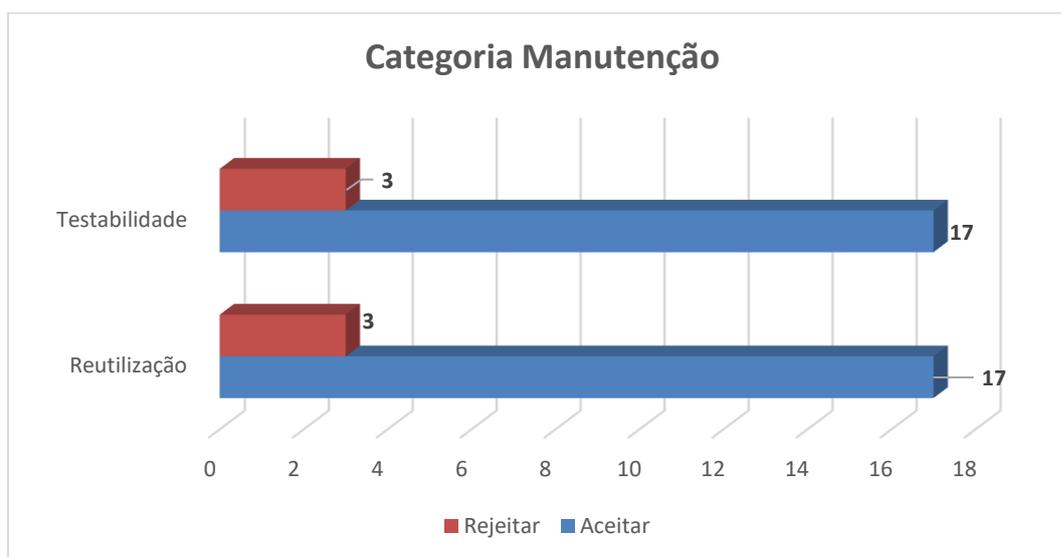
Fonte: Elaborado pelo autor (2021).

A métrica **Confidencialidade** foi rejeitada por dois especialistas, porém, assim como aconteceu com as demais, eles não recomendaram excluí-la, mas fazer ajustes no texto. O Esp_18 comenta que a questão ficaria melhor em acesso ou acessibilidade, no entanto as métricas têm funcionalidades e objetivos diferentes e, por essa razão, não será considerada a sugestão deste especialista. Já o Esp_05 sugere alterar o texto da métrica para “O *chatbot* garante a confidencialidade de dados do usuário por algum tipo *login* com senha ou, ainda, redireciona para um *link* específico de acesso”. Esta sugestão será acolhida.

As métricas **Integridade** e **Autenticidade** também não foram rejeitadas pelos especialistas, mas houve um pequeno retorno deles que não justifica a alteração da métrica e de seu texto.

A próxima categoria é a **Manutenção** e seus resultados são apresentados na figura 5.18.

Figura 5.18 - Resultados métricas categoria Manutenção



Fonte: Elaborado pelo autor (2021).

A Categoria **Manutenção** teve apenas 3 rejeições para ambas as métricas. Os detalhes e transcrição dos especialistas podem ser conferidos na tabela 5.8.

Tabela 5.7 – Transcrição das avaliações dos especialistas - categoria Manutenção

Especialista	Transcrição das respostas dos Especialistas
Métrica: Reutilização	
ID Esp_05	“Informações da base de conhecimentos poderão ser reaproveitadas para a construção/expansão de outras bases.”
ID Esp_14	“Retroalimentar o modelo para ser um item de funcionalidade.”
ID Esp_19	“Acho que vai além das bases de conhecimento. O ideal é que pudéssemos reusar todos os artefatos e esforços usados durante o estabelecimento do <i>chatbot</i> (e.g.,

	máquina de inferência, modelo treinado que representa algum comportamento/expertise etc.).”
Métrica: Testabilidade	
ID Esp_20	“Testabilidade poderia ser um item de Confiabilidade. Manutenção parece-me estar mais relacionado a capacidade de diagnóstico (<i>troubleshooting</i>) ou rotinas de limpeza (<i>logging, tracing, events</i>).”
ID Esp_17	“Não tenho uma sugestão direta, mas este item ficou confuso, deveria ser repensado.”
ID Esp_19	“Confuso, não consegui entender.”

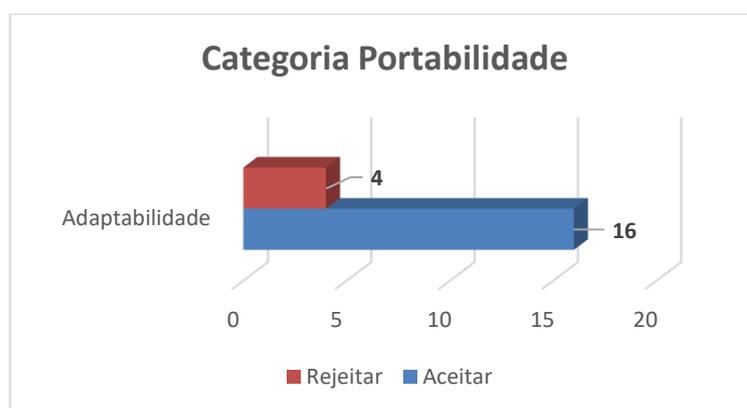
Legenda Identificador do participante – ID.
Fonte: Elaborado pelo autor (2021).

No que se refere à métrica **Testabilidade**, o texto provocou dúvidas nos Esp_17 e Esp_19 sobre a clareza de seu objetivo. Embora confusos, não conseguiram oferecer uma contribuição. Já o Especialista Esp_20, sugere que esse item seja realocado para a categoria de Confiabilidade. Assim, a métrica **Testabilidade** vai fazer parte da Categoria Confiabilidade, conforme sugestão do Esp_20. Os demais especialistas que rejeitaram a respectiva métrica não a invalidaram, como é possível constatar na tabela 5.8.

Já a categoria **Reutilização** também teve 3 rejeições. Tais rejeições não foram para excluir a métrica, mas para refletir sobre possibilidades de reutilização, como sugere o Esp_19: “...reusar todos os artefatos e esforços usados durante o estabelecimento...”. O Esp_14 comenta que parece ser um item da Categoria Funcionalidade. De fato, a métrica Reutilização não está relacionada com a função do produto *software* (agente conversacional) como um todo. Assim, após a análise dos especialistas, a métrica **Reutilização** será realocada para a Categoria **Compatibilidade**, visto que tem maior relação com as métricas já disponíveis na respectiva categoria. Assim, a partir da realocação da métrica Reutilização e Testabilidade para a Categoria Compatibilidade, a categoria Manutenção será excluída do *framework*.

A próxima categoria a ser apresentada é a **Portabilidade** (figura 5.19).

Figura 5.19 – Resultados métrica categoria Portabilidade



Fonte: Elaborado pelo autor (2021).

Conforme a figura 5.19, a categoria Portabilidade obteve 4 rejeições, segundo a avaliação dos especialistas. Tais rejeições são para melhorias no texto. Os especialistas não apontam a exclusão da respectiva métrica. A tabela 5.9 apresenta a transcrição das avaliações dos especialistas da Categoria Portabilidade.

Tabela 5.8 – Transcrição das avaliações dos especialistas - categoria Portabilidade

Especialista	Transcrição das respostas dos Especialistas
Métrica: Adaptabilidade	
ID Esp_05	“O <i>chatbot</i> funciona em diferentes sistemas operacionais e em diferentes configurações de <i>hardware</i> ou de <i>software</i> .”
ID Esp_07	“Exemplificar os S.O.”
ID Esp_16	“Não seria importante verificar se o <i>chatbot</i> funciona em diferentes dispositivos? Ou diferentes configurações de <i>hardware</i> são dispositivos?”
ID Esp_20	“Acredito que deva acrescentar - <i>cross browser</i> .”

Legenda: Identificador do participante - ID.

Fonte: Elaborado pelo autor (2021).

De acordo com a tabela 5.9, o especialista Esp_05 sugere que o texto para avaliação da métrica seja reajustado para “O *chatbot* funciona em diferentes sistemas operacionais e em diferentes configurações de *hardware* ou de *software*.”; assim, o texto ficou mais adequado do que o texto proposto. O Especialista Esp_07 sugere exemplificar o S.O., no entanto a ideia do *framework* é basicamente deixar uma métrica como sugestão e apontar que o profissional que avaliará um determinado agente conversacional educacional tenha a liberdade de especificar o sistema operacional que será avaliado.

Devido a não constar sugestão para realocação da respectiva métrica por parte dos especialistas para outra categoria, a métrica Adaptabilidade permanecerá na Categoria Portabilidade, assim como sugere o modelo ISO/IEC 25010.

A próxima categoria a ser apresentada é a **Satisfação**. Esta categoria é a que contém mais métricas no *framework*. A figura 5.20 apresenta os resultados da respectiva métrica. Referida.

Figura 5.20 – Resultados métricas categoria Satisfação



Fonte: Elaborado pelo autor (2021).

Destaca-se a métrica Ética e Comportamento com total aprovação por parte dos especialistas. As transcrições das demais avaliações são apresentadas na tabela 5.10.

Tabela 5.9 – Transcrição das avaliações dos especialistas - categoria Satisfação

Especialista	Transcrição das respostas dos Especialistas
Métrica: Utilidade / Motivação para utilização	
ID Esp_05	“Após usar o <i>chatbot</i> , sinto-me satisfeito em relação ao uso baseado em meus objetivos, tais como os resultados do uso.”
ID Esp_13	“Falta detalhes aos pré-requisitos dos questionamentos iniciais. Muitas vezes, a opção indicada não encaminha para a pergunta real e, portanto, não temos as respostas, voltamos aos canais tradicionais, como: telefone ou e-mail.”
ID Esp_16	“Achei a frase um pouco confusa. Não entendi o que quer avaliar.”
ID Esp_17	“em relação ao uso de meus objetivos’ - eu não entendi bem o que seria ao uso de meus objetivos. Talvez poderia ser reescrito. ‘tais como, os resultados do uso’ - também poderia ser reescrito.”
ID Esp_19	“Esse item tem a mesma função que o item reconhecimento de adequação.”
Métrica: Confiança	
ID Esp_05	“O <i>chatbot</i> é confiável ao fornecer os serviços/interações ao qual foi proposto.”
ID Esp_07	“Não sei se o termo ‘confiar’ é o mais correto”
ID Esp_16	“Achei confuso. O que seria ‘o pretendido?’”
ID Esp_19	“Esse item, a meu ver, tem a mesma função do item maturidade.”
Métrica: Prazer/satisfação	
ID Esp_19	“Esse item é igual ao item ‘Reconhecimento de adequação’. Acho que não precisa.”
Métrica: Afetividade	
ID Esp_5	“O <i>chatbot</i> apresenta diálogos com saudações, cordialidade e é capaz de responder, com respeito e cordialidade, as interações afetivas dos participantes.”
ID Esp_19	“Veja, isso toca o item métricas de mensagem. Será que precisa do métricas de mensagem? Acho melhores essas subdivisões. Mas aí tem outra coisa. Quem garante que não está falando algum item?”
Métrica: Humanidade	
ID Esp_20	“Acrescentar a possibilidade de informação ao usuário que o mesmo está conversando com uma IA desde o início da conversa.”

Legenda: Identificador do participante - ID.

Fonte: Elaborado pelo autor (2021).

De todas as métricas da Categoria Satisfação, a métrica **Utilidade / Motivação para utilização** foi a métrica que apresentou o maior número de rejeições. As cinco rejeições dos especialistas se referem mais ao texto. O Esp_05 deixou uma sugestão que deve contemplar as sugestões e comentários dos demais especialistas. Assim, o texto será ajustado para “Após usar o *chatbot*, sinto-me satisfeito em relação ao uso baseado em meus objetivos, tais como, os resultados do uso”.

A métrica **Confiança** foi aceita por todos, embora tenham sugestões de alterações no texto de validação, sendo esse o motivo de 3 rejeições. Novamente, o *framework* irá adequar a sugestão do Esp_05 alterando o texto de validação da métrica para “O *chatbot* é confiável ao fornecer os serviços/interações aos quais foi proposto”.

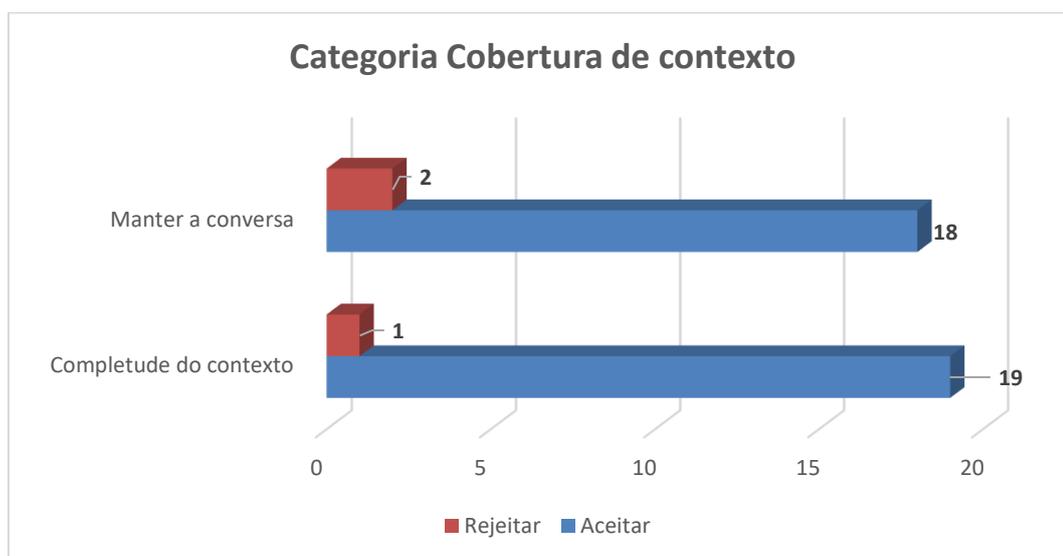
Já a métrica **Prazer/satisfação** apresentou uma rejeição. Trata-se do Esp_19, que sugere que a métrica seja descartada. Considerando que houve apenas uma rejeição por um especialista e os demais não argumentaram, optou-se por seguir com a métrica “prazer/satisfação” no *framework* proposto.

Dando continuidade às métricas da categoria Satisfação, a **Afetividade** obteve duas rejeições. Destaca-se aqui a contribuição do Esp_5 para a alteração do texto de validação da métrica para “O *chatbot* apresenta diálogos com saudações, cordialidade e é capaz de responder, com respeito e cordialidade, as interações afetivas dos participantes”.

A métrica **Humanidade** teve a sugestão do Esp_20 de “Acrescentar a possibilidade de informação ao usuário que o mesmo está conversando com o uma IA desde o início da conversa”. A sugestão deste especialista traz a discussão de que se pode informar ao usuário que está interagindo com uma máquina, pois, em muitos atendimentos, é possível ter a impressão de se estar falando com um ser humano, enquanto se está, de fato, conversando com um agente conversacional. Por outro lado, o objetivo desta métrica é avaliar a interação do agente conversacional de maneira satisfatória e natural; assim, este estudo não vai considerar a sugestão do especialista.

A penúltima Categoria do *framework* é a **Cobertura de Contexto**. Os resultados das métricas podem ser conferidos na figura 5.21.

Figura 5.21 – Resultados métricas categoria Cobertura de Contexto



Fonte: Elaborado pelo autor (2021).

A tabela 5.11 apresenta a transcrição das avaliações dos especialistas em relação à Categoria Cobertura de Contextos.

Tabela 5.10 – Transcrição das avaliações dos especialistas - categoria Cobertura de Contexto

Especialista	Transcrição das respostas dos Especialistas
Métrica: Completude de contexto	
ID Esp_16	“O que seria a completude de contexto?”
Métrica: Manter a Conversa	
ID Esp_05	“O <i>chatbot</i> é capaz de manter o fluxo de diálogo.”
ID Esp_17	“Em qualquer situação? Manter o diálogo até o fim de seu uso? Achei que faltou algo na frase.”

Legenda: Identificador do participante - ID.

Fonte: Elaborado pelo autor (2022).

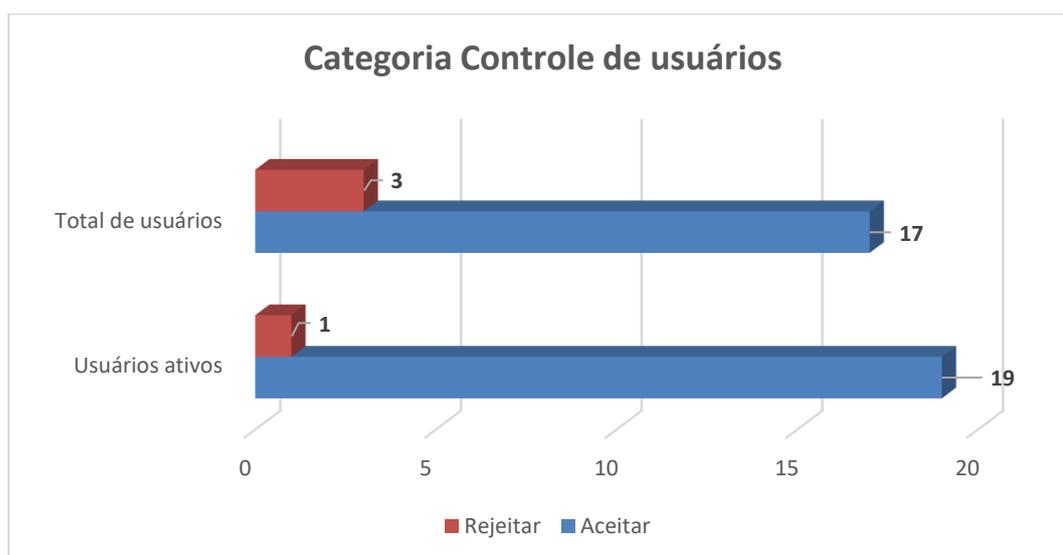
No que se refere à métrica **Completude de contexto**, o Especialista Esp_16 questionou o nome da métrica “O que seria a completude de contexto?”, no entanto foi solicitado aos especialistas que rejeitassem e deixassem suas contribuições. A dúvida do Esp_16 poderá ser encontrada quando os especialistas utilizarem o *ChatAval*, que vai conter a descrição da respectiva métrica. Como foi uma única rejeição e os dezenove especialistas aceitaram a respectiva métrica, essa será utilizada e mantida no *framework*.

Já em relação à métrica **Manter a Conversa**, o Esp_17 argumenta: “...manter o diálogo até o fim de seu uso?...”. Já o especialista Esp_05 sugere ajustar o texto de validação da métrica para “O *chatbot* é capaz de manter o fluxo de diálogo”. Assim, acredita-se que

alterando o texto de validação da métrica para a sugestão do Esp_05, a dúvida do Esp_17 será respondida. Dessa forma, este estudo vai acatar a sugestão de texto do Esp_05.

A última categoria do *framework* é o **Controle de usuários**. A figura 5.22 apresenta o resultado das duas métricas desta categoria.

Figura 5.22 – Resultados métricas categoria Controle de Usuários



Legenda: Identificador do participante - ID.
Fonte: Elaborado pelo autor (2021).

A tabela 5.12 registra a transcrição das avaliações dos especialistas.

Tabela 5.11 – Transcrição das avaliações dos especialistas - categoria Controle de Usuários

Especialista	Transcrição das respostas dos Especialistas
Métrica: Usuários ativos	
ID Esp_30	“Acredito que seja necessário justificar a relevância do controle de usuários.”
Métrica: Total de Usuários	
ID Esp_05	“O <i>chatbot</i> permite identificar o total de usuários que interagiram com ele?”
ID Esp_14	“Essa questão ficou confusa. Seria o total de usuário que conversou com o <i>bot</i> ?”
ID Esp_17	“Achei este item um pouco confuso, poderia ser repensado.”

Legenda: Identificador do participante - ID.
Fonte: Elaborado pelo autor (2021).

A métrica **Usuários ativos** obteve apenas a rejeição do especialista Esp_30: “Acredito que seja necessário justificar a relevância do controle de usuários”. O fato é que este especialista apenas diz acreditar que não acha necessário, mas não justifica. Considerando que os dezenove especialistas aceitaram a respectiva métrica, ela será mantida no *framework*.

Por fim, a métrica **Total de Usuários** teve 3 rejeições pelos especialistas. As rejeições não se dão no sentido de excluir a métrica, mas de reformular a questão de validação, conforme sugestão dos Esp_14 e Esp_15. Assim, a questão de validação da métrica acolherá a validação do Esp_05: “O *chatbot* permite identificar o total de usuários que interagiram com ele?”.

Em síntese, a análise das categorias e métricas realizadas pelos especialistas foi de grande importância para consolidar as métricas do *framework* e entregar a versão final com as considerações e apontamentos de cada métrica avaliada por eles. Vale reforçar que todas as métricas disponibilizadas para os especialistas estão fundamentadas na ISO/IEC 25010:2011 e nos estudos de trabalhos relacionados, que apontaram uma diversidade de métricas de avaliação de agentes conversacionais educacionais. Salienta-se que a tabela 4.2, já apresentada, traz a compilação das métricas ISO/IEC 9126, ISO/IEC 25010 e trabalhos relacionados, demonstrando a origem de cada métrica.

A tabela 5.13 apresenta as métricas do *framework* após as considerações dos especialistas.

Tabela 5.12 – Compilação das métricas ISO/IEC 9126, 25010, Figura 4.1 e trabalhos relacionados após a avaliação e validação dos especialistas

Fatores e Dimensões		Itens
Categoria	Métrica	Itens para avaliação da métrica
Funcionalidade	Funcionalidade completa	É apresentada a funcionalidade completa descrita pelo <i>chatbot</i> em análise.
	Qualidade das respostas	O <i>chatbot</i> apresenta respostas adequadas durante a conversação.
	Respostas de <i>fallback</i>	O <i>chatbot</i> é capaz de identificar a quantidade de vezes que falhou ou passou por uma situação de falha.
	Métricas de mensagens	O <i>chatbot</i> apresenta mensagens iniciais/finais de conversação ou saudações, a fim de iniciar ou encerrar um diálogo.
Eficiência	Comportamento em relação ao tempo	O <i>chatbot</i> apresenta um tempo adequado de respostas.
	Eficácia	O <i>chatbot</i> atende ,com precisão e integridade, quanto ao objetivo do diálogo.
Compatibilidade	Interoperabilidade	O <i>chatbot</i> é capaz de interagir com outro sistema ou plataforma específica.
	Reutilização	Informações da base de conhecimentos poderão ser aproveitadas para outros futuros <i>chatbots</i> .

Usabilidade	Reconhecimento de adequação	O <i>chatbot</i> é apropriado para suas necessidades.
	Aprendizagem	O <i>chatbot</i> oferece mecanismos para que o usuário possa aprender a usá-lo em situações de emergência ou em situação de uso imediato.
	Operabilidade	O <i>chatbot</i> é fácil de usar.
	Proteção contra erros de usuário	Em caso de algum possível erro na resposta do <i>chatbot</i> , o mesmo segue funcionando para novos diálogos.
	Acessibilidade	O <i>chatbot</i> pode ser utilizado por pessoas com diferentes características físicas e intelectuais?
Confiabilidade		
Confiabilidade	Maturidade	O <i>chatbot</i> atende às necessidades de confiabilidade em operação normal.
	Disponibilidade	Quando você precisou, o <i>chatbot</i> estava disponível?
	Tolerância a falha	Caso algum <i>software</i> do seu notebook, computador ou telefone apresente um possível erro de atualização de outro aplicativo, o <i>chatbot</i> segue funcionando normalmente.
	Testabilidade	O <i>chatbot</i> possibilita estabelecer critérios de testes a serem realizados, a fim de determinar se os critérios estabelecidos foram atendidos.
Segurança		
Segurança	Confidencialidade	O <i>chatbot</i> garante a confidencialidade de dados do usuário por algum tipo de login com senha, ou ainda, redireciona para um <i>link</i> específico de acesso.
	Integridade	O <i>chatbot</i> garante que o usuário não irá alterar determinadas configurações, tais como programações, botões, informações na base de conhecimento.
	Autenticidade	O <i>chatbot</i> permite sua identificação, a fim de garantir o sigilo ou diálogos já estabelecidos.
Portabilidade		
Portabilidade	Adaptabilidade	O <i>chatbot</i> funciona em diferentes sistemas operacionais e em diferentes configurações de <i>hardware</i> ou de <i>software</i> .
Satisfação		
Satisfação	Utilidade / Motivação para utilização	Após usar o <i>chatbot</i> , sinto-me satisfeito em relação ao uso baseado em meus objetivos, tais como os resultados do uso.
	Confiança	Confio que o <i>chatbot</i> vai comportar-se conforme o pretendido.
	Prazer/satisfação	O <i>chatbot</i> atendeu às suas necessidades pessoais.
	Afetividade	O <i>chatbot</i> apresenta diálogos com saudações, cordialidade e é capaz de responder com respeito e cordialidade as interações afetivas dos participantes.
	Humanidade	O <i>chatbot</i> é capaz de interagir de maneira convincente, satisfatória e natural.
	Ética e Comportamento	O <i>chatbot</i> apresenta respeito, inclusão e preservação da dignidade durante os diálogos.

Cobertura de Contexto	Completeness do contexto	O <i>chatbot</i> apresentou informações envolvendo uma completeness de contexto.
	Manter a conversa	O <i>chatbot</i> é capaz de manter o diálogo.
Controle de usuários	Usuários ativos	O <i>chatbot</i> permite identificar os usuários ativos ou on-line.
	Total de usuários	O <i>chatbot</i> permite identificar o total de usuários que conversaram com os usuários.

Fonte: Elaborado pelo autor, com base nas considerações dos especialistas.

De acordo com a tabela 5.13, os resultados apresentados na coluna Categoria, Métrica e Itens para a avaliação da métrica tratam do resultado para compor o *framework*. O percurso percorrido para encontrar métricas de avaliação vem ao encontro da abordagem de Goh, Kumar e Choon(2016), destacando a crescente necessidade de métricas para avaliar e medir a qualidade dos sistemas de agentes conversacionais em diferentes abordagens e domínio.

Nesta análise sobre a avaliação das métricas, constata-se que houve uma boa aceitação de todas as métricas sinalizadas para o *framework*. Destaca-se, desse modo, que nenhuma das métricas apresentadas serão excluídas por recomendação dos especialistas. Phillips (2018) aponta que é importante ter informações sobre o *chatbot*, para que se tenha conhecimento sobre oportunidades de crescimento e aprimoramento. Nesse sentido, as métricas propostas e revisadas por especialistas certamente apontarão as informações de crescimento para o *chatbot* educacional a ser avaliado. Neste estudo, houve uma série de sugestões para os textos já mencionados nas tabelas apresentadas, em que a maioria foi acolhida para o ajuste no *framework*, o que não desclassificou as métricas propostas, mas contribuiu para melhorar os textos a serem apresentados ao usuário para avaliá-las.

A partir da análise dos especialistas, algumas métricas foram realocadas para diferentes Categorias. A Categoria Manutenção não será mantida no *framework*, pois a métrica Testabilidade foi redirecionada para a Categoria Confiabilidade. Já a métrica Reutilização foi redirecionada para a Categoria Compatibilidade. Justifica-se a alteração de categorias em razão de as métricas terem uma maior relação com as categorias já existentes, facilitando no momento de selecionar, atribuir pesos e alterar os textos para a avaliação de um agente conversacional.

A métrica Tolerância à Falha foi substituída por Tolerância ao Erro, conforme recomendação do *Esp_19*. A alteração do nome da métrica se deu devido ao especialista ser um profissional de tecnologia da informação e atuar em testes de *softwares*. Assim, acredita-se que a sugestão do especialista ajudará no futuro em outras avaliações.

Vale destacar que os resultados da análise e validação das métricas realizadas pelos especialistas contribuíram para reorganizar as métricas dentro de suas respectivas categorias. Além disso, colaboraram para melhorar o texto e a forma de abordagem de uma determinada métrica, destacando-se que nenhum deles recomendou excluir nenhuma das métricas apresentadas.

A aceitação da compilação das métricas para compor o *framework*, as quais têm origem na NBR-ISO IEC 25010 e nos estudos relacionados de avaliação de agentes conversacionais, também é contemplada em recentes estudos de *chatbots*, como é o caso do estudo proposto por Almalki e Azeez (2020). Os autores abordaram o uso de *chatbots* no contexto da saúde, utilizando uma avaliação no contexto das métricas de Usabilidade e Confiabilidade. Destaca-se que tais métricas aparecem na avaliação dos autores e que já estão contempladas no *framework*.

Outro recente estudo envolveu a avaliação do *chatbot* como ferramenta de alfabetização em saúde proposto por Mokmin e Ibrahim (2021). Eles também realizaram uma avaliação envolvendo as métricas de Satisfação do Usuário, Performance e Usuários Ativos.

Já em relação aos estudos tratam de questões de tecnologias e implementação, abordando *chatbots*, Caldarini, Jaf e Macgarry (2022) discutem questões abrangendo diálogos de conversação. O *framework* com métricas de avaliação de agentes conversacionais educacionais proposto nesta tese indica as seguintes métricas para esse tipo de avaliação: Métrica de Mensagens e Manter a Conversa. Nesse sentido, as métricas apresentadas no *framework* ainda propõem e especificam novas possibilidades e direcionamentos para a avaliação envolvendo o contexto dos diálogos de conversação.

Em um outro contexto que abrange a avaliação de agentes conversacionais, Acheampong, Wenyu e Nunoo-Mensah (2020), em sua discussão do desafio na empatia nas respostas emocionalmente ajustadas, em que o *chatbot* deve detectar o estado emocional do humano e fornecer uma resposta à altura, constata-se um grande desafio para a construção de *chatbots*, envolvendo diversas questões técnicas para o desenvolvimento e o uso de algoritmos apropriados. No entanto, realizar avaliações em diferentes etapas do desenvolvimento com métricas que presumem questões previstas no *framework* proposto neste estudo pode ser um grande diferencial para uma melhoria contínua, estabelecendo-se uma evolução também contínua. A questão de avaliação que considera a empatia nas respostas propostas por Acheampong, Wenyu e Nunoo-Mensah (2020) pode ser avaliada a partir das métricas da categoria Satisfação do *Chatbot* (Humanidade e Afetividade) propostas no *framework*.

Nesse sentido, constatou-se que as métricas validadas pelos especialistas continuam fortemente sendo utilizadas por recentes estudos que contêm a avaliação de agentes conversacionais, especificamente, agentes educacionais.

5.3.4 Público do *ChatAval*

A seção 4 desta tese apresentou o percurso de construção do *framework* e sugeriu potenciais públicos que poderão usufruir do resultado deste estudo, como os usuários, testadores e desenvolvedores.

O público classificado como Usuários são os estudantes que desejam utilizar as métricas de avaliação dos agentes conversacionais. Os Testadores são usuários avançados, tais como alunos, professores e alunos avançados que desejam avaliar métricas de funcionamento dos agentes conversacionais. Já os Desenvolvedores são os programadores, analistas de sistemas, profissionais de tecnologia da informação, que irão atuar no desenvolvimento dos *chatbots* educacionais e precisam de métricas como base para desenvolver os respectivos parâmetros. Nesse sentido, a tabela 5.14 reclassifica e sugere as métricas para os possíveis públicos.

Tabela 5.13 – Mapeamento das métricas para o público do *ChatAval*

Fatores e Dimensões		Público do <i>ChatAval</i>		
Categoria	Métrica	Usuários	Testadores	Desenvolvedores
Funcionalidade	Funcionalidade completa	X	X	
	Qualidade das respostas	X	X	
	Respostas de <i>fallback</i>		X	X
	Métricas de mensagens	X		
Eficiência	Comportamento em relação ao tempo		X	
	Eficácia		X	
Compatibilidade	Interoperabilidade		X	X
	Reutilização			X
Usabilidade	Reconhecimento de adequação	X	X	

	Aprendizagem	X	X	
	Operabilidade	X		
	Proteção contra erros de usuário	X	X	
	Acessibilidade	X		
Confiabilidade	Maturidade		X	
	Disponibilidade	X	X	
	Tolerância à falha	X	X	X
	Testabilidade		X	X
Segurança	Confidencialidade		X	X
	Integridade		X	X
	Autenticidade			X
Portabilidade	Adaptabilidade	X	X	X
Satisfação	Utilidade / Motivação para utilização	X	X	
	Confiança	X	X	
	Prazer/satisfação	X	X	
	Afetividade	X	X	
	Humanidade	X		
	Ética e Comportamento	X	X	
Cobertura de Contexto	Completeness do contexto		X	
	Manter a conversa	X	X	
Controle de usuários	Usuários ativos		X	X
	Total de usuários		X	X

Fonte: Elaborado pelo autor, com base nas considerações dos especialistas.

A tabela 5.14 sugere métricas selecionadas e validadas por especialistas para serem utilizadas por possíveis diferentes públicos. Destaca-se que uma mesma métrica pode atender, ao menos, um ou até todos os públicos sugeridos neste estudo, os quais poderão utilizar o

framework, como é o caso das métricas Funcionalidade Completa e Qualidade das Respostas, que podem ser utilizadas por um público de Usuários ou Testadores.

Destaca-se, por exemplo, a métrica Tolerância à Falha da Categoria Confiabilidade, que é sugerida para ser avaliada para os mais diversos públicos, pois um usuário pode usar um determinado agente conversacional e este apresentar falhas, seja de comunicação com um *link*, ou ao navegar de um *tablet* ou computador, por exemplo. A mesma métrica pode ser utilizada por Testadores de aplicativos. Podem, ainda, ser detectadas falhas diversas de funcionamento, de acesso, entre tantas as possibilidades que existem no meio tecnológico. Por fim, a métrica Tolerância à Falha também pode ser utilizada por grupos de desenvolvedores, que estão criando seus códigos, configurando com banco de dados etc., para detectar falhas quando se constrói um agente conversacional.

Outras métricas parecem estar mais direcionadas aos Testadores, como por exemplo, Comportamento em Relação ao Tempo, pois, a partir de um olhar de testes, essa métrica está mais para um testador do que para um usuário. Um usuário pode ser a figura principal para a Métrica de Mensagens, por exemplo, pois trata-se da figura que vai conversar, fazer as perguntas não previstas pelos desenvolvedores dos agentes conversacionais.

Assim, vale salientar que a tabela 5.14 está apenas sugerindo métricas para possíveis públicos, no entanto qualquer um deles poderá utilizar qualquer métrica já validada no *framework*. Nesta conjuntura, suponha-se que haja um usuário classificado como Usuário. Este usuário é um aluno de qualquer curso de ensino superior, e imagine-se como esse aluno utiliza a métrica Testabilidade da Categoria Confiabilidade. O estudante classificado como Usuário não tem acesso a dados de acesso, senha, segurança, pois possivelmente essa informação estará sob a responsabilidade dos Desenvolvedores ou, em alguns casos, também dos Testadores.

Por fim - e também reforçando, a tabela 5.14 sugere algumas métricas para públicos do Protótipo. As métricas foram propostas e sugeridas pelo autor da tese, tendo em vista a experiência durante o desenvolvimento do aplicativo *ChatAval* e a experiência em aplicar as métricas do *ChatAval* com os alunos do Curso Técnico em Informática para Internet, no entanto todos os usuários terão todas as métricas à sua disposição.

5.4 VALIDAÇÃO DO *FRAMEWORK* COM O APLICATIVO *CHATAVAL*

Esta subseção apresenta a validação do *framework* e, para isso, foi construído o *ChatAval*. Reitere-se que o *framework* é um conjunto de métricas para a avaliação de *chatbot* no contexto educacional. Para facilitar e agilizar esta validação, foi implementado o aplicativo

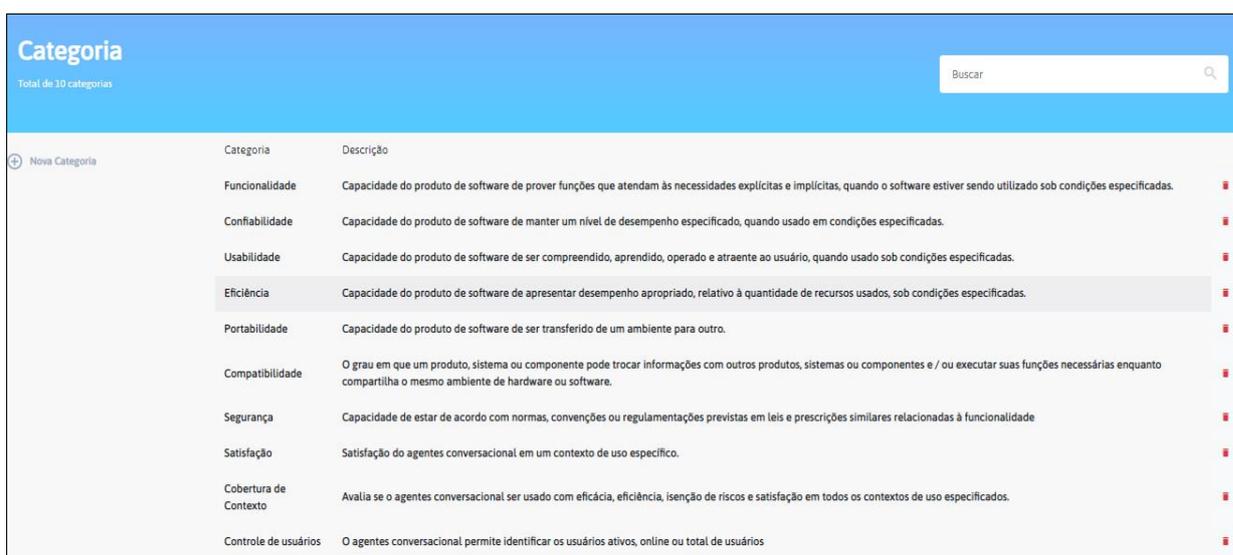
ChatAval, que é um sistema *web* construído para automatizar o uso e o *framework*, proporcionando o cadastro de categorias, subcategorias, questões, elaboração de questionários, distribuição de pesos nas categorias e suas respectivas métricas, além de gerar um *link* para a avaliação das questões e apresentar os resultados validados automaticamente.

O *ChatAval* passou por um teste piloto, conforme apresentado na subseção 5.2.2 (Resultados da Fase 2.2), em que foram utilizadas as métricas da ISO-9126, a fim de validar o aplicativo com alunos do Curso Técnico em Informática para Internet, sendo que eles tiveram a oportunidade de realizar uma série de atividades envolvendo conceitos de agentes conversacionais, práticas de programação nas linguagens HTML e AIML, além de construção e validação dos agentes conversacionais METIS, ATENA e ALTEIA.

Neste momento, com o *ChatAval* construído e validado, com as métricas do *framework* examinadas sobre o olhar crítico dos especialistas, houve a parametrização das novas Categorias e subcategorias (métricas) do *framework* para o Sistema *ChatAval*. A seguir, apresentam-se as principais telas que receberam as novas categorias, métricas (subcategorias), texto de apresentação da métrica e o texto final para ser validado.

A figura 5.23 apresenta a tela do *ChatAval* com todas as categorias e suas respectivas descrições validadas pelos especialistas.

Figura 5.23 – Tela inicial do *ChatAval* com as categorias validadas pelos especialistas



The screenshot shows the 'Categoria' screen of the ChatAval system. At the top, there is a blue header with the title 'Categoria' and a search bar labeled 'Buscar'. Below the header, there is a table listing 10 categories. The table has three columns: 'Categoria', 'Descrição', and a red square icon on the right. The categories listed are: Nova Categoria, Funcionalidade, Confiabilidade, Usabilidade, Eficiência, Portabilidade, Compatibilidade, Segurança, Satisfação, Cobertura de Contexto, and Controle de usuários. The 'Eficiência' row is highlighted in grey.

Categoria	Descrição	
Nova Categoria		
Funcionalidade	Capacidade do produto de software de prover funções que atendam às necessidades explícitas e implícitas, quando o software estiver sendo utilizado sob condições especificadas.	
Confiabilidade	Capacidade do produto de software de manter um nível de desempenho especificado, quando usado em condições especificadas.	
Usabilidade	Capacidade do produto de software de ser compreendido, aprendido, operado e atraente ao usuário, quando usado sob condições especificadas.	
Eficiência	Capacidade do produto de software de apresentar desempenho apropriado, relativo à quantidade de recursos usados, sob condições especificadas.	
Portabilidade	Capacidade do produto de software de ser transferido de um ambiente para outro.	
Compatibilidade	O grau em que um produto, sistema ou componente pode trocar informações com outros produtos, sistemas ou componentes e / ou executar suas funções necessárias enquanto compartilha o mesmo ambiente de hardware ou software.	
Segurança	Capacidade de estar de acordo com normas, convenções ou regulamentações previstas em leis e prescrições similares relacionadas à funcionalidade	
Satisfação	Satisfação do agentes conversacional em um contexto de uso específico.	
Cobertura de Contexto	Avalia se o agentes conversacional ser usado com eficácia, eficiência, isenção de riscos e satisfação em todos os contextos de uso especificados.	
Controle de usuários	O agentes conversacional permite identificar os usuários ativos, online ou total de usuários	

Fonte: Elaborado pelo autor (2021).

De acordo com a figura 5.23, todas as 10 categorias, com suas respectivas descrições, estão cadastradas no *ChatAval*. A partir do cadastro das categorias, é possível cadastrar as métricas (subcategorias), conforme se pode observar a seguir (figura 5.24).

Figura 5.24 – Tela *ChatAval* com parte das métricas (Subcategorias)

Satisfação	- Utilidade / Motivação para utilização	Grau com que o chatbot apresenta satisfação em relação ao uso.	Após usar o chatbot, sinto-me satisfeito em relação ao uso baseado em meus objetivos, tais como, os resultados do uso.	■
Satisfação	- Confiança	Grau em que um usuário ou outra parte interessada tem confiança de que um produto ou sistema se comportará conforme pretendido.	Confoio que o Chatbot vai comportar-se conforme o pretendido.	■
Satisfação	- Prazer/satisfação	Grau em que um usuário obtém prazer em atender às suas necessidades pessoais	O chatbot atendeu às suas necessidades pessoais.	■
Satisfação	- Afetividade	Grau em que um usuário obtém em relação as saudações, cordialidade, responder com respeito e cordialidade as interações afetivas e não afetivas dos participantes.	O chatbot apresenta diálogos com saudações, cordialidade e é capaz de responder com respeito e cordialidade as interações afetivas dos participantes.	■
Satisfação	- Humanidade	Grau de capacidade do agente conversacional em interagir de maneira convincente, satisfatória e natural.	O chatbot é capaz de interagir de maneira convincente, satisfatória e natural.	■
Satisfação	- Ética e Comportamento	Capacidade do agente conversacional em apresentar respeito, inclusão e preservação da dignidade durante os diálogos.	O chatbot apresenta respeito, inclusão e preservação da dignidade durante os diálogos.	■
Cobertura de Contexto	- Completude do contexto	Grau em que um agente conversacional pode ser usado com eficácia, eficiência, isenção de risco e satisfação em contextos além daqueles inicialmente especificados.	O chatbot apresentou informações envolvendo uma completude de contexto.	■
Cobertura de Contexto	- Manter a conversa	Grau em que o agente conversacional é capaz de manter a conversação.	O chatbot é capaz de manter o diálogo.	■
Controle de usuários	- Total de usuários	Grau com que o agente conversacional consegue informar e controlar o total de usuários.	O chatbot permite identificar o total de usuários que conversaram com os usuários.	■
Controle de usuários	- Usuários ativos	Grau com que o agente conversacional possibilita ter o controle de usuário ativos e não ativos.	O chatbot permite identificar os usuários ativos ou online.	■

Fonte: Elaborado pelo autor (2021).

Conforme a figura 5.24, todas as métricas propostas pelo autor da tese e validadas pelos especialistas estão atualizadas no *ChatAval*. Esta figura apresenta parte de algumas categorias e de suas respectivas métricas. Pode-se observar as categorias Satisfação, Cobertura de Contexto e Controle de Usuários com suas respectivas métricas, descrições e pergunta inicial sugerida para a avaliação dos agentes conversacionais.

A figura 5.25 apresenta o resumo do *ChatAval* com as respectivas categorias e a quantidade de métricas (subcategorias) cadastradas.

Figura 5.25 – Quantidade de métricas (subcategorias) para cada categoria

Subcategoria	
Funcionalidade	4
Confiabilidade	4
Usabilidade	5
Eficiência	2
Portabilidade	1
Compatibilidade	2
Segurança	3
Satisfação	3
Cobertura De Contexto	2
Controle De Usuários	3

+ Nova Subcategoria

Fonte: Elaborado pelo autor (2021).

De acordo com a figura 5.25, constata-se que o *ChatAval* está parametrizado com todas as categorias e respectivas métricas propostas para o *framework*, o que vai se refletir diretamente nas demais funcionalidades do sistema, conforme se pode constatar na figura 5.26.

Figura 5.26 - Cadastro de um formulário no *ChatAval* com as métricas validadas

Cobertura de Contexto

Peso Cobertura de Contexto

- Completude do contexto O chatbot apresentou informações envolvendo uma completude de contexto.

- Manter a conversa O chatbot é capaz de manter o diálogo.

Controle de usuários

Peso Controle de usuários

- Total de usuários O chatbot permite identificar o total de usuários que conversaram com os usuários.

- Usuários ativos O chatbot permite identificar os usuários ativos ou online.

Criar

Fonte: Elaborado pelo autor (2021).

A figura 5.26 apresenta um recorte das categorias do *ChatAval*, com suas respectivas categorias, métricas e questões de avaliação. Ao todo, são 10 categorias com 31 métricas cadastradas. Ainda conforme a figura 5.26, constata-se nas Categorias Cobertura de Contexto e Controle de Usuários, a funcionalidade para informar o peso de cada métrica. Cada categoria e cada métrica cadastrada contém um campo exatamente igual que faz com que a validação das notas das categorias e subcategorias sejam iguais a 10 para um correto funcionamento.

Por fim, esta seção teve o objetivo de demonstrar o sistema *ChatAval* parametrizado com todas as categorias e suas respectivas métricas validadas por especialistas. O Sistema *ChatAval* já foi validado em um estudo piloto, como mencionado na subseção 5.2.2 - Resultados da Fase 2.2, e as métricas que compõem o *framework* também foram validadas, como registrado na subseção 5.3.3 – Resultado da análise e validação das métricas.

Assim, esta seção demonstrou a interseção das métricas propostas para o *framework* desenvolvido para avaliar agentes conversacionais educacionais cadastradas e devidamente configuradas no *ChatAval*. Em vista disso, caso um usuário, especialista ou desenvolvedor tenha necessidade ou interesse em avaliar um determinado *chatbot*, o *ChatAval* está apto a auxiliá-lo com todas as suas funcionalidades.

6 CONCLUSÃO

A presente pesquisa abordou métricas de avaliação de agentes conversacionais educacionais. Partiu-se das Tecnologias de Informação e Comunicação, destacando as ferramentas que possibilitam uma comunicação em tempo real, que permitem a participação e intervenção dos sujeitos, como é o caso dos *chatbots*. Constatou-se que as TICs podem contribuir, de maneira significativa, com a interação entre os pares a partir do uso de *chatbots*, destacando a característica de, segundo Leonhardt et al. (2005), permanecer 24 horas por dia à disposição. Entre as diferentes finalidades do uso dos *chatbots*, Satu, Parvez e Mamun (2015) apontam aplicações relacionadas ao *e-learning*, governo eletrônico, modelo de diálogo, humor, entre outros.

Já no contexto educacional, constatou-se que o uso do *chatbot* aparece de forma significativa e relevante, apontando diversos resultados positivos quando utilizado no contexto educacional (SANTOS; ROLAND; SILVA, 2019; PASCHOA; CHICON; FALKEMBACH, 2017; CLEMENTE, 2016; SANTANA, 2017). Paralelamente, o relatório *Hype Cycle for Digital Government Technology 2019* (GARTNER'S, 2019) destaca o crescente uso de *chatbots* com excelentes expectativas para implementação e desenvolvimento até 2024. Em verdade, embora o uso de *chatbots* seja promissor e relevante para a educação, além de estar com excelentes perspectivas de uso e crescimento para os próximos anos, a implantação e avaliação de um *chatbot* oferece inúmeros desafios. Entre eles, destaca-se a avaliação com métricas adequadas sob a perspectiva de diversos parâmetros para a avaliação, em que, nos estudos de agentes conversacionais educacionais, identificou-se que grande parte dos estudos envolve a avaliação a partir de um questionário com perguntas elaboradas aleatórias e específicas, mas sem fundamentação de métricas de avaliação dos respectivos agentes.

Nesse sentido, a avaliação de *chatbots* ganha destaque para enfatizar a necessidade de métricas para avaliar e medir a qualidade das respostas produzidas pelos sistemas de agentes conversacionais, com base em diferentes abordagens, domínios e funcionalidades e categorias de contexto de conversa (GOH; KUMAR; CHOON, 2016; HUSSAIN; SIANAKI; ABABNEH, 2019). Assim, realizar a avaliação de um *chatbot* educacional é um desafio, considerando que há poucos estudos que abordam tais avaliações.

Para dar conta de uma investigação, com o intuito de avaliar os agentes conversacionais educacionais a partir de métricas de avaliação, dividiu-se esta tese em três fases, com o propósito de construir um percurso para a avaliação. Na primeira fase, confirmou-se a hipótese de que os alunos do Curso Técnico em Informática para Internet têm acesso a

computadores ou *notebooks*, boa velocidade de internet e espaços físicos para realizar estudos em suas residências. No entanto, a maioria deles não têm conhecimentos de algoritmos, de uma linguagem de programação ou de desenvolvimento de um *site*, mas manifestaram excelentes expectativas para a aprendizagem de programação. Esta fase possibilitou conhecer o público que irá testar um protótipo do *framework* com as métricas de avaliação de agentes conversacionais.

A segunda fase do estudo dividiu-se em 2.1 e 2.2 e foi muito significativa, porque introduziu e agregou conhecimentos de linguagem de programação AIML e HTML ao alunos do curso Técnico em Informática, sendo que grande parte deles não possuía tais conhecimentos de construção de *sites*, conforme se constatou na fase 1. A fase 2 oportunizou aos alunos conhecer a linguagem de programação HTML e AIML na prática, a partir de desenvolvimento de *sites* e da elaboração de códigos em AIML para serem inseridos no site. A prática de 35 horas de programação em HTML e AIML capacitou os alunos a compreender ambas as linguagens de programação, além de adicionar, no *site*, algumas informações referentes aos conhecimentos em que o agente conversacional é capaz de responder, afinal, uma das limitações dos agentes é ter respostas adequadas para todas as perguntas realizadas pelos alunos. Houve resultados muito positivos em relação à construção do *site* final e desenvolvimento da base de conhecimento. A proposta de elaboração do *site* e da elaboração de diálogos em AIML foi realizada individualmente.

Destaca-se que, até este momento, já havia conhecimento do perfil dos alunos, sabendo-se que eram capazes de elaborar um *site* a partir de *tags* básicas de HTML e de construir base de conhecimentos de agentes conversacionais na linguagem AIML. Na sequência, foi necessário possibilitar aos alunos um momento de conversação com os agentes conversacionais METIS, ALTEIA e ATENA, a fim de realizar um teste e avaliar suas funcionalidades.

Neste estudo, entendeu-se que, para avaliar um agente conversacional, é preciso saber o que é um agente conversacional, visto que se deve compreender as possibilidades e limitações no que tange ao desenvolvimento e implementação dos diálogos na programação dos agentes. Assim, na fase 2.2, a duração de uma hora de experiência de conversação e avaliação para cada agente conversacional foi fundamental para que os alunos fossem capazes de avaliar os agentes conversacionais. Basicamente, a Fase 1, a Fase 2.1 e a Fase 2.2 foram preparações para que os alunos estivessem capacitados para responder e avaliar os agentes a partir de métricas de avaliação de *softwares*.

Para dar conta da investigação das métricas e continuidade da Fase 2.2, realizou-se uma avaliação baseada nas métricas da NBR ISO/IEC 9126. Esta avaliação foi um estudo piloto para modelagem do *framework* proposto, para avaliar agentes conversacionais educacionais. Os agentes que compuseram os testes foram METIS, ATENA e ALTEIA. A possibilidade de utilizar algumas das categorias e subcategorias, bem como de elaborar questões a partir da definição das métricas permitiu que os alunos realizassem a avaliação dos agentes sem perceber que estavam utilizando métricas específicas. Além disso, a distribuição do peso no questionário para cada métrica de avaliação possibilitou direcionar a avaliação para a funcionalidade que se desejou avaliar.

A avaliação ocorreu no *ChatAval*, que foi desenvolvido especificamente para cadastrar e gerenciar categorias, métricas e descrição de cada métrica. Os agentes avaliados também são cadastrados, o que permite aos usuários realizar diversas avaliações a partir de cada métrica. Além disso, o *ChatAval* possui funcionalidades para automatizar o envio de questionários dos alunos a partir de um *link*, e para apresentar automaticamente o resultado da avaliação de um ou mais agentes conversacionais, à medida que o questionário é preenchido. Nas pesquisas realizadas, não se constatou nenhum aplicativo com tal finalidade ou funcionalidade.

O *ChatAval* é totalmente gerenciável e isso significa que é possível adicionar categorias, métricas, alterar as questões iniciais de abordagens e adicionar, alterar ou excluir a relação dos agentes conversacionais já cadastrados. Trata-se de um sistema padronizado e adaptável para calcular métricas de avaliação de agentes conversacionais. Destaca-se, aqui, que a construção do aplicativo *ChatAval*, compondo as métricas de avaliação de *softwares*, tratou de um resultado extremamente relevante deste estudo, pois não se identificou um aplicativo com as mesmas finalidades que ele apresenta.

Embora tenha sido uma proposta de validar o estudo piloto do *framework*, os resultados que envolveram a avaliação com as abordagens das métricas, com base na NBR ISO/IEC 25010, sinalizaram um cenário positivo em relação às categorias Confiabilidade, Eficiência, Funcionalidade e Manutenibilidade, cujos agentes ALTEIA e METIS ficaram com média final de 7.2 e a agente ATENA com média final de 6.9, considerando-se uma escala de 0 a 10. Acredita-se que os resultados pudessem ser diferentes, caso outras métricas fossem selecionadas para avaliar os respectivos agentes.

A funcionalidade do *ChatAval* ocorreu perfeitamente em relação à aplicação do questionário e demais funcionalidades. Uma desvantagem quanto ao uso do *ChatAval* é que ocorreram travamentos quando usado por 40 alunos ao mesmo tempo, principalmente no momento de se entrar no *link* para realizar as respostas e, também, no momento de enviá-las. A

solução foi gerenciar o número de alunos que acessavam por vez o *ChatAval*. Vale destacar que o problema sobre os travamentos estava relacionado com a capacidade de quantidade de acessos do servidor, no qual estava hospedado o *ChatAval*, onde um *upgrade* no número de acessos solucionaria esse problema.

A respeito da fase 3 do estudo, houve o percurso para a construção das métricas com a origem da ISO/IEC 9126, ISO/IEC 25010 e trabalhos relacionados, contabilizando 31 métricas estrategicamente distribuídas em categorias para compor o *framework* para avaliar e elencar o que um agente conversacional pode e deve ter para ser funcional. Ainda na fase 3, a validação por 20 especialistas que atuam nos campos da Educação, Computação e Informática na Educação contribuiu para validar cada métrica proposta pelo autor da tese, sendo que as principais contribuições e considerações referem-se à realocação de métricas para outra categoria e alteração do texto para a avaliação da métrica. Ressalte-se que nenhum dos especialistas rejeitou ou sugeriu que fosse excluída uma determinada métrica, o que demonstra que aquelas com origem dos trabalhos relacionados e as métricas da ISO/IEC 9126 e da ISO/IEC 25010 atenderam, de forma satisfatória, a composição do *framework*. Assim, pode-se concluir que o objetivo da tese em desenvolver um *framework* com métricas de avaliação de *software*, a fim de elencar o que um agente conversacional educacional deve ter para ser utilizado como referência para a avaliação de *chatbots* de contexto educacional foi concluído com sucesso.

Sob a perspectiva do *ChatAval*, atualizaram-se os textos de avaliação, realocaram-se as métricas e alteraram-se as categorias a partir das recomendações dos especialistas. O *ChatAval* está atualizado com todas as métricas que compõem o *framework* e já pode ser utilizado por diferentes públicos. Entende-se diferentes públicos os grupos de usuários classificados com distintos níveis de interesse ou de conhecimentos, como sugerido neste estudo. Assim, os Usuários são os alunos, estudantes, que desejam utilizar as métricas de avaliação; os Testadores são usuários mais avançados, como os professores e alunos que desejam avaliar as métricas de um *chatbot*; e os Desenvolvedores, os quais são os profissionais programadores, analistas de sistemas, entre outros, que desejam consultar as possíveis métricas que um agente conversacional pode e deve ter. Enfatiza-se que esse grupo de profissionais é apenas uma sugestão para categorizar o uso do *ChatAval* para diferentes públicos, afinal, todos os usuários terão todas as métricas pesquisadas e validadas à sua disposição para a avaliação de um determinado agente conversacional educacional.

Por fim, reitera-se que a contribuição desta tese consistiu em entregar um *framework* automatizado, com o apoio do aplicativo *ChatAval*, o qual consiste em um aplicativo *web*

desenvolvido e testado especificamente para automatizar a avaliação de agentes conversacionais educacionais, tendo como base métricas de avaliação de *softwares*.

Constatou-se que, ao longo dos diversos trabalhos relacionados abordados neste estudo, todos os agentes conversacionais desenvolvidos apresentam um determinado contexto de conversação, para tratar sobre um assunto educacional específico, que pode estar relacionado à saúde, à condição física, à conversação sobre um domínio turístico, ou a qualquer outro entre tantos assuntos que um *chatbot* é capaz de ser treinado para responder.

Verificou-se, ainda, que os resultados das métricas utilizadas para o desenvolvimento do *framework* de avaliação de agentes conversacionais educacionais ratificam resultados e métricas já citados no referencial teórico, embora, para a grande maioria dos trabalhos que abordam ou apresentam uma determinada avaliação sobre agentes conversacionais, observa-se que realizam avaliações de maneiras distintas, como, por exemplo, os estudos de Goh, Kumar e Choon(2016) e Sedoc et al. (2019). Estes autores realizaram suas análises dos *chatbots* desenvolvidos, focando, especificamente, na análise das métricas de Testes de Conversação ou Qualidade das Respostas.

Já o foco de avaliação específico dos estudos de Niculescu et al. (2014) e O'Shea, Bandar e Crockett (2011) debruçou-se sobre as métricas de Usabilidade e Funcionalidade. Por outro lado, Phillips (2018) e Sedoc et al. (2019) abordaram apenas a importância de se conhecer as possibilidades de métricas de mensagens e usuários engajados.

Nesse sentido, percebe-se que os autores realizam suas avaliações, considerando seus interesses. No entanto, poderiam ampliar e complementar suas avaliações para além de suas percepções e interesses, com o apoio do *framework* desenvolvido nesta tese enriquecido com métricas validadas especificamente para a avaliação de agentes conversacionais educacionais, proporcionando, assim, uma avaliação completa sobre diversos aspectos de avaliação de agentes conversacionais educacionais.

Cumprе salientar que as métricas que compõem a avaliação do *framework* para a avaliação de agentes conversacionais educacionais também têm sido apontadas como relevantes em trabalhos recentes, tais como, o de Joseph Raj, Anton e Johnson Raj (2021), que desenvolveram um *chatbot* para atender a sociedade que sofre com a Covid-19 e focaram seus testes nas métricas de Testes de Conversação e Qualidade; ou, ainda, os estudos de Almalki e Azeez (2020), que abordaram a experiência do usuário de seu *chatbot*.

Ainda em estudos recentes, Mokmin e Ibrahim (2021) realizaram avaliações envolvendo as métricas de Satisfação do usuário, Performance, Usuários ativos. Já Drouin et al. (2022) avaliaram Confiabilidade, Ética e Comportamento, Humanidade e Afetividade.

Por fim, pode-se concluir que o *framework* desenvolvido com métricas de avaliação contém métricas bastante completas para avaliar agentes conversacionais educacionais, pois as que compõem o *framework* foram identificadas em estudos mais recentes. O produto desta tese é, portanto, o *framework* com métricas de avaliação de agentes conversacionais educacionais que está automatizado pelo aplicativo *web ChatAval*. As métricas que compõem o *framework* poderão também ser utilizadas como referência e apoio para a construção de agentes conversacionais educacionais ou *chatbots* com outras finalidades além das educacionais.

6.1 TRABALHOS FUTUROS

Ao longo do desenvolvimento e a partir dos resultados desta tese, identificaram-se algumas possibilidades que poderão ser continuadas em trabalhos futuros.

As métricas de avaliação têm origem na ISO/IEC 25010 e em 12 trabalhos relacionados que abordaram a avaliação de agentes conversacionais educacionais. Entretanto, novas métricas poderão ser adicionadas a partir da investigação de novos estudos que envolvem a avaliação de *Chatbots*. Nos estudos recentes apresentados especificamente na subseção 2.3.13, não foram identificadas métricas que não tenham sido contempladas no *framework*, mas foram constatados agentes desenvolvidos para atender a demandas recentes, como a Covid-19, por exemplo. Porém, as análises e testes estão contempladas nas métricas já utilizadas no *framework* proposto, o que demonstra que estão bastante completas.

O protótipo *ChatAval* está funcionando a partir das características e métricas investigadas para o desenvolvimento e automação do *framework*, todavia, para trabalhos futuros, sugere-se a implementação de funcionalidades com permissões de leitura, execução e edição para os recursos desenvolvidos, tais como o cadastro e gerenciamento das categorias, métricas, textos que envolvem os questionários e pesos de cada questão.

Ademais, o protótipo *ChatAval* poderá ser migrado para um novo servidor que suporte o acesso ao mesmo tempo de, no mínimo, 50 usuários para responder as avaliações dos agentes conversacionais que serão avaliados. Isso ajudará a automatização e funcionalidade do aplicativo.

Registra-se que o protótipo *ChatAval* foi validado a partir de métricas da ISO/IEC 9126 por estudantes do Ensino Médio com conhecimentos básicos de linguagem de programação HTML e com noções básicas de programação de diálogos em AIML. Sugere-se, para trabalhos futuros, a avaliação com todas as métricas validadas a partir da ISO/IEC 25010

e dos 12 trabalhos relacionados, os quais compõem o *framework*, em outros agentes conversacionais educacionais.

O referido *framework* contém uma diversidade de métricas para a avaliação de agentes conversacionais educacionais. Este estudo sugeriu um público que foi identificado como Usuários, Testadores e Desenvolvedores. Sugere-se, por fim, realizar a validação das métricas sugeridas para os respectivos públicos em outros agentes conversacionais, a fim de validar a classificação destas para cada perfil de usuário inicialmente proposto.

6.2 CONTRIBUIÇÕES

As principais contribuições desta tese podem ser sintetizadas da seguinte forma:

- a) Um *framework* com métricas de avaliação de agentes conversacionais educacionais;
- b) Desenvolvimento de um aplicativo *web* (*ChatAval*) com funcionalidades específicas para a avaliação a partir de métricas de avaliação de agentes conversacionais;
- c) Realização de um estudo com alunos de um Curso Técnico em Informática envolvendo técnicas de programação em HTML e AIML, com o propósito de testar as funcionalidades do Protótipo *ChatAval*;
- d) As publicações científicas em periódicos nacionais e internacionais e capítulos de livros (Apêndice A).

REFERÊNCIAS

- ABUSHAWAR, Bayan.; ATWELL, Eric. ALICE Chatbot: Trials and Outputs. **Computación y Sistemas**, v. 19, n. 4, p. 625–632, 2015. Disponível em: < researchgate.net/publication/289684788_ALICE_chatbot_Trials_and_outputs>. Acesso em: jan. 2021.
- ACHEAMPONG, F. A.; WENYU, C.; NUNOO-MENSAH, H. Text-based emotion detection: Advances, challenges, and opportunities. **Engineering Reports**, v. 2, n. 7, jul./2020, p. 3-24. <https://doi.org/10.1002/eng2.12189>. 2020. Disponível em: < <https://onlinelibrary.wiley.com/doi/full/10.1002/eng2.12189> >. Acesso em: jan. 2021.
- AGUIAR, Eliane V. B., VITORINO, N. Ferramentas e métodos para aperfeiçoamento do funcionamento de um agente conversacional. **Tecnologias Digitais na educação: Pesquisas e Práticas Pedagógicas**. Essentia Editora, 2015, p. 106-114, DOI: 10.19180/978-85-99968-49-9.7. Disponível em: <<http://www.bibliotekevirtual.org/livros/registrados/pdfs/978-85-99968-49-9/c07.pdf>>. Acesso em: jan. 2021.
- ALMALKI, Manal. Perceived Utilities of COVID-19 Related Chatbots in Saudi Arabia: a Cross-sectional Study. **Acta Inform. Med.** 28(3): 218-223, 2020. Disponível em: <<https://pesquisa.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/resource/pt/covidwho-1016349>>. Acesso em: jan. 2021.
- ALMALKI, Manal; AZEEZ, Fahad. Health Chatbots for Fighting COVID-19: a Scoping Review. **Acta Inform. Med.** 28(3): 241-247, dez./2020. Disponível em: < <https://www.ncbi.nlm.nih.gov/pmc/issues/375324/>>. Acesso em: jan. 2021.
- ALTEIA (2020). **ALTEIA** (Agente Lexical de apoio a Tarefas Educacionais usando Inteligência Artificial). Disponível em: <<http://avatar.cinted.ufrgs.br/ALTEIA/gui/plain/index.php#end>>. Acesso em: 18 abr. 2020.
- ARAUJO, D. et al. Autoria automática de objetos de aprendizagem a partir de captura multimídia e associação a estilos de aprendizagem. **III Congresso Brasileiro de Informática na Educação (CBIE 2014). XXV Simpósio Brasileiro de Informática na Educação (SBIE 2014)**, Faculdade de Computação, Universidade Federal de Uberlândia, p. 229-238, 2014. Disponível em: <https://www.researchgate.net/publication/269400383_Autoria_Automatica_de_Objeto_de_Aprendizagem_a_Partir_de_Captura_Multimidia_e_Associacao_a_Estilos_de_Aprendizagem>. Acesso em: jan. 2021.
- ATENA (2020). **ATENA** (Agente Tutora para Ensino e Navegação no Ambiente de Física). Disponível em: <<http://avatar.cinted.ufrgs.br/ATENAweb/gui/jquery/?>>. Acesso em: 15 mar. de 2020.
- AUGELLO, A. et al. An approach to enhance chatbot semantic power and maintainability: Experiences within the frasi project. **IEEE**. Out./2012. DOI: 10.1109/ICSC.2012.26. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/6337103>>. Acesso em: 2 jun., 2017.
- BLOOM, B. S. The 2 Sigma problem: The Research for Methods of Group Instruction as Effective as One-to-one Tutoring. **Educational Researcher**. v. 13, n. 6, p. 4-16, Jun./Jul.1984. Disponível em: < <https://www.jstor.org/stable/1175554>>. Acesso em: 2 jun., 2021.

- BRANDÃO, C.; REIS, L. P.; ROCHA, A. P. **Evaluation of Embodied Conversational Agents**. IEEE. 8th Information Systems and Technologies (CISTI), Iberian Conference on. 2013
- CALDARINI, G.; JAF, S.; MCGARRY, K. **A Literature Survey of Recent Advances in Chatbots**. Information. <https://doi.org/10.3390/info13010041>. 2022.
- CASAS, Jacky. et al. **Emotional Paraphrasing Using Pre-trained Language Models. International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)**. DOI: 10.1109/ACIIW52867.2021.9666309. 2021. Disponível em: <https://www.researchgate.net/publication/269400383_Autoria_Automatizada_de_Objeto_de_Aprendizagem_a_Partir_de_Captura_Multimidia_e_Associacao_a_Estilos_de_Aprendizagem>. Acesso em: jan. 2021.
- CHO, Y. H., LIM, K. Y. **Effectiveness of collaborative learning with 3D virtual worlds**. British Journal of Educational Technology, 48(1), 202-211. 2017.
- CLARK, L.M.H.; PANTIDI, N.; COONEY, O.; GARAIALDE, P.R.D.D.; Edwards, J.; Spillane, B.; Gilmartin, E.; Murad, C.; Munteanu, C. **What Makes a Good Conversation?: Challenges in Designing Truly Conversational Agents**. In Proceedings of the 2019 CHI Conference, Glasgow, UK. 2019.
- CLEMENTE, B. G. R. **Um ambiente para a construção de chatterbot com evolução dinâmica da base de conhecimento**. Dissertação. Disponível em encurtador.com.br/nuFN5. Acesso em 10 de fevereiro de 2020, 2016.
- COLBY, K. **Modeling a paranoid mind**, 1981. Disponível em: <<https://goo.gl/YJJuNo>>. Acesso em: 19 mai. 2017.
- CROWN, S. et al. **Anne G. Neering: Interactive chatbot to engage and motivate engineering students**. Computers in Education Journal, 21(2), 24-34. 2011
- DALE, R. The return of the chatbots. **Natural Language Engineering**, 22 (5): 811–817, Cambridge University Press 2016. doi:10.1017/S1351324916000243 811 I. Disponível em: <<https://www.cambridge.org/core/journals/natural-language-engineering/article/return-of-the-chatbots/0ACB73CB66134BFCA8C1D55D20BE6392>>. Acesso em: jan. 2022.
- DIAS, C. **Usabilidade na web: Criando portais mais acessíveis**. 2. ed. Rio de Janeiro: Alta Books, 2007.
- DILLENBOURG, Pierre. **What do you mean by collaborative learning?**. P. Dillenbourg. Collaborative-learning: Cognitive and Computational Approaches., Oxford: Elsevier, pp.1-19, 1999.
- DROUIN, M.; SPRECHER, S.; NICOLA, R.; PERKINS, T. **Is chatting with a sophisticated chatbot as good as chatting online or FTF with a stranger?** Computer in Human Behavior. 2022.
- FERREIRA, Thiagus. **ChatAval: Avaliador de chatbot voltado para à educação**. Trabalho de Conclusão de Curso. Graduação. Curso Sistemas de Informação. Gravataí-RS. 2021.
- FOLSTAD ET AL., A. et al. Future directions for chatbot research: an interdisciplinary research agenda. **Computing**, (2021) 103: 2915–2942. <https://doi.org/10.1007/s00607-021-01016-7>. Disponível em: <<https://link.springer.com/content/pdf/10.1007/s00607-021-01016-7.pdf>>. Acesso em: jan. 2022.

FRYER, Luke K., et al. **Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners.** Computers in Human Behavior, 2017.

GARTNER, L. G. **Hype Cycle for Digital Government Technology.** Disponível em <https://www.gartner.com/en/newsroom/press-releases/2019-08-28-gartner-2019-hype-cycle-shows-cloud-office-has-hit-ma>. Acesso em: dez. 2019.

GHOSE, S., BARUA, J. J. **Toward the implementation of a topic specific dialogue based natural language chatbot as an undergraduate advisor.**" Informatics, Electronics & Vision (ICIEV), 2013 International Conference on. IEEE. 2013.

GIL, A. C. **Métodos e Técnicas de Pesquisa Social.** 6.ed. São Paulo: Atlas, 2008.

GOH, O. S.; KUMAR, Y. J.; CHOON, N. H.; Leong, P. H. L.; Safar, M. **An Evaluation of the Conversational Agent System.** Springer-Verlag. Heidelber. DOI: 10.1007/978-3-662-49390-8)35. 2016.

GRAESSER, A. C.; CHIPMAN, P.; HAYNES, B. C.; Olney, A. **An Intelligent Tutoring System With Mixed-Initiative Dialogue.** IEEE Transaction on Education. DOI: 10.1109/TE.2005.856149. 2005.

GREIS, L. K.; REATEGUI, E. **Um simulador educacional para disciplina de física em mundos virtuais.** Renote. Disponível em <https://seer.ufrgs.br/renote/article/view/15220>. Acesso em fevereiro de 2020. 2010

GRIOL, D.; Molina, J. M. Do Human-Agent Conversations Resemble Human-Human Conversations? D12th Int. Conference, Advances in Intelligent Systems and Computing 373. DOI: 10.1007/978-3-319-19638-1_18. Springer. 2015.

HARSLEY, R., FOSSATI, D., DI EUGENIO, B., & GREEN, N. **Interactions of individual and pair programmers with an intelligent tutoring system for computer science. Paper presented at the Proceedings of the Conference on Integrating Technology into Computer Science Education, ITiCSE, 285-290.** doi:10.1145/3017680.3017786. 2017.

HAYASHI, Y. **Analyzing Lexical Expressions in an Human-Agent Online Explanation Task: Influence of Affect and Characteristics.** Department of Psychology, Ritsumeikan University Toji-in Kitamachi, Kita-ku, Kyoto. 2016.

HERPICH, F.; NUNES, F. B.; PETRI, G.; NICOLETE, P.; Tarouco, L. Modelo de Avaliação de abordagens Educacionais em realidade aumentada. Renote. p. 355-364. V. 17, DOI: 10.22456/1679-1916.95842. 2019.

HILL, J.; FORD, W. R.; FARRERAS, I. G. **Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations.** Computers in Human Behavior 49. p. 245-250. 2015.

HUSSAIN, S.; SIANAKI, O. A.; ABABNEH, N.. **A Survey on Conversational Agents/Chatbots Classification and Design Techniques.** DOI: 10.1007/978-3-030-15035-8_93. Springer Nature Switzerland. 2019.

ISO/IEC 25010 - **System and Software engineering - System and softwares Quality Requirements and Evaluation (SQuaRE)** - System and softwares quality models. Switzerland; 2011. Disponível em <https://iso25000.com/index.php/en/iso-25000-standards/iso-25010>. Acesso em outubro de 2020.

ISO/IEC 9126 - Engenharia de software - Qualidade de produto - Parte 1: Modelo de Qualidade - NBR ISO/IEC9126-1 DE 02/2013. Switzerland; 2011. Disponível em

<https://www.normas.com.br/visualizar/abnt-nbr-nm/22804/abnt-nbriso-iec9126-1-engenharia-de-software-qualidade-de-produto-parte-1-modelo-de-qualidade>. Acesso em outubro de 2020.

JACOB Jr, A. F., BARROS, F. A., FRANCÊS, C. R., & CWA, J. **Processo de Criação de um Modelo de Computação Afetiva para Chatterbots**. Anais do XXII SBIE - XVII WIE Aracaju, 2011.

KERLY, A., HALL, P., BULL, S.: **Bringing chatbots into education: Towards natural language negotiation of open learner models**. Know.-Based Syst. 20(2), 177–185, 2007.

KONZEN, A.; et al. (2011). **Maga Vitta – agente conversacional aplicado ao jogo educacional Città**. Disponível em: <https://goo.gl/gp5om6>. Acesso em 27 de abril, 2017.

KRASSMANN, A. L; HERPICH, F; SILVA, A. S. P; SILVA, A. R.; ABREU, C. S. Souza; SCHMITT, M.; BERCHT, M.; TAROUCO, L. M. R. **FastAIML: uma ferramenta para apoiar a geração de base de conhecimento para chatbots educacionais**. Renote – Revista novas tecnologias na educação, Porto Alegre, v. 15, n. 2, 2017.

KRASSMANN, Aliane Loureiro; FLACH, João Marcos. **AGATA – Automatic Generation of AIML from Text Acquisition**. Disponível em: <<http://agata.pgie.ufrgs.br/>>. Acesso em: 11 fevereiro 2018.

LEONHARDT, M.; DUTRA, R. L. S. D.; GRANVILLE, L. Z.; TAROUCO, L. M. R. **DOROTY: An extension in the architecture of a chatterbot for academic and professional training in the field of network management**. In: **IFIP World Conference on Computers in Education**, 2005, Stellenbosch - Cape Town. Conference on Computers in Education. Stellenboch: IFIP, v. 1, 2005.

LEONHARDT, Michelle Denise; CASTRO, Daiane Dorneles de; DUTRA, Renato Luís de Souza; TAROUCO, Liane Margarida Rockenbach. **ELEKTRA: Um Chatterbot para Uso em Ambiente Educacional**. **Novas Tecnologias na Educação**, v. 1, n. 2, Setembro, 2003.

MANFIO, E. R. et al. **Professor Tical e ALiB: Interação Humano Computador em Diferente Campo**. XIX Conferência Internacional sobre Informática na Educação, TISE. v. 1. 10. ISBN: 978-956-19-0889-5. 9-11, dez, 2014.

MAULDIN, M. L. **Chatterbots, tinymuds, and the turing test: Entering the loebner prize competition**. AAI. v. 94. 1994.

MAYER, Richard. E., MORENO, Roxana. **Nine Ways to Reduce Cognitive Load in Multimedia Learning**. Educational Psychologist, Lawrence Erlbaum Associates, Inc. 2003, 38(1), 43–52. Disponível em: <https://www.uky.edu/~gmswan3/544/9_ways_to_reduce_CL.pdf>. Acesso em: jan. 2022.

MAYER, Richard. **Multimedia Learning**. Cambridge: Cambridge University, 2001.

MCNEAL, M. L.; NEWYEAR, D. Chatbot creation options. **Library Technology Reports**, 49(8), pp. 11-17, 2013.

MCTEAR, Michael. **Conversational AI: Dialogue Systems, Conversational Agents e Chatbots**. Morgan & Claypool, Williston. Ulster University. <https://doi.org/10.2200/S01060ED1V01Y202010HLT048>. 2020.

METIS (2020). **METIS (Mediadora de Educação em Tecnologia Informática e Socializadora)**. Disponível em: <<http://avatar.cinted.ufrgs.br/METIS/>>. Acesso em: 15 mar. 2020.

- MOKMIN, N.A.M., IBRAHIM, N.A. **The evaluation of chatbot as a tool for health literacy education among undergraduate students.** *Educ Inf Technol* 26, 6033–6049. <https://doi.org/10.1007/s10639-021-10542-y>. 2021
- MORAES, S.; SANTOS, A.; REDECKER, M. **Prototipação de Chatterbots como Método de Aprendizagem.** *Anais do XXII Workshop de Informática na Escola*. DOI: <https://doi.org/10.5753/cbie.wie.2016.91>. 2016.
- MOU, Yi; XY, Kun. **The media inequality: Comparing the initial human-human and human-AI social interactions.** *Computers in Human Behavior*, 72: 432-440, 2017
- NICULESCU, Andrea Iona. et al. Design and Evaluation of a Conversational Agent for the Touristic Domain. Department of Human Language Technology, **Institute for Infocomm Research**, Singapore, 2014. Disponível em: <https://www.researchgate.net/publication/283101179_Design_and_evaluation_of_a_conversational_agent_for_the_touristic_domain>. Acesso em: jan. 2021.
- OKONKWO, C. W., and A. ADE-IBIJOLA. Chatbots Applications in Education: A Systematic Review. *Computers and Education: Artificial Intelligence*, v. 2, 2021. <https://doi.org/10.1016/j.caeai.2021.100033>. Disponível em: <<https://reader.elsevier.com/reader/sd/pii/S2666920X21000278?token=265DDAF6E1D072F0A9D882B9108625FD5C8E18EED55219D27FAE7A1DE276AD180D356F16CE9F9F0F39ACA2668E0920ED&originRegion=us-east-1&originCreation=20220612194512>>. Acesso em: jan. 2021.
- O'SHEA, J.; BANDAR, Z.; CROCKETT K. **Systems Engineering and Conversational Agents.** School of Computing, Mathematics and Digital Technology, Manchester Metropolitan University, Chester St., Manchester M15GD, United Kingdom. pp. 201–232. Springer. 2011.
- PASCHOAL, L. N.; CHICON, P. M. M.; FALKEMBACH, G. A. M. **Concepção, Implementação e Avaliação de uma Agente Conversacional com Suporte à Aprendizagem Ubíqua.** CINTED-UFRGS. Novas Tecnologias. 2017.
- PHILLIPS, C. **Chatbot Analytics 101: The Essential Metrics You Need to Track.** *Chatbots Magazine*. 2018. Disponível em <https://chatbotsmagazine.com/chatbot-analytics-101-e73ba7013f00>. Acesso em abril de 2020.
- PRESS, G. **AI Stats News: 62 % of US Consumers like using chatbots to interact with businesses.** *Forbes*. Disponível em: <https://www.forbes.com/sites/gilpress/2019/10/25/ai-stats-news-usconsumers-interest-in-using-chatbots-to-interact-with-businesses-rise-to-62/>. Acesso em março de 2019.
- PRODANOV, Cleber Cristiano. **Metodologia do trabalho científico [recurso eletrônico] : métodos e técnicas da pesquisa e do trabalho acadêmico / Cleber Cristiano Prodanov, Ernani Cesar de Freitas. – 2. ed. – Novo Hamburgo: Feevale, 2013.**
- RADZIWILL, N. M.; BENTON, M. C. **Evaluating Quality of Chatbots and Intelligent Conversational Agents.** arXiv preprint arXiv:1704.04579, 2017.
- RAJ, Joseph; ANTON, Joliz; RAJ, Johnson P. Durai. **Detection of Recovery of Covid-19 Cases using Machine Learning.** *International Journal of Current Research and Review*. Scopus. DOI: <http://dx.doi.org/10.31782/IJCRR.2021.SP183>. 2021.
- ROCHA, R., Leite, A.; COUTINHO, F e ARAUJO, C. **Avaliação de objetos de aprendizagem sobre o sistema digestório com base nos princípios da Teoria Cognitiva de Aprendizagem.** *Revista Ciência e Educação, Bauru*, v. 20, n. 4, p. 1003-1017. 2014. Disponível em <http://www.scielo.br/pdf/ciedu/v20n4/1516-7313-ciedu-20-04-1003.pdf>

ROGERS, C. **Client-Centered Therapy: Its Current Practice, Implications and Theory**. Houghton Mifflin, Boston, 1951.

SANTANA, A. O. Geração de Ambientes Virtuais de Ensino-Aprediagens para Cursos Baseados na Realização de Projetos. Dissertação. UFRGN. Departamento de Informática, 2017.

SANTOS, P. S.; ROLAND, L. A.; SILVA, D. R. Recomendação de Aplicativos Educacionais Mediada por Interfaces Conversacionais. VII Congresso Brasileiro de Informática na Educação (CBIE 2019). **Anais do XXV Workshop de Informática na Escola (WIE 2019)**.

SATU, Shahriare; PARVEZ, Hasnat; MAMUM, Shamim. **Review of integrated applications with AIML based chatbot**. 1 st International Conference on Computer & Information Engineering. 2015. Disponível em <https://goo.gl/M2QFop>. Acesso em 25 de agosto de 2017.

SEDOC, J.; IPPOLITO, D.; ARUN K.; THIRANI, J.; UNGAR, L.; BURCH, C. C. **ChatEval: A Tool for Chatbot Evaluation**. Anais da Conferência de 2019 do Capítulo Norte-Americano da Association for Computational Linguistics (Demonstrations). DOI 10.18653 / v1 / N19-4011. 2019.

SGOBBI, Fabiana S.; TAROUCO, Liane M. R.; REATEGUI, Eliseu. Mundo virtual 3D e Internet das Coisas para motivar mudança de comportamento saudável. **Revista Iberoamericana de Educación en Tecnología y Tecnología en Educación**, n. 19. p 7-15, Jun., 2017. Disponível em: <http://www.scielo.org.ar/scielo.php?script=sci_abstract&pid=S1850-99592017000100002>. Acesso em: jan. 2021.

SGOBBI, Fabiana S.; HERPICH, F.; NUNES, F. B. As ferramentas de autoria para criar um mundo Virtual Imersivo. In: SGOBBI, Fabiana S.; HERPICH, F.; NUNES, F. B. (Orgs.). **Cognição e aprendizagem em Mundo Virtual Imersivo**. Porto Alegre: URGS, 2019, p. 42-94. Disponível em: <<http://penta3.ufrgs.br/AVATAR/Livro/AVATAR-Cogni%C3%A7%C3%A3o-Aprendizagem-MundoVirtual-2018.pdf>>. Acesso em: 16 mar. 2020.

SILVEIRA, C. Easy AIML - Gerador de códigos AIML. **Revista CNEC Educação**, v. 1, n. 1, 2018. Disponível em: <<http://sys.facos.edu.br/ojs/index.php/cneceduacao/article/view/281>>. Acesso em: fev. 2020.

TIBOLA, L.; VOSS, G.; AVILA, B., TAROUCO, L., & SGOBBI, F. **Virtual laboratory for promoting engagement and complex learning**. In E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (pp. 1933-1938). Association for the Advancement of Computing in Education (AACE), 2014.

VYGOTSKY, L.S. Mind in society: The development of higher psychological processes. (1978) . Harvard University Press Cambridge, MA

WALLACE, R.S.; TOMABECHI, H., Aimless, D.: **Chatterbots Go Native: Considerations for an eco-system fostering the development of artificial life forms in a human world** (2003), <http://www.pandorabots.com/pandora/pics/chatterbotsgonative.doc> (accessed February 2021)

WALLACE, Richard S. **The Elementes of AIML Style**. 2003 ALICE A. I. Foundation, Inc. Disponível em <http://www.alicebot.org/style.pdf>. Acesso em 25 de fevereiro de 2020.

WEIZENBAUM, J. ELIZA. **A computer program for the study of natural language communication between man and machine**. Communications of the ACM, v. 9, n. 1, p. 3644, jan. 1966.

WOLLNY, Sebastian et al. Are We There Yet? A Systematic Literature Review on Chatbots in Education. **Front. Artif. Intelligence**. 2021. <https://doi.org/10.3389/frai.2021.654924>. Disponível em: <<https://www.frontiersin.org/articles/10.3389/frai.2021.654924/full>>. Acesso em: jan. 2021.

YU, Z.; XU, Z.; BLACK, A. W.; Rudnicky, A. I. **Chatbot Evaluation and Database Expansion via Crowdsourcing**. Carnegie Mellon University. 2016.

ZDRAVKOVA, K. **Conceptual framework for an intelligent chatterbot**. In: Proceedings of the 22nd International Conference on Information Technology Interfaces, p. 189-194, 2000.

APÊNDICES

APÊNDICE A - PUBLICAÇÕES

Algumas produções científicas de relevância elaboradas durante o doutorado do autor desta tese (autoria e coautoria):

Artigos completos publicados em periódicos:

1. SILVEIRA, Clóvis da; DA SILVA, Anita Raquel; HERPICH, Fabrício; TAROUCO, Liane Margarida Rockenbach . Uso de Agente conversacional como recurso de aprendizagem socioeducacional. *RENOTE. Revista Novas Tecnologias na Educação*, v. 17, p. 668-678, 2019.
2. KRASSMANN, Aliane Loureiro; PAZ, Fábio Josende; SILVEIRA, Clóvis da; TAROUCO, Liane Margarida Rockenbach; BERCHT, Magda. Conversational Agents in Distance Education: Comparing Mood States with Students? Perception. *Creative Education*, v. 9, p. 1726-1742, 2018.
3. SILVEIRA, Clóvis da; SALADA, J. Easy AIML - Gerador de códigos AIML. *Revista CNEC Educação*, v. 1, p. 86-99, 2018.
4. PAZ, F. J.; KRASSMANN, A.; SILVEIRA, Clóvis da; TAROUCO, L. M. R. Perspectivas tecnológicas para o aprimoramento de chatbots educacionais em AIML. *TE&ET Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, v. 20, p. 7-15, 2017.
5. MACHADO, L. R.; MENDES, J. S. S.; KRIMBERG, L.; SILVEIRA, Clóvis; BEHAR, P. A. Competência digital de idosos: mapeamento e avaliação. *ETD: Educação Temática Digital*, v. 21, p. 941-959, 2019.

Capítulos de livros publicados:

6. SILVA, A. R.; SILVEIRA, Clóvis da; TAROUCO, Liane Margarida Rockenbach. Agentes conversacionais no mundo virtual imersivo. In: TAROUCO, Liane Margarida Rockenbach; SILVA, Patrícia Fernanda, HERPICH, Fabrício. (Org.). **Cognição e Aprendizagem em Mundo Virtual Imersivo**. Porto Alegre: UFRGS, 2019, v. 1, p. 95-121.
7. TAROUCO, Liane M. R.; SILVEIRA, Clóvis da; KRASSMANN, Aliane L. Collaborative Learning with Virtual Entities. **Lecture Notes in Computer Science**. Porto Alegre: Springer International Publishing, 2018, v., p. 480-493.
8. TAROUCO, Liane M. R.; SILVEIRA, Clóvis; KRASSMANN, Aliane L. Collaborative Learning with Virtual Entities. **Lecture Notes in Computer Science**. Porto Alegre: Springer International Publishing, 2018, v., p. 480-493.

Artigos aguardando parecer:

9. DESCOVI, L. M. G.; HERPICH, F.; PHILERENO, J. A.; SILVEIRA, C.; TAROUCO, L. M. R. Avaliação da usabilidade do Chatbot IBM Watson Agente Conversacional Cerebrum de Equações Algébricas de Primeiro Grau com uma Variável. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología.
10. SILVEIRA, C.; SILVA, A.; HERPICH, F.; TAROUCO; HYGIA, L. M. R.: A conversational agent as support for the context of the covid-19 pandemic in Brazil. Revista Tecne.

APÊNDICE B – PERFIL DO PÚBLICO DA FASE 1 DA PESQUISA

O objetivo da tese consiste em criar um *framework* com sugestões de métricas de avaliação de *software*, a fim de elencar o que um agente conversacional educacional pode e deve ter para ser relevante, com vistas a ser utilizado como referência para a avaliação e construção de *chatbots* desenvolvidos para o contexto educacional.

O objetivo deste questionário é conhecer um pouco mais de você, aluno do Curso superior de Tecnologia em Sistemas para Internet, e saber o que conhece sobre agentes conversacionais.

Por favor, preencha os dados abaixo:

Questões gerais:

1. Nome:
2. Idade:
3. Série:

Experiência com o computador:

- Você tem computador em sua casa?
- Você tem acesso à Internet?
- Quais meios você usa para acessar a Internet (celular, *tablet*, computador, *notebook*)?
- Onde você acessa a Internet (sua casa, casa do amigo, *lan house*, casa de amigos ou familiares)?
- Quantas horas por dia você fica conectado à Internet?
- No que passa maior tempo utilizando a Internet (redes sociais, *blogs*, navegando)?
- Atualmente, quando acessa a Internet, você a utiliza mais para o seu lazer e entretenimento ou para aprender algum conteúdo de seu interesse?

Marque a alternativa de acordo com o uso de cada ferramenta

METIS:

- () desconheço () já ouvi falar, mas nunca o utilizei
() criei meu *login*, mas não utilizo () já utilizei () utilizo com frequência
() uso para minha diversão () uso para meus estudos () ambos

ATENA:

- () desconheço () já ouvi falar, mas nunca o utilizei
 () criei meu *login*, mas não utilizo () já utilizei () utilizo com frequência
 () uso para minha diversão () uso para meus estudos () ambos

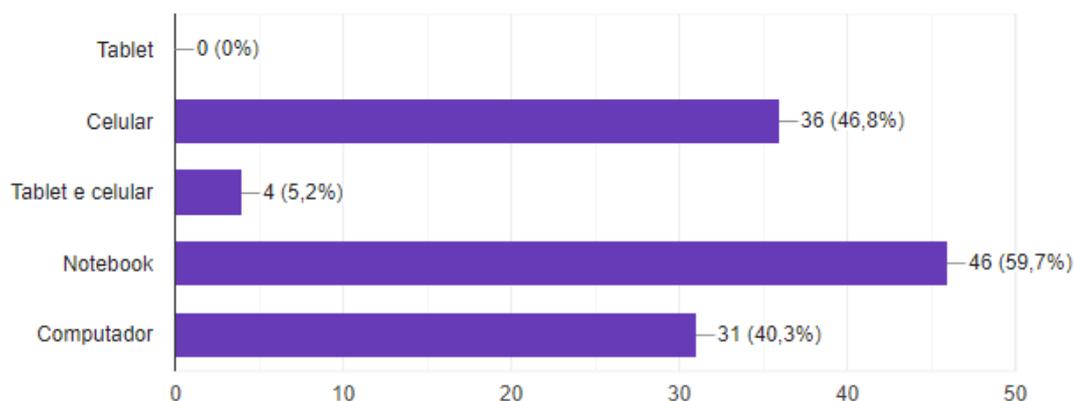
ALTEIA:

- () desconheço () já ouvi falar, mas nunca o utilizei
 () criei meu *login*, mas não utilizo () já utilizei () utilizo com frequência
 () uso para minha diversão () uso para meus estudos () ambos

A seguir, são apresentadas as informações iniciais para que possamos saber quanto conhece das ferramentas que serão utilizadas na pesquisa.

Assim, procurou-se questionar os alunos sobre os dispositivos que usam para acessar as aulas on-line, conforme o gráfico abaixo:

Quais os dispositivos você usa para as aulas on-line?

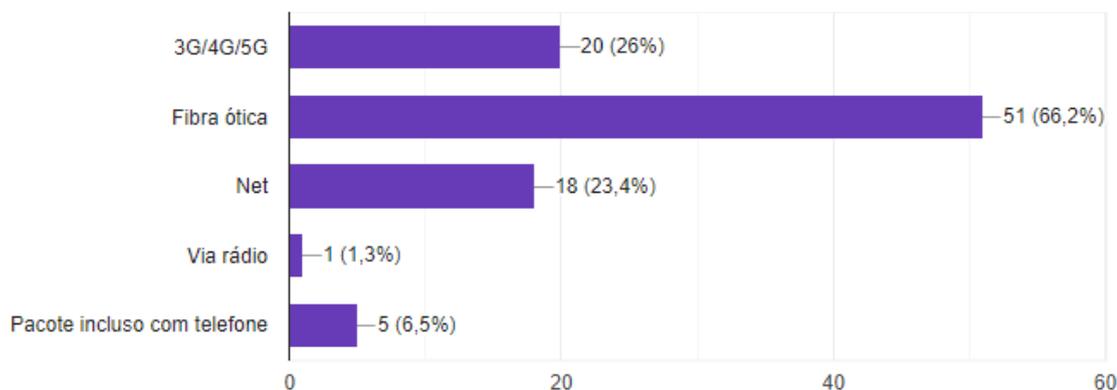


Fonte: Elaborado pelo autor (2021).

Nesta questão, os alunos puderam marcar mais de um dispositivo, que é usado para acessar as aulas on-line. Observa-se que o *notebook*, seguido do celular, compõem os dispositivos mais usados para acessar as aulas on-line. Além desses meios, 31 alunos informaram o computador como dispositivo para acessar as aulas. Essa informação foi bem importante para organizar os conteúdos e as possibilidades de instalações. Constatou-se que todos eles têm acesso a dispositivos para acessar as aulas. Curiosamente, essa geração, que varia entre 15 e 17 anos, não usa o *tablet* como dispositivo para acessar as aulas on-line.

Os alunos também foram questionados sobre o meio de conexão da internet mais utilizado por eles. As respostas foram registradas no gráfico a seguir:

Meio de conexão da internet dos estudantes:

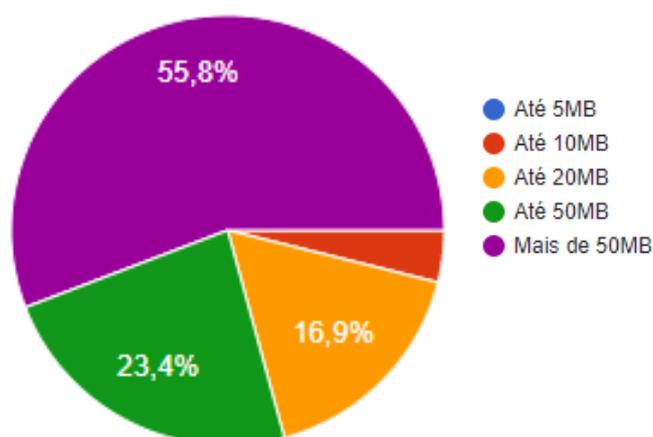


Fonte: Elaborado pelo autor (2021).

No gráfico, constata-se que 66,2% dos estudantes têm acesso à internet via fibra ótica, e 26% deles têm acesso à internet 3G/4G/5G. Nesta questão, eles podiam assinalar todos os meios de acesso à sua internet. Essa pergunta foi importante para entender o meio de conexão com a internet, porque, no período do experimento com os alunos, estavam assistindo às aulas no formato híbrido.

Outro ponto importante foi entender a velocidade da internet disponível nas residências dos estudantes. Por meio do gráfico, a seguir, constata-se que mais da metade da turma tem internet a uma velocidade de 50MB ou mais.

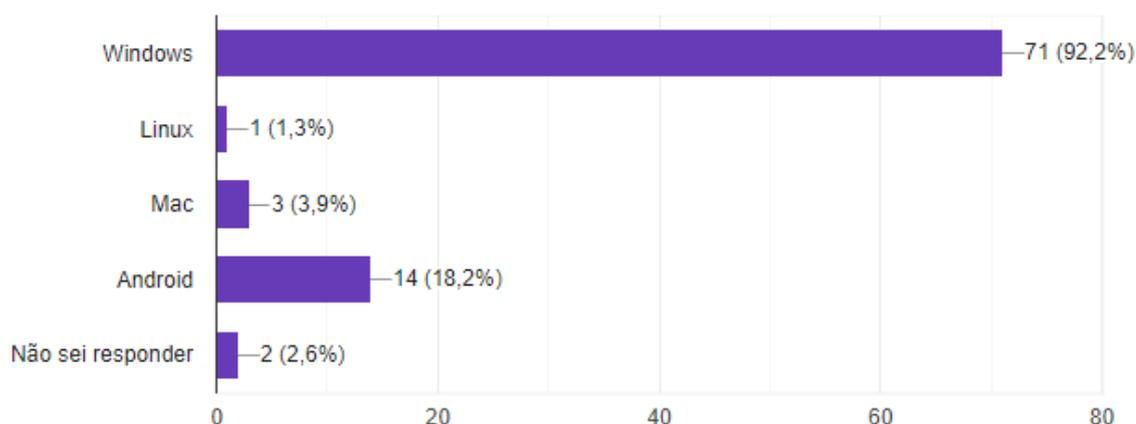
Velocidade da internet na residência dos alunos



Fonte: Elaborado pelo autor (2021).

Questionou-se, também, qual o Sistema Operacional que os estudantes estavam familiarizados, e obteve-se o seguinte resultado:

Sistemas Operacionais utilizados pelos alunos



Fonte: Elaborado pelo autor (2021).

Ter conhecimento do Sistema Operacional que os alunos utilizam é relevante para a escolha de alguns aplicativos utilizados para a construção dos códigos, como é o caso do *Visual Studio Code* usado para escrever os códigos na linguagem HTML, AIML e CSS. Ainda conforme o gráfico anterior, não seria possível deixar os 3 alunos que usam o Sistema Operacional MAC com aplicativos em que fosse permitida a instalação apenas para a *Microsoft*.

Em razão de metade dos alunos estarem em suas residências enquanto outro grupo permanecia nas aulas presenciais, buscou-se saber se tinham um ambiente bom de estudos para acompanhar as aulas on-line. Explicou-se que se considera um bom ambiente para estudo um espaço específico em suas residências, onde não sejam interrompidos por outras pessoas que moram com eles.

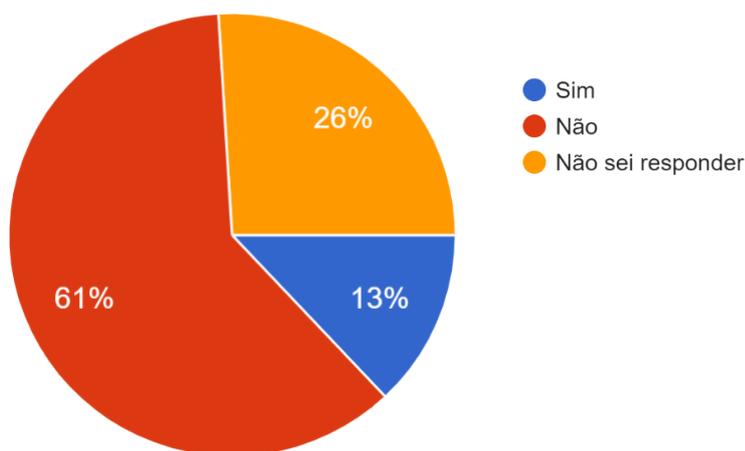
Quanto ao ambiente de estudos das aulas síncronas



Fonte: Elaborado pelo autor (2021).

Partindo-se do princípio de que os alunos estão em um curso Técnico em Informática para Internet - e que o desenvolvimento de códigos é algo que deverá acompanhá-los durante o Ensino Médio -, questionou-se se conheciam algo sobre algoritmos. As respostas dos estudantes estão contempladas no gráfico que segue:

Você conhece algo sobre algoritmos?

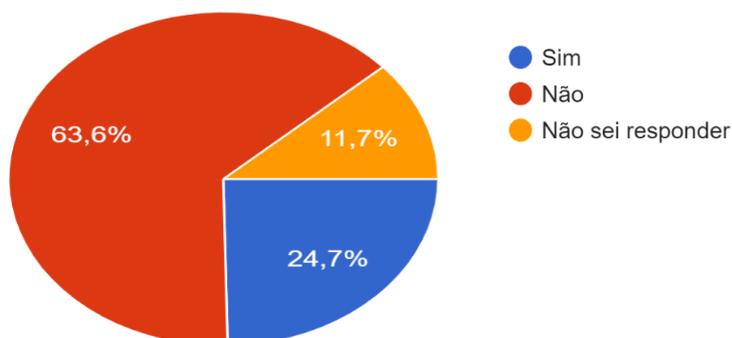


Fonte: Elaborado pelo autor (2021).

Conforme este gráfico, constata-se que os alunos não têm conhecimento sobre algoritmos, visto que apenas 13% deles conhecem algo sobre algoritmos, ou seja, dos 77 alunos de ambas as turmas que participam da pesquisa, apenas 10 afirmaram conhecer. Assim, identificou-se um público com pouca experiência em lógica de programação.

Na sequência, buscou-se entender se os alunos conheciam alguma linguagem de programação e obteve-se as seguintes respostas:

Você conhece alguma linguagem de programação?

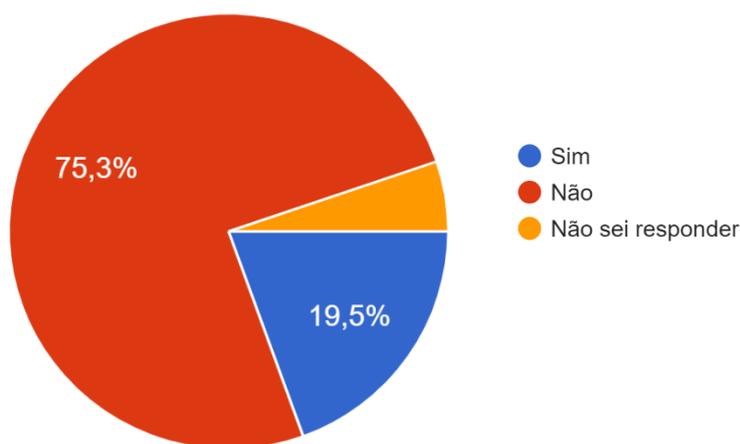


Fonte: Elaborado pelo autor (2021).

Embora grande parte dos alunos não tivesse conhecimento sobre uma linguagem de programação, ou não sabia responder se conhecia sobre linguagem de programação, alguns sinalizaram conhecer alguma linguagem, informando que já tinham desenvolvido alguns códigos mais básicos.

Ao se questionar se os estudantes já haviam programado, obtiveram-se os seguintes resultados, conforme está demonstrado no gráfico a seguir:

Você já programou?

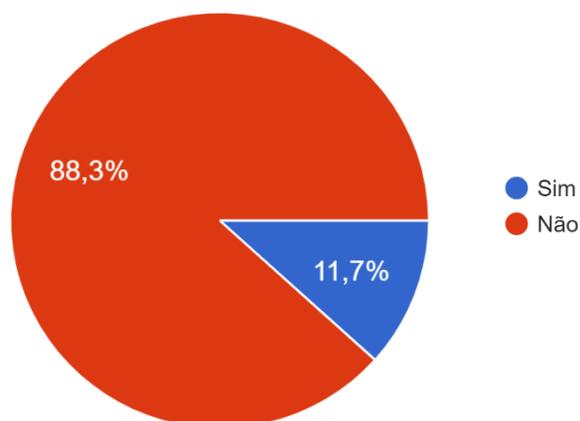


Fonte: Elaborado pelo autor (2021).

Constata-se que alguns alunos já haviam realizado o desenvolvimento de trechos de pequenos códigos.

Por fim, os alunos foram questionados se já tinham desenvolvido algum *site*. Nesse caso, chegou-se ao seguinte resultado:

Você já desenvolveu algum site?



Fonte: Elaborado pelo autor (2021).

Na pergunta aberta: “Quais suas expectativas com o Curso Técnico em Informática para Internet?”, houve uma diversidade de respostas; desde respostas com duas palavras, até frases mais elaboradas. No entanto, destaca-se que todos os registros sinalizaram expectativas positivas em relação à aprendizagem no curso Técnico em Informática.

Vale destacar que, embora grande parte dos alunos tenha informado que não tinha contato com programação, muitos têm interesse em aprender, conforme demonstram algumas respostas que registram, na tabela que segue, essas expectativas dos alunos.

Tabela com as Respostas abertas de alunos referentes às expectativas dos mesmos

Participante	Transcrição das respostas dos alunos
ID A39	“Espero aprender bastante sobre programação e informática e sair do ensino médio com emprego.”
ID B28	“Espero aprender muitas coisas sobre informática e me preparar para o futuro.”
ID A08	“Por ser uma área que não tenho domínio, mas tenho interesse, tenho como expectativa que o curso acrescente na minha vida e que, de certa forma, me auxilie na escolha de profissão.”
ID B20	“Tenho grandes expectativas, principalmente, na parte de programação, algoritmos etc.”
ID A34	“Estou bastante animada em relação a esse curso, pois sei muito pouco sobre esse assunto e quero muito aprender mais. Minhas expectativas com o curso técnico em informática é para que eu consiga aprender bastante coisas novas.”
ID A23	“Minhas expectativas são que eu consiga entender melhor como funciona o uso do computador, celular, para coisas que, no momento, eu não sei o que abrange quase tudo sobre informática, além de que o que eu aprender com o curso, eu possa usar futuramente na minha carreira.”
ID A15	“Aprender sobre programação <i>web</i> para no futuro trabalhar na área.”
ID B25	“Muito boas, não sei muito, porém quero aprender pelo menos o básico.”
ID B15	“Muito altas, acho que vou estar muito mais preparado para o futuro e vai ser muito bom para mim também.”
ID A33	“Eu espero que eu consiga aprender mais e realmente me aprofundar nos programas, para que eu atinja a maestria e consiga entender mais sobre este assunto que eu sempre tive interesse.”
ID B27	“Espero sair com uma boa base em tecnologia, creio que, independentemente do curso que irei escolher na faculdade, vou utilizar bastante o conhecimento que vou adquirir ao longo desses três anos.”
ID B11	“Aprender a programar, pois eu acho bem interessante tudo sobre programação.”
ID B24	“Minha expectativa é conhecer esse mundo novo para mim, e aprender o máximo que conseguir, pois sei que vou usar a tecnologia para absolutamente tudo (ainda mais no futuro).”
ID A01	“Gostaria de aprender algo que me permitisse trabalhar em outro país, usando meus conhecimentos em inglês.”
ID B30	“Minhas expectativas são muito boas, até porque eu não tenho total domínio na área de informática para internet e hoje em dia é algo que é necessário saber e entender, pois faz parte da nossa rotina e vai ser algo obrigatório no futuro.”
ID A31	“Como não entendo nada dessa questão de programação, informática, essas coisas, a minha expectativa é aprender bastante sobre isso.”
ID A24	“Estou bem motivada, pois é uma proposta muito interessante. Espero que eu consiga alcançar o que desejo e que aulas sejam dinâmicas e super aproveitadas.”
ID B16	“Tenho muita curiosidade de aprender sobre, e espero conseguir entender o conteúdo que vai ser passado.”
ID A14	“Minhas expectativas são ótimas, acredito que vou me adaptar e aprender muito.”

ID A35	“Minha expectativa é de sair do ensino médio com conhecimento suficiente para sair com emprego.”
ID A18	“São bastante. Sempre amei o mundo da tecnologia e está sendo uma grande realização, principalmente por ser no Senac. Quero que sejam bem dinâmicas e aulas animadas.”
ID B12	“Eu espero conseguir aprender muito sobre a linguagem da programação e, pelo menos, um pouco sobre programação de jogos.”
ID A22	“Estou com uma expectativa bem grande, porque gostaria de aprender mais sobre a matéria, para que, no futuro, eu possa trabalhar com isso.”
ID A19	“Aprender muito sobre programação, desenvolver bons sites, apps e softwares, além de me divertir bastante.”
ID B29	“Eu espero conseguir aprender várias coisas do curso esse ano, pois eu realmente quero aprender sobre as coisas dessa área, e eu tenho expectativas altas sobre a matéria e tenho expectativas altas sobre mim aprendendo elas”.

Legenda: Identificador do Participante - ID.

Fonte: Elaborado pelo autor (2021).

Esta tabela registra parte das respostas abertas referentes às expectativas dos alunos. As demais respostas não apresentadas na tabela contêm textos muito parecidos com as respostas transcritas. Enfatiza-se que houve registros de expectativas negativas ou de falta de interesse em relação ao curso.

Devido aos alunos terem pouca experiência em relação ao mundo de lógica de programação, construção de *sites* ou desenvolvimento, elaborou-se uma nuvem de palavras para constatar o interesse de uma forma geral, a partir da fala dos alunos, conforme apresentado na figura abaixo.

Expectativas dos alunos com o Curso Técnico



Fonte: produzido pelo autor na plataforma Voyant Tools (2019).

Diante do exposto, conclui-se que os alunos do Curso Técnico em Informática para Internet compõem um público com boas condições de acesso a recursos de computadores, com boa velocidade de internet para navegar, e que a grande maioria conta com local propício para os estudos.

APÊNDICE C – PERGUNTAS PARA ANÁLISE DAS MÉTRICAS – FASE 2.2

Adaptação dos Questionamentos da NBR ISO/IEC 25010:

- 1) É possível fazer testes e validações no agente conversacional antes do uso da versão final ou versão final?
- 2) O agente conversacional informa diagnóstico de falhas para o usuário?
- 3) O agente conversacional possui eficiência para abrir, fechar e carregar as telas de diálogos e de respostas, tendo em vista a configuração do seu computador e velocidade de sua internet?
- 4) O agente conversacional apresenta um bom acesso em relação ao tempo para trazer a resposta durante os diálogos?
- 5) Enquanto você utiliza o agente conversacional você o considera atrativo?
- 6) O agente conversacional possibilita que você consiga operá-lo intuitivamente?
- 7) O agente conversacional apresenta características de fácil compreensão e pode ser aprendido e operado facilmente?
- 8) Caso o servidor ou computador apresente problemas de travamento, os dados permanecem quando o serviço é restabelecido?
- 9) O agente conversacional retorna respostas de acordo com os diálogos e de forma correta, considerando o conteúdo digitado?
- 10) O agente conversacional continua funcionando enquanto algumas alterações na base de conhecimento estão sendo implementadas?
- 11) O agente conversacional possibilita funções apropriadas para o usuário, tais como botões, caixas de diálogos?

APÊNDICE D – TERMO DE CONSENTIMENTO LIVRE ESCLARECIDO – FASE 1 (PAIS E RESPONSÁVEIS DOS ESTUDANTES DO ENSINO MÉDIO)

Pelo presente, convido-o a participar da pesquisa intitulada: **“Estratégias de Avaliação de Agentes Conversacionais Educacionais”**. Esta pesquisa faz parte da Tese de Doutorado, desenvolvida no programa de Pós-Graduação em Informática na Educação (PPGIE), do Centro Interdisciplinar de Novas Tecnologias na Educação (CINTED), da Universidade Federal do Rio Grande do Sul (UFRGS).

A pesquisa tem como objetivo criar um *framework* com sugestões de métricas de avaliação de *software*, a fim de elencar o que um agente conversacional educacional pode e deve ter para ser relevante, com vistas a ser utilizado como referência para a avaliação e construção de *chatbots* desenvolvidos para o contexto educacional. Ela será aplicada no ano de 2021, remotamente, durante as aulas do Técnico em Informática para Internet com os alunos das Turmas 1A e 1B, oferecidas pela Instituição SENAC São Leopoldo/RS, e ministrada pelo professor Clóvis da Silveira. Destaca-se que o referido docente e pesquisador é o mesmo professor da disciplina do Curso Técnico em Informática para Internet.

Algumas dessas ações poderão ser fotografadas para possíveis visualizações futuras e acervo documental. Ressalta-se que a identidade dos participantes nas imagens será mantida em sigilo, assim como todos os instrumentos a serem aplicados, servindo apenas para fins de pesquisa.

As aulas ocorrerão normalmente para todos os alunos no período da manhã, alterando-se somente a forma de interação com o conteúdo. Serão promovidas atividades extras e de caráter lúdico pelo professor e, nesses momentos, os alunos usarão um aplicativo para dispositivo móvel ou *desktop*.

Cabe informar que a sua participação na pesquisa não oferece risco algum. Caso seja verificado algum constrangimento durante os encontros, o pesquisador irá intervir, direcionando o assunto tratado. Também lhe é garantido o que segue:

- Receber resposta a qualquer pergunta, ou esclarecimento a qualquer dúvida acerca dos procedimentos, riscos, benefícios e outros assuntos relacionados com a pesquisa;
- Poder retirar seu consentimento a qualquer momento, deixando de participar do estudo, sem que isso traga qualquer tipo de prejuízo;
- Não ser identificado(a), quando da divulgação dos resultados, sendo todas as informações obtidas apenas para fins científicos vinculados à pesquisa;

- No caso de gastos adicionais, estes serão absorvidos pelo orçamento da pesquisa.

Este documento deverá ser assinado em duas vias, sendo que uma delas será retida pelo sujeito da pesquisa e a outra pelo pesquisador. O responsável pela pesquisa é o aluno de doutorado Clóvis da Silveira e seu e-mail é: csclovis@gmail.com.

Pelo presente Termo de Consentimento Livre e Esclarecido, declaro que autorizo a participação de meu responsabilizado nesta pesquisa, pois fui devidamente informado(a), de forma clara e detalhada, livre de qualquer constrangimento e coerção, dos objetivos e da justificativa do estudo, dos instrumentos de coletas de informação que serão utilizados, dos riscos e benefícios, conforme já citados neste termo.

<data>

Nome do aluno:

Nome do responsável:

Data de Nascimento:

E-mail:

Assinatura do responsável:

Pesquisador: Clóvis da Silveira

Assinatura:

APÊNDICE E - SOLICITAÇÃO DE DESENVOLVIMENTO DE PESQUISA FECOMÉRCIO-RS/SESC/SENAC

Para que a sua solicitação seja analisada, todos os dados abaixo devem ser informados. Após a análise, o resultado será comunicado, por e-mail ou telefone, pela unidade/escola ou departamento onde foi registrada a solicitação.

1 DADOS PARA ANÁLISE DE DESENVOLVIMENTO DE PESQUISA NO SISTEMA FECOMÉRCIO-RS/SESC/SENAC

1.1 DADOS DE IDENTIFICAÇÃO

1.1.1 Nome Completo do pesquisador: Clóvis da Silveira

1.1.2 E-mail e telefone de contato: cdsilveira@senacrs.com.br | 51 984632968

1.1.3 Instituição de origem do aluno/pesquisador: Universidade Federal do Rio Grande do Sul (UFRGS)

1.1.4 Curso do aluno/pesquisador: Programa de Pós-Graduação em Informática na Educação (PPGIE)

1.1.5 Nome completo do Professor Orientador: Liane Margarida Rockenbach Tarouco

1.1.6 E-mail e telefone de contato do professor orientador: liane@penta.ufrgs.br | 51 99981-5577

1.2 DADOS DA PESQUISA

1.2.1 Tema da pesquisa: Avaliação de Agentes Conversacionais Educacionais

1.2.2 Objetivo da pesquisa: Apresentar recursos computacionais de três agentes conversacionais educacionais aos alunos do Técnico em Informática para Internet Integrado ao Ensino Médio - Senac São Leopoldo/RS, a fim de registrar as percepções dos estudantes, em relação ao uso e funcionalidades dos respectivos agentes.

1.2.3 Metodologia: O pesquisador buscará saber com os alunos se conhecem algum tipo de agente conversacional educacional. Parte da hipótese de que conhecem pouco sobre agentes conversacionais ou não conhecem. Com base nessa informação, será abordado o tema.

Na sequência, serão apresentados os agentes Conversacionais abordados na pesquisa. Destaca-se que os três agentes conversacionais são desenvolvidos pelo Programa de Pós-Graduação de Informática na Educação, da Universidade Federal do Rio Grande do Sul, para fins educacionais. Sobre os agentes METIS, ATENA e ALTEIA:

A Agente conversacional METIS (Mediadora de Educação em Tecnologia Informática e Socializadora) foi desenvolvida com a finalidade de dialogar com os alunos, por meio de uma interface que simula uma conversação on-line (*chat*) entre pessoas. A implementação da METIS é baseada no sistema A.L.I.C.E., desenvolvido por Wallace (2003), que organiza a base de conhecimento, notação baseada em XML (*Extensible Markup Language*), AIML. Foi desenvolvida utilizando o *software Open Source Program-O* e possui uma interface em *Cascading Style Sheets* (CSS), o que permite que a tela seja ajustada de acordo com o dispositivo, facilitando que estudantes interajam até mesmo por um aparelho celular.

A base de dados conversacionais da METIS foi construída a partir de perguntas e respostas elaboradas por especialistas na área da educação, mídias e tecnologias. A base de conhecimento do sistema foi alimentada por especialistas em mídias e informática na educação, possibilitando um conhecimento mínimo dos conteúdos e retorno positivo às dúvidas apresentadas pelos alunos.

A METIS²⁷ possui, na sua base de conhecimento, um conjunto de informações relacionadas com tecnologias educacionais.

A Agente Tutor para Ensino e Navegação no Ambiente – ATENA²⁸ foi desenvolvida para conversação na área de física. Também possui capacidade de diálogo, além de acompanhar o estudante em seu percurso no cenário imersivo virtual.

A Agente Lexical de apoio a Tarefas Educacionais – ALTEIA²⁹ é um *chatbot* especialista em Metodologias Ativas, Sala de Aula Invertida e Ferramentas de Autoria. Tem capacidade de entregar respostas aos estudantes em relação a conceitos, além de responder a diversas dúvidas relacionadas à sua especialidade.

Após a exploração dos agentes conversacionais, será aplicado um questionário de avaliação de *software*, no qual o aluno responderá sobre suas percepções de uso da ferramenta,

²⁷ A METIS pode ser acessada por meio do site: <http://avatar.cinted.ufrgs.br/metis/>.

²⁸ A agente ATENA pode ser acessada em <http://avatar.cinted.ufrgs.br/atenaweb/gui/jquery/?>

²⁹ A agente ALTEIA pode ser acessada em <http://avatar.cinted.ufrgs.br/alteia/gui/plain/index.php#end>

no que se refere à funcionalidade, à navegabilidade, aos erros e ao entendimento das respostas. Será questionado, ainda, se os agentes ajudaram na aprendizagem.

No contexto do curso técnico, a parte de exploração das ferramentas será muito bem-vinda e totalmente compatível com a lógica de programação e a linguagem HTML, as quais serão abordadas ao longo do primeiro ano do ensino médio.

Salienta-se que tal pesquisa está alinhada com o perfil do egresso. De acordo com o Projeto Político Pedagógico de São Leopoldo:

O formando do Ensino Médio Técnico em Informática para Internet é um jovem consciente de sua atuação, interagindo e agindo de maneira colaborativa com sujeitos de diferentes contextos sociais e culturais. É pesquisador, comprometido com as transformações da sociedade, flexível às novas formas de ser e estar no mundo do trabalho.

Outro ponto de alinhamento do Curso Técnico com a ideia da pesquisa, de acordo com o Projeto Político Pedagógico de São Leopoldo:

Trata-se de um profissional cidadão proativo, ético e socialmente comprometido com as exigências atuais em relação à convivência social e laboral, capaz de pensar criticamente e com condições de prosseguimento de estudos posteriores.

Ainda conforme o Projeto Político Pedagógico, essa pesquisa vai possibilitar aos alunos a utilização e exploração dos agentes conversacionais, a fim de terem uma prática com os critérios de usabilidade e acessibilidade, conforme espera-se do perfil do egresso:

É o profissional que desenvolve aplicações para internet (*Web* e *Web Mobile*), seguindo as normas e especificações da lógica e das linguagens de programação. Utiliza ferramentas de auxílio na análise e desenvolvimento de sistemas para construção de soluções que auxiliam o processo de criação de interfaces e aplicativos, seguindo os critérios de usabilidade e acessibilidade na interação com o usuário final.

O experimento envolve a análise de *softwares*, categorizados com agentes conversacionais. Os agentes conversacionais são desenvolvidos na linguagem *Artificial Intelligence Markup Language* (AIML), que é muito semelhante à linguagem *HyperText Markup Language* (HTML), a qual será utilizada com os alunos do Curso Técnico em Informática e tem aderência com a proposta do curso Técnico em Informática.

Por fim, trata-se de uma fase prévia que visa realizar um experimento com os agentes e coletar as funcionalidades e percepções dos alunos em relação a esses agentes. Salienta-se que não há nenhum tipo de impacto para os participantes na coleta de dados em virtude da fase de validação dos agentes. Os dados serão coletados, com base nas percepções dos alunos em

relação às funcionalidades dos agentes. Essas funcionalidades envolvem percepções como funcionalidade, confiabilidade, usabilidade, eficácia, dificuldade de uso, facilidade de usar o agente, entre outros requisitos de avaliação de *software*.

Como anexo à solicitação de desenvolvimento de pesquisa Fecomércio-RS/SESC/SENAC, está o modelo de termo dos participantes do estudo.

1.2.4 Dados do Sistema Fecomércio-RS/Sesc/Senac necessários para a pesquisa: A pesquisa ocorrerá de forma totalmente remota e não haverá nenhum custo para o Sistema Fecomércio-RS/Sesc/Senac. Será utilizada a plataforma *Teams*. Os agentes conversacionais estão hospedados na UFRGS e um formulário de pesquisa *Google Forms*, desenvolvido pelo pesquisador, estará disponível.

1.2.5 Local de realização da coleta de dados: Senac São Leopoldo - RS

1.2.6 Outras informações/solicitações que julgar pertinentes: A pesquisa ocorrerá no contexto das aulas do Curso Técnico em Informática para Internet não prejudicando o andamento dos conteúdos previstos.

2 ORIENTAÇÕES AO PESQUISADOR

2.1 O nome do Sistema Fecomércio-RS/Sesc/Senac não deve constar na pesquisa, salvo se houver autorização prévia para sua divulgação, mediante solicitação específica e justificada.

2.2 Os resultados da pesquisa poderão ser utilizados pelo Sistema Fecomércio-RS/Sesc/Senac, respeitando-se o crédito intelectual do pesquisador.

2.3 Os sujeitos da pesquisa deverão assinar Termo de Consentimento Livre e Esclarecido, elaborado e preservado pelo pesquisador.

2.4 A pesquisa deverá ser disponibilizada, em formato eletrônico, junto ao local onde foi realizada (unidade/escola ou departamento/Sistema Fecomércio-RS/Sesc/Senac) em até 30 dias após sua aprovação.

2.5 Declaro que estou ciente e de acordo com os termos e requisitos estabelecidos neste documento.

Assinatura do pesquisador

Data: ____ / ____ / ____

APENDICE F- AUTORIZAÇÃO DE PESQUISA FECOMÉRCIO RS/SESC/SENAC

Juliane Monticelli dos Santos

Seg, 29/03/2021 14:09



Para: Clovis da Silveira; Fulvia da Silva Spohr

Cc: Jozilda Berenice Candido Fogaca; Antônia Regina Gomes Neves; Lianamar da Silveira Rosa +1 outra pess



Pesquisa Sistema Fecomércio ...

38 KB

Prezado, Clóvis

Acusamos o recebimento e análise da sua solicitação. Informamos que poderá proceder com a pesquisa, observando as orientações descritas no documento enviado. A realização das atividades de pesquisa deverá ser negociada junto ao setor pedagógico da escola EM Senac São Leopoldo.

Desejamos sucesso e permanecemos à disposição.
Atenciosamente.

Juliane Monticelli dos Santos
Gerência de Educação Profissional
Sistema Fecomércio-RS | Porto Alegre/RS
(51) 3375-7973



APÊNDICE G – EMAIL ENVIADO AOS PAIS PARA PARTICIPAÇÃO DA PESQUISA

São Leopoldo, 23/04/21.

Prezados pais,

Pelo presente, convidamos seu(sua) filho(a) a participar de uma pesquisa durante as aulas do itinerário formativo do Curso Técnico em Informática.

A pesquisa faz parte dos estudos de doutorado do Professor Coordenador do Curso Técnico, Professor Me. Clóvis Silveira, junto ao programa de Pós-Graduação em Informática na Educação da UFRGS, sob a orientação da Professora Dra. Liane Tarouco. A pesquisa está aprovada e autorizada pelo Comitê de Pesquisa e Ética do SENAC-RS.

A pesquisa visa avaliar tecnicamente a percepção dos estudantes sobre o uso e funcionalidades do recurso computacional: agentes conversacionais educacionais.

Destacamos que a pesquisa tem aderência ao conteúdo curricular do Curso Técnico em Informática, no que se refere ao desenvolvimento de *sites* na linguagem HyperText Markup Language (HTML), no contexto da *Artificial Intelligence Markup Language* (AIML).

A coleta de dados consistirá no preenchimento de um questionário sem identificação de dados pessoais do aluno. O questionário será disponibilizado durante o mês de maio, no período de realização de aulas do Ensino Técnico. Os dados dos estudantes serão mantidos em sigilo para uso exclusivo do estudo.

O instrumento consistirá em perguntas abertas (dissertativas) e fechadas (múltipla escolha), com duração de preenchimento de, aproximadamente, 20 minutos. As perguntas versarão exclusivamente sobre a experiência de usuário com os agentes avaliados.

Solicitamos aos pais/responsáveis informar se autorizam ou não o(a) estudante a participar da pesquisa até a **sexta-feira**, dia 30/04/21.

Em anexo, enviamos o Termo de Consentimento Livre e Esclarecido para conhecimento.

Em caso de dúvidas e esclarecimentos, contatar o Professor Coordenador do Curso Técnico Clóvis Silveira: cdsilveira@senacrs.com.br (whats 51984632968).

[Clique aqui para informar se autoriza a participação da pesquisa.](#)

APÊNDICE H – INSTRUMENTO MODELO DE AVALIAÇÃO DAS MÉTRICAS POR ESPECIALISTAS - METODOLOGIA FASE 3

Nome do Avaliador:	
Profissão / Instituição:	
Formação:	

Contextualização
<p>Nesta pesquisa, estamos propondo o desenvolvimento de um <i>framework</i>, com métricas de avaliação de <i>software</i>, baseado na ISO/IEC 25010, e outras métricas identificadas em estudos nacionais e internacionais, envolvendo a avaliação de agentes conversacionais.</p> <p>Cada categoria tem uma ou mais subcategorias. Cada subcategoria tem uma afirmação relacionada à métrica específica.</p> <p>Diretrizes de avaliação:</p> <ol style="list-style-type: none">1) A questão é relevante para a avaliação de <i>chatbots</i>? Caso não, rejeite-a (você pode escrever uma justificativa no campo "Ajustes ao item").2) A questão condiz com a categoria/subcategoria indicada? Caso não, sugira qual categoria/subcategoria seria mais apropriada (você pode escrever uma justificativa no campo "Ajustes ao item").3) A questão segue a estrutura padrão de questionamento? Caso não, sugira ajustes no campo "Ajustes ao item".4) Este questionamento já foi realizado (de outra forma, nessa ou em outra dimensão)? Caso sim, rejeite-a (de preferência, escrevendo uma justificativa ao lado do campo "Ajustes ao item").

Fatores e Dimensões		Nº	Itens	Ajustes ao item		
Categoria	Subcategoria		Itens para avaliação da subcategoria	Aceitar	Rejeitar	Reescreva o item com o ajuste sugerido
Funcionalidade	Funcionalidade completa	1	O <i>chatbot</i> apresenta funções que atendam a sua funcionalidade completa.			
	Qualidade das respostas	2	O <i>chatbot</i> apresenta respostas adequadas durante a conversação.			
	Respostas de <i>fallback</i> *	3	O <i>chatbot</i> é capaz de apresentar a quantidade de vezes que falhou ou passou por uma situação de falha.			
	Métricas de mensagens	4	O <i>chatbot</i> apresenta mensagens iniciais de conversação ou saudações a fim de iniciar um diálogo.			
Eficiência	Comportamento em relação ao tempo	5	O <i>chatbot</i> apresenta um tempo rápido de respostas.			
	Eficácia	6	O <i>chatbot</i> atende, com precisão e integridade, quanto ao objetivo do diálogo.			
Compatibilidade	Interoperabilidade	7	O <i>chatbot</i> é capaz de interagir com outro sistema ou plataforma específica.			
Usabilidade	Reconhecimento de adequação	8	É possível reconhecer que o <i>chatbot</i> é apropriado para suas necessidades.			
	Aprendizagem	9	O <i>chatbot</i> permite aprender como usá-lo em situações de emergência, ou em situação de uso imediato.			
	Operabilidade	10	O <i>chatbot</i> é fácil de operar e controlar.			
	Proteção contra erros de usuário	11	Em caso de algum possível erro na resposta, o <i>chatbot</i> segue funcionando para novos diálogos.			
	Acessibilidade	12	O <i>chatbot</i> pode ser utilizado por pessoas com diferentes características físicas.			
Confiabilidade	Maturidade	13	O <i>chatbot</i> atende às necessidades de confiabilidade em operação normal.			
	Disponibilidade	14	Quando você precisou, o <i>chatbot</i> estava disponível, ou há informação que estará disponível no momento que você precisar?			

	Tolerância ao erro	15	Caso algum <i>software</i> do seu <i>notebook</i> , computador ou telefone apresente um possível erro de atualização de outro aplicativo, o <i>chatbot</i> segue funcionando normalmente.			
Segurança	Confidencialidade	16	O <i>chatbot</i> tem algum tipo de usuário e senha ou, ainda, um <i>link</i> específico de acesso para que seja possível acessar o mesmo.			
	Integridade	17	O <i>chatbot</i> garante que o usuário não irá alterar determinadas configurações, tais como programações, botões, informações na base de conhecimento.			
	Autenticidade	18	O <i>chatbot</i> permite sua identificação, a fim de garantir o sigilo ou os diálogos já estabelecidos.			
Manutenção	Reutilização	19	Informações da base de conhecimentos poderão ser aproveitadas para outros futuros <i>chatbots</i> .			
	Testabilidade	20	O <i>chatbot</i> possibilita estabelecer critérios de testes a serem realizados, a fim de determinar se os critérios estabelecidos foram atendidos.			
Portabilidade	Adaptabilidade	21	O <i>chatbot</i> funciona em diferentes sistemas operacionais e em diferentes configurações de hardware.			
Satisfação	Utilidade / Motivação para utilização	22	Após usar o <i>chatbot</i> , sinto-me satisfeito em relação ao uso de meus objetivos, tais como os resultados do uso.			
	Confiança	23	Confio que o <i>chatbot</i> vai comportar-se conforme o pretendido.			
	Prazer/satisfação	24	O <i>chatbot</i> atendeu às suas necessidades pessoais.			
	Afetividade	25	O <i>chatbot</i> apresenta diálogos com saudações, cordialidade e é capaz de responder, com cordialidade, às interações humoradas dos participantes.			
	Humanidade	26	O <i>chatbot</i> é capaz de interagir de maneira convincente, satisfatória e natural.			
	Ética e Comportamento	27	O <i>chatbot</i> apresenta respeito, inclusão e preservação da dignidade durante os diálogos.			
Cobertura de Contexto	Compleitude do contexto	28	O <i>chatbot</i> apresentou informações envolvendo uma completude de contexto.			
	Manter a conversa	29	O <i>chatbot</i> é capaz de manter o diálogo.			
Controle de Usuários	Usuários ativos	30	O <i>chatbot</i> permite identificar os usuários ativos ou on-line.			

	Total de usuários	31	O <i>chatbot</i> permite identificar o total de usuários que conversaram com os usuários.			
--	-------------------	----	---	--	--	--

* *Fallback*: Captura a quantidade de vezes que o *chatbot* falhou ou passou por uma situação de quase falha.

Observações e outras sugestões:

APÊNCIDE I - TERMO DE CONSENTIMENTO LIVRE ESCLARECIDO - ESPECIALISTAS

Pelo presente, convido-o a participar da pesquisa intitulada: “**Estratégias de Avaliação de Agentes Conversacionais Educacionais**”. Trata-se da avaliação de métricas de avaliação de agentes conversacionais, que será utilizada como parte do estudo de tese do doutorando Clóvis da Silveira, orientado pela Prof.^a Dra. Liane Tarouco, e Coorientado pelo Prof. Dr. Fabrício Herpich, no programa de Pós-Graduação em Informática na Educação (PPGIE), do Centro Interdisciplinar de Novas Tecnologias na Educação (CINTED), da Universidade Federal do Rio Grande do Sul (UFRGS).

A pesquisa tem como objetivo criar um *framework* com sugestões de métricas de avaliação de *software*, a fim de elencar o que um agente conversacional educacional pode e deve ter para ser relevante, com vistas a ser utilizado como referência para a avaliação e construção de *chatbots* desenvolvidos para o contexto educacional. Ela será aplicada de maneira on-line, no ano de 2021, a partir do envio de e-mails, de mensagens a grupos específicos em redes sociais. O público que pode contribuir com essa pesquisa são professores desenvolvedores, pesquisadores, programadores, profissionais de Tecnologia da Informação e autores de artigos de *chatbots* ou agentes conversacionais.

Todos os instrumentos a serem aplicados serão mantidos em sigilo, servindo apenas para fins de pesquisa, não se revelando os nomes dos participantes. A sua participação não oferece risco algum. Caso seja verificado algum constrangimento, o pesquisador irá intervir. É-lhe garantido, também:

- Receber resposta a qualquer pergunta, ou esclarecimento a qualquer dúvida acerca dos procedimentos, riscos, benefícios e outros assuntos relacionados com a pesquisa;
- Poder retirar seu consentimento a qualquer momento, deixando de participar do estudo, sem que isso ofereça qualquer tipo de prejuízo;
- Não ser identificado quando da divulgação dos resultados, tendo a garantia de que todas as informações obtidas serão utilizadas apenas para fins científicos vinculados à pesquisa;
- Gastos adicionais serão absorvidos pelo orçamento da pesquisa.

Este documento deverá ser assinado em duas vias, sendo que uma delas será retida pelo sujeito da pesquisa e a outra pelo pesquisador. O responsável pela pesquisa é o aluno de doutorado Clóvis da Silveira, seu e-mail é csclovis@gmail.com, e o telefone é: 5198463.2968.

Pelo presente Termo de Consentimento Livre e Esclarecido, concordo com a participação nesta pesquisa, pois fui devidamente informado(a), de forma clara e detalhada, livre de qualquer constrangimento e coerção, dos objetivos e da justificativa do estudo, dos instrumentos de coleta de informação que serão utilizados e dos riscos e benefícios, conforme já citados neste termo.

Nome do participante da pesquisa:

E-mail:

Assinatura do participante da pesquisa: _____

Pesquisador: Clóvis da Silveira

Assinatura: _____