



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CAMPUS RUSSAS**  
**CURSO DE GRADUAÇÃO EM ENGENHARIA DE SOFTWARE**

**ELIS IONARA VIEIRA DO NASCIMENTO E SILVA**

**UMA AVALIAÇÃO DA USABILIDADE E EXPERIÊNCIA DO USUÁRIO NO  
GOOGLE ASSISTANT**

**RUSSAS**

**2021**

ELIS IONARA VIEIRA DO NASCIMENTO E SILVA

UMA AVALIAÇÃO DA USABILIDADE E EXPERIÊNCIA DO USUÁRIO NO GOOGLE  
ASSISTANT

Trabalho de Conclusão de Curso apresentado ao  
Curso de Graduação em Engenharia de Software  
do Campus Russas da Universidade Federal do  
Ceará, como requisito parcial à obtenção do  
grau de bacharel em Engenharia de Software.

Orientadora: Prof<sup>a</sup>. Dr<sup>a</sup>. Marília Soares  
Mendes

RUSSAS

2021

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Biblioteca Universitária  
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

S579a Silva, Elis Ionara Vieira do Nascimento.  
Uma Avaliação da Usabilidade e Experiência do Usuário no Google Assistant / Elis Ionara Vieira do Nascimento Silva. – 2021.  
75 f. : il.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Campus de Russas, Curso de Engenharia de Software, Russas, 2021.  
Orientação: Profa. Dra. Marília Soares Mendes.

1. Assistentes por Voz Inteligentes. 2. Usabilidade. 3. Experiência do Usuário. I. Título.

CDD 005.1

---

ELIS IONARA VIEIRA DO NASCIMENTO E SILVA

UMA AVALIAÇÃO DA USABILIDADE E EXPERIÊNCIA DO USUÁRIO NO GOOGLE  
ASSISTANT

Trabalho de Conclusão de Curso apresentado ao  
Curso de Graduação em Engenharia de Software  
do Campus Russas da Universidade Federal do  
Ceará, como requisito parcial à obtenção do  
grau de bacharel em Engenharia de Software.

Aprovada em:

BANCA EXAMINADORA

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Marília Soares Mendes (Orientadora)  
Universidade Federal do Ceará (UFC)

---

Thiago Hellen Oliveira da Silva  
Deloitte Touche Tohmatsu

---

Prof<sup>a</sup>. Dr<sup>a</sup>. Patrícia Freitas Campos de Vasconcelos  
Universidade Federal do Ceará (UFC)

Dedico este trabalho, a minha família em especial aos meu pais Elis Regina e Oséas, ao meu padrinho e aos meus amigos.

## AGRADECIMENTOS

Primeiramente a Deus, por me dar a capacidade para realizar meus sonhos, e nos momentos mais difíceis me dar força e coragem para continuar essa jornada.

Aos meus amados pais, Elis Regina e Oséas por todos os ensinamentos que me formaram e me ensinaram a ser uma pessoa de caráter e princípios. Por me apoiarem desde sempre, sendo pais, mas também sendo amigos. Por sempre estarem ao meu lado, seja presencial ou à distância, me dando força e sempre me dizendo que tudo ia dar certo. Por sempre me colocarem em suas orações e dizendo pro cara lá de cima iluminar meus passos. Por todo o sacrifício feito em prol da minha felicidade. Ao meu irmão Emanuel Víctor, por ser um dos meus maiores amigos, me fazendo sorrir, me ensinando coisas novas e sempre do meu lado.

Ao meu padrinho e segundo pai, Edilberto Coelho por toda, amizade, companheirismo, e contribuição em todos os momentos de minha formação. Por sempre me tratar de forma carinhosa, acolhedora. Agradeço por todo apoio, compreensão e incentivo em todas as fases. Por todas as conversas e mensagens de apoio, o senhor é uma pessoa muito importante pra mim.

A minha amiga e irmã, Sabrina de Oliveira, por nunca deixar que eu desistisse, você sabe que não foi fácil né amiga, já passamos por muita coisa. Por sempre me trazer uma paz e equilíbrio de forma inimaginável, você sabe que sempre pode contar comigo, assim como sei que posso contar com você. Por me tratar como parte de sua família.

Ao meu amigo e namorado Thomas Dillan, que cuidou e me acolheu durante essa jornada que não é fácil. Por pegar na minha mão e dizer que eu ia conseguir, e em muitas vezes me mostrando como seguir nas dificuldades. Por ser meu mentor durante essa fase importante que é encontrar um trabalho que seja satisfatório e faça a diferença. Á sua família pelo acolhimento e força.

Ao meu amigo, Carlos Victor Dantas, pela amizade incondicional, por todos os ensinamentos e por toda a dedicação. Por ser uma fonte de inspiração e mesmo que tão longe se fazendo presente.

A minha orientadora, prof<sup>a</sup>. Marília Soares Mendes por me orientar, me incentivar e principalmente inspirar como mulher, pesquisadora e professora. Por me acolher no projeto DICA VI e no grupo de pesquisa LINCE, me dando a chance de ter muitas oportunidades.

Aos meus amigos Nathalia Gonçalves, Luis Filipe e Paulla Gama por serem meus primeiros amigos na UFC e me ensinarem um pouco todos os dias o que é fazer parte de um grupo.

Aos meus amigos Tagila Lima, Vinicius Almeida, Isaias Soares e Isaac Rahel que me acompanharam mesmo por pouco tempo, mas que me ensinaram tanto sobre as decisões dentro e fora da universidade. Por se fazerem presentes mesmo depois da universidade, me mostrando que as amizades independem dos lugares e que não possuem barreiras. Amo vocês!

Obrigado a todos!

## RESUMO

Assistentes por Voz Inteligentes (AVIs) são um novo tipo de tecnologia que atuam como assistentes pessoais, gerenciadores de tarefas e em algumas situações podem atuar até como "amigo pessoal", pois eles estão embutidos nos dispositivos mais usados no mundo, os smartphones. A forma como esses AVIs funcionam podem causar estímulo ou desestímulo ao uso dessa e de outras novas tecnologias, por esse motivo, a avaliação de sistemas se torna uma ferramenta fundamental para a experiência do usuário e para a projeção e construção da interface dos sistemas. Designers, gerentes de projeto e todos aqueles que se envolvem em um projeto tem por objetivo oferecer uma boa experiência de uso ao usuário. Para isso, foram usados dois métodos avaliativos para obtenção de visões diferentes sobre os problemas de usabilidade existentes no assistente pessoal Google Assistant. O primeiro método avaliativo usado na primeira fase deste trabalho, foi a metodologia MALTU, aplicada em comentários deixados na loja de aplicativos Google Play Store para obter informações relacionadas ao AVI, no caso, obter Postagens Relacionadas ao Uso - PRUs. Ao final da aplicação da metodologia foram gerados gráficos que interpretados indicaram os principais problemas de usabilidade do AVI. Para a segunda fase deste trabalho, foram considerados os resultados já obtidos na aplicação do primeiro método avaliativo, e dessa forma, foi possível realizar uma avaliação remota focada nos problemas de usabilidade já destacados na primeira fase. Por esse motivo, a criação de cenários de uso como forma de guiar os avaliadores durante a realização dos testes, como forma de não deixar brechas no momento de execução dos testes. E assim, obter resultados comparativos dos problemas de usabilidade relacionados ao uso do AVI.

**Palavras-chave:** Assistente por Voz Inteligente. Usabilidade. Experiência do Usuário. Avaliação Textual. Avaliação Remota. Postagens Relacionadas ao Uso. Classificação de Postagens.

## ABSTRACT

Intelligent Voice Assistants (IVAs) are a new type of technology that act as personal assistants, task managers and in some situations even act as a "personal friend", as they are embedded in the most used devices in the world, smartphones. The way these IVAs work can stimulate or discourage the use of this and other new technologies, for this reason, systems assessment becomes a tool fundamental for the user experience and for the design and construction of the systems interface. Designers, project managers and all those who are involved in a project aim to offer a good user experience. For this, two evaluation methods were used to obtain different views on the usability problems existing in the personal assistant Google Assistant. The first evaluation method used in the first phase of this work, was the MALTU methodology, applied in comments left in the Google Play Store app store to obtain information related to IVA, in this case, to obtain Posts Related to Usage - PRUs. At the end of the application of the methodology, graphs were generated that interpreted indicated the main usability problems of the IVA. For the second phase of this work, the results obtained in the application of the first evaluation method were considered, with the objective of a remote evaluation focused on the usability problems already highlighted in the first phase. For this reason, the creation of usage scenarios as a way to guide the evaluators during the performance of the tests, as a way of not leaving gaps in the moment of the tests' execution. And so, obtain comparative results of usability problems related to the use of IVA.

**Keywords:** Intelligent Voice Assistant. Usability. Textual Evaluation. Remote Assessment. Posts Related to Usage. Post Rating.

## LISTA DE FIGURAS

Figura 1 – Procedimento Metodológico . . . . .	16
Figura 2 – Shoebox . . . . .	19
Figura 3 – Alexa . . . . .	20
Figura 4 – Logo Bixby . . . . .	21
Figura 5 – Cortana . . . . .	22
Figura 6 – Google Assistant . . . . .	22
Figura 7 – Sara . . . . .	23
Figura 8 – Logotipo da Siri - versão atual . . . . .	24
Figura 9 – Procedimento da Aplicação da Metodologia MALTU . . . . .	39
Figura 10 – Gráfico de PRUs e Não-PRUs . . . . .	41
Figura 11 – Gráficos da Avaliação Textual . . . . .	45
Figura 12 – Procedimento da Aplicação da Avaliação Remota . . . . .	48
Figura 13 – Storyboards dos Cenários . . . . .	50
Figura 14 – Imagens Durante a Execução dos Testes . . . . .	53

## LISTA DE TABELAS

Tabela 1 – Sugestões de Funcionalidades . . . . .	46
---	----

## LISTA DE QUADROS

Quadro 1 – Quadro Comparativo de AVIs . . . . .	25
Quadro 2 – Quadro Comparativo dos Métodos de IHC . . . . .	31
Quadro 3 – Quadro Comparativo dos Trabalhos Relacionados . . . . .	38
Quadro 4 – Quadro de Postagens Antes do Processo de Sentenciação . . . . .	40
Quadro 5 – Quadro de Postagens Depois do Processo de Sentenciação . . . . .	40
Quadro 6 – Exemplos de Classificação por Tipo . . . . .	41
Quadro 7 – Exemplos de Classificação por Intenção do Usuário . . . . .	42
Quadro 8 – Exemplos de Classificação por Informação do Sentimento - Polaridade . . . . .	42
Quadro 9 – Exemplos de Classificação por Qualidade de Uso - Usabilidade . . . . .	42
Quadro 10 – Exemplos de Classificação por Qualidade de Uso - UX . . . . .	43
Quadro 11 – Exemplos de Classificação por Funcionalidade . . . . .	43
Quadro 12 – Exemplos de Classificação por Plataforma . . . . .	43
Quadro 13 – Classificação por Tipo x Funcionalidades . . . . .	46
Quadro 14 – Critérios de Qualidade de Uso - Usabilidade x Funcionalidades . . . . .	46
Quadro 15 – Critérios de Qualidade de Uso - UX x Funcionalidades . . . . .	46
Quadro 16 – Quadro de Perfil dos Usuários . . . . .	48
Quadro 17 – Quadro de Descrição das Funcionalidades . . . . .	49
Quadro 18 – Quadro de Funcionalidades mais Criticadas . . . . .	49
Quadro 19 – Cenários x Funcionalidades . . . . .	50
Quadro 20 – Quadro com os Tempos de Execução dos Testes . . . . .	52
Quadro 21 – Quadro Descritivo do Teste - Participante 1 . . . . .	53
Quadro 22 – Quadro Descritivo do Teste - Participante 2 . . . . .	54
Quadro 23 – Quadro Descritivo do Teste - Participante 3 . . . . .	54
Quadro 24 – Quadro Descritivo do Teste - Participante 4 . . . . .	54
Quadro 25 – Quadro Descritivo do Teste - Participante 5 . . . . .	55

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> . . . . .	13
<b>1.1</b>	<b>Objetivos</b> . . . . .	15
<i>1.1.1</i>	<i>Objetivo Geral</i> . . . . .	15
<i>1.1.2</i>	<i>Objetivos Específicos</i> . . . . .	15
<b>1.2</b>	<b>Procedimentos Metodológicos</b> . . . . .	15
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> . . . . .	18
<b>2.1</b>	<b>Assistentes por Voz Inteligentes</b> . . . . .	18
<b>2.2</b>	<b>Avaliação de Sistemas</b> . . . . .	26
<i>2.2.1</i>	<i>Tipos de Avaliação</i> . . . . .	27
<i>2.2.1.1</i>	<i>Métodos de Inspeção</i> . . . . .	27
<i>2.2.1.2</i>	<i>Métodos de Observação</i> . . . . .	28
<i>2.2.1.3</i>	<i>Métodos de Investigação</i> . . . . .	29
<b>2.3</b>	<b>Métodos de Avaliação Utilizados neste Trabalho</b> . . . . .	31
<i>2.3.1</i>	<i>Metodologia MALTU</i> . . . . .	31
<i>2.3.2</i>	<i>Teste Remoto de Usabilidade</i> . . . . .	33
<b>3</b>	<b>TRABALHOS RELACIONADOS</b> . . . . .	35
<b>4</b>	<b>APLICAÇÃO DA METODOLOGIA MALTU</b> . . . . .	39
<b>5</b>	<b>TESTE REMOTO</b> . . . . .	48
<b>6</b>	<b>DISCUSSÃO</b> . . . . .	56
<b>7</b>	<b>CONCLUSÃO</b> . . . . .	58
<b>7.1</b>	<b>Trabalhos Futuros</b> . . . . .	58
	<b>REFERÊNCIAS</b> . . . . .	59
	<b>ANEXOS</b> . . . . .	61
	<b>ANEXO A</b> – Relatório de Avaliação . . . . .	62
	<b>ANEXO B</b> – Termo de Consentimento . . . . .	73
	<b>ANEXO C</b> – Roteiro de teste . . . . .	76

## 1 INTRODUÇÃO

O advento da Inteligência Artificial (IA) e o uso exponencial de smartphones promoveram o uso dos Assistentes por Voz Inteligentes (AVIs), sistemas que auxiliam pessoas com tarefas do dia a dia, realizando buscas por voz como gatilhos de ação (ALENCAR *et al.*, 2013). Os AVIs buscam oferecer ao usuário maior praticidade e conforto por meio de esforço mínimo durante a realização das tarefas (West *et al.*, 2019 apud MOTTA; QUARESMA, 2019), além da intuitividade, uma vez que os usuários utilizam a fala, que é a linguagem natural aos seres humanos (AMAZON, 2019; PEARL, 2016, KLEINER PERKINS, 2016 apud MOTTA; QUARESMA, 2019).

O crescente uso dos AVIs no Brasil, deve-se em grande parte à compatibilidade com o sistema Android, sendo o sistema mais popular no mundo e no Brasil (THIAGO LAVADO, 2019). Com esse crescente uso e evolução dos AVIs, modificações no comportamento e no cotidiano de muitas pessoas, desde o modo de execução de tarefas simples, como uma pesquisa no Google, até tarefas que exigem um maior esforço, como uma compra, que pode ser feita utilizando um comando por voz, são notados como consequência da interação sistema-usuário (IMPrensa Mercado e Consumo, 2019).

Uma definição de interação é apresentada por Prates e Barbosa (2003), como o processo de comunicação entre pessoas e sistemas interativos, onde o usuário e o sistema trocam turnos em que um fala e o outro "ouve", interpreta e realiza uma ação. Essa interação sistema-usuário revela aspectos como facilidade de uso, como o usuário se sente durante o uso do sistema, se o sistema está adequado ao uso correspondendo as expectativas do usuário, que acabam por compor um cenário para que possa acontecer uma experiência boa e fluida para o usuário, fazendo com que ele alcance os objetivos desejados durante a interação com o sistema, e volte a pensar naquele sistema como uma solução para outros problemas que possa vir a ter. Por isso, um estudo dos aspectos da interação, as formas de avaliação de sistemas e a forma como influenciam para uma experiência fluida é importante para analisar estes aspectos.

O estudo dos aspectos do uso de sistema, da experiência e satisfação do usuário são importantes para a identificação de problemas que surgem durante o processo de comunicação e interação entre o sistema e o usuário. Alguns trabalhos têm se dedicado à avaliação de AVIs a fim de avaliar o desempenho entre dois AVIs (Google Assistant e Siri) como o de Lima *et al.* (2019b) que realizou uma análise de falas com problemas linguísticos (cacoépie e disfluência), concluindo por meio de experimentos que participantes das regiões Sul (S) e Sudeste (SE)

tendem a ter interações mais fluidas como AVI, por terem menos problemas linguísticos. E o trabalho de Lima *et al.* (2019b) que analisa o desempenho entre os dois AVIs mencionados anteriormente, relacionado a presença de viés na interação com uma análise realizada por meio de áudios, concluindo que o Google Assistant atendeu melhor as frases faladas por pessoas do sexo feminino, diferença que não foi relevante para a Siri.

Diante da análise desses trabalhos, é possível reforçar que problemas decorrentes da usabilidade acabam causando a exclusão de usuários, e assim como cacoécia e disfluência e o entendimento de falas do sexo feminino, a baixa escolaridade, os fatores socioeconômicos também são problemas na comunicação que podem causar a exclusão e causar impactos negativos nos usuários. Por esse motivo, as fases de avaliação e testes com usuários são consideradas fundamentais para a produção de sistemas interativos com alta qualidade de uso, ajudando na detecção e resolução de problemas antes da inserção do sistema no cotidiano dos usuários (BARBOSA; SILVA, 2010).

Neste trabalho, problemas de interação serão investigados como forma de entender a influência positiva ou negativa do uso dos AVIs no cotidiano dos usuários, com foco na usabilidade e na experiência do usuário com esses tipos de sistema. E, como forma de detecção de problemas durante o uso, a investigação da satisfação e da experiência do usuário será realizada de duas formas. A primeira por meio de opiniões e avaliações dos usuários sobre o assistente, no caso, comentários no aplicativo Google Play, e a segunda, por meio de uma avaliação remota, para que o uso do assistente seja avaliado em tempo real pelo avaliador. Uma das motivações para este trabalho, além da detecção desses problemas, é a participação da autora deste trabalho em um projeto denominado DICAVI (Design Inclusivo para Cidadãos de Assistentes Virtuais Inteligentes), que traz como objetivo principal projetar e implementar um assistente de voz inteligente para o atendimento em serviços públicos do Estado do Ceará, compreendendo cidadãos em diversos cenários para obtenção de uma abordagem mais inclusiva possível.

As técnicas de avaliação de sistemas que são utilizadas são: a avaliação por texto, uma avaliação que utiliza dados subjetivos expressados pelos usuários do sistema, como opiniões e preferências, no caso, os textos (BARBOSA; SILVA, 2010), como forma de coletar a interação dos usuários com o AVI via texto. É realizada uma avaliação textual dos comentários deixados na página do aplicativo na loja Google Play Store, por meio da metodologia MALTU (Modelo para a Avaliação da Interação em Sistemas Sociais a partir da Linguagem Textual do Usuário)

(MENDES, 2015). E a avaliação remota semi-instrumentada, um tipo específico da avaliação remota, que usa a coleta seletiva de dados acionada diretamente pelos usuários que executam tarefas em seu contexto normal de trabalho (HARTSON *et al.*, 1996). Com o auxílio de um aplicativo de monitoramento, observar os usuários quando estão realizando tarefas com o AVI, e dessa forma, obter resultados com o usuário no atual contexto de pandemia em que vivemos, onde o isolamento social se tornou necessário como modo de combate à pandemia por COVID-19.

Este trabalho se divide em 7 seções, organizadas da seguinte maneira, este capítulo apresenta uma contextualização do tema, a problemática, a motivação para este trabalho, bem como os objetivos (Tópico 1.1) e o procedimento metodológico (Tópico 1.2). O Capítulo 2 apresenta conceitos que irão auxiliar a compreensão do trabalho proposto. O Capítulo 3 apresenta alguns trabalhos relacionados. O Capítulo 4 apresenta uma descrição de como foi realizada a aplicação da metodologia MALTU. O Capítulo 5 descreve de forma detalhada como a aplicação do teste remoto foi realizado. O Capítulo 6 apresenta uma comparação entre os resultados obtidos. O Capítulo 7 apresenta conclusão do trabalho e trabalhos futuros.

## **1.1 Objetivos**

### ***1.1.1 Objetivo Geral***

Investigar os principais problemas do Google Assistant relacionados à usabilidade.

### ***1.1.2 Objetivos Específicos***

- Apresentar um estado da arte sobre assistentes por voz inteligentes;
- Apresentar resultados de duas avaliações da usabilidade e experiência do usuário em um assistente por voz inteligente;

## **1.2 Procedimentos Metodológicos**

Nesta seção são apresentados os procedimentos adotados durante a pesquisa, juntamente com a descrição de cada fase ressaltada na Figura 1.



Fonte: Elaborado pelo Autor (2021)

### 1. Estudo dos Conceitos Relacionados

Na fase inicial desse trabalho foi realizada uma revisão da literatura dos trabalhos relacionados abordando avaliação de AVIs, com o objetivo de obter conhecimento necessário para realizar este trabalho. Inicialmente, foram definidos os principais temas e aspectos relacionados a esse tipo de avaliação, tais como: avaliação, desempenho e comportamento dos sistemas por voz, bem como as formas de experimentos com os usuários.

### 2. Análise dos Trabalhos Relacionados

Nesta fase, foi realizado um estudo dos trabalhos relacionados com o objetivo de entender os tipos de avaliações realizadas em AVIs. Esse procedimento serviu para escolher os métodos mais adequados para este trabalho, no caso, na fase de avaliação textual, a metodologia MALTU, aplicada na primeira parte deste trabalho. E, o outro método escolhido para a avaliação remota, foi a aplicação de um teste remoto, na segunda parte do trabalho corrente.

### 3. Definição do Método de IHC

Nesta fase, foram definidos os métodos avaliativos mais adequados ao momento em que estamos inseridos, um momento de pandemia de Covid-19 e mesmo assim, seguir as recomendações como forma de prevenção e contenção, como o isolamento e distanciamento social. Por esse motivo, a avaliação textual, realizada com a aplicação da metodologia MALTU, que foi seguida durante o processo indo da coleta à interpretação dos resultados relacionados a classificação das postagens no sistema. E, o outro método escolhido, foi a aplicação de testes remotos com o intuito de obter resultados mais detalhados de uso.

#### **4. Aplicação da Metodologia MALTU**

A metodologia foi aplicada com o objetivo de detectar problemas de usabilidade e experiência do usuário no Google Assistant. A aplicação da MALTU, a extração dos dados, a aplicação do processo de sentencição para a obtenção de informações relacionadas a usabilidade e experiência do usuário serão apresentados mais adiante.

#### **5. Avaliação Remota**

A avaliação remota foi realizada com o objetivo de validar diretamente com os usuários alguns problemas de usabilidade, que foram detectados com a aplicação da metodologia MALTU. Foi utilizada uma avaliação remota semi-instrumentada, realizada com o auxílio de um software para gravação das reuniões com os usuários, para o acompanhamento das ações dos usuários à medida que estes tentam realizar as tarefas que foram repassadas, antes do início do teste.

#### **6. Análise dos Resultados**

A análise dos resultados foi realizada de formas diferentes dentro de cada tipo de avaliação. Com a aplicação da metodologia MALTU, os resultados foram analisados a partir da combinação de Tabelas e Gráficos como forma de filtrar informações relacionadas a usabilidade e experiência do usuário. Já nos Testes Remotos, a análise foi realizada com a interpretação dos vídeos das reuniões com os usuários, e assim, gerar possíveis tabelas, com um relato dos problemas encontrados durante o uso, também relacionados a usabilidade.

## 2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção são apresentados alguns conceitos introdutórios que auxiliarão no entendimento do trabalho. Como o principal foco é apresentar resultados de uma avaliação da usabilidade e experiência do usuário em AVIs, são apresentados AVIs, avaliação de sistemas, usabilidade e experiência do usuário.

### 2.1 Assistentes por Voz Inteligentes

Budiu e Laubheimer (2018) definem os AVIs como sistemas que recebem como entrada do usuário comandos de voz em linguagem natural, que por sua vez, são interpretados de forma inteligente por meio de informações adicionais como o contexto da conversa e dessa forma, retornar como resposta ao usuário informações na tela do dispositivo. Os AVIs aprimoram seus serviços por meio de aprendizagem diária sobre o usuário a partir de ações como pesquisas, compras on-line, preferências feitas nos dispositivos em que os assistentes estão presentes.

O Brasil é o país que apresenta 235 milhões de smartphones ativos atuando como assistentes pessoais (ALINE PRADO, 2018) capazes de aumentar a potencialidade da convivência entre dispositivos, usuários a partir da coleta de dados e do contexto de uso. Por conseguinte, o uso dos AVIs causam grandes transformações em diversos cenários na atualidade. Os AVIs auxiliam de diversas formas os usuários, como: a) apenas o uso da voz quando interagem com o assistente, gerando o mínimo de esforço por parte dos usuários, b) a falta de necessidade da aproximação física do usuário quando profere comandos para o assistente, pela linguagem natural, o que não exige que os usuários leiam manual para uso (WEST *et al.*, 2019).

De acordo com a IBM (2011), um dos primeiros esforços em relação a um sistema de reconhecimento por voz foi o Shoebox, apresentado na Figura 1, uma máquina desenvolvida por William C. Derisch, desenvolvedor da IBM. Essa máquina foi apresentada pela primeira vez na feira mundial de 1962, em Seattle, apresentando capacidades como o de reconhecimento e resposta a 16 palavras. A interação com essa máquina era feita com o uso de um microfone que recebia sons e os convertia em impulsos elétricos. A máquina de calcular era ativada pelo reconhecimento dos números de 0 a 9 e por palavras como "mais", "menos" e "total".

Figura 2 – Shoebox



Fonte: IBM (2011)

Atualmente existem assistentes mais avançados e com capacidade de realização de tarefas mais complexas que o Shoebox, descritos a seguir.

A Amazon (2019) lançou em 2014 o assistente Alexa, apresentado na Figura 3, que surgiu com o objetivo de atender os usuários em tarefas simples do dia a dia, como por exemplo, configurar um alarme. Por não ser atrelado a um sistema operacional, é compatível com Android, Windows 10 e com smartphones da marca Apple, ainda com algumas limitações de uso em aplicativos, como exemplo, colocar para executar o aplicativo Apple Music com Alexa.

Os dados recebidos pelo AVI são processados por um servidor na nuvem da Amazon, o Alexa Voice Services (AVS), e assim, transformados em texto que, por sua vez, são analisados retornando ao usuário como comandos baseados na interpretação feita, como uma pesquisa na internet (AMAZON, 2019). Esse processamento de dados feito em nuvem traz como limitação ao AVI a necessidade de acesso à internet para funcionar, não sendo capaz de acessar o AVS sem conexão à rede mundial (AMAZON, 2019).

A Amazon (2019) acrescenta como diferencial para Alexa o "Skill", ou "habilidade", no qual o usuário pode desenvolver uma aplicação para um determinado propósito, sendo controlada por meio de comandos de voz dados ao assistente, e possibilitando aos usuários experiências de voz naturais como uma maneira mais natural de interagir com o assistente. Algumas marcas e serviços famosos do mundo digital aderiram ao lançamento da Alexa como o Uber, para solicitações de corridas, o iFood, para pedidos delivery de comida. Um de seus lançamentos iniciais que também chama atenção, foi o alto-falante domiciliar Echo Dot, como uma aposta para assistentes domiciliares, podendo controlar desde a rotina diária das tarefas de casa até o acender e apagar de uma lâmpada.

Figura 3 – Alexa



Fonte: Amazon (2019)

O assistente Bixby, apresentado na Figura 4-a, é o AVI criado pela Samsung, lançado em 2017, inicialmente integrado a linha de smartphones Galaxy, a partir dos modelos S8 e S8+ (Figura 4 -b). Vem com Inteligência Artificial, tendo como principal função ajudar nas ações do dispositivo, como abrir um aplicativo (FELIPE PAYÃO, 2017). Além disso, oferece um auxílio personalizado, baseado no registro frequente dos interesses, situações e locais dos usuários, possibilitando a realização de compras por meio da busca por imagens e obtenção de detalhes sobre locais próximos com a tecnologia de reconhecimento de imagens (FELIPE PAYÃO, 2017).

Segundo Ry Crist (2019) os dados recebidos pelo Bixby são enviados para um servidor pertencente a Samsung, mas primeiro é necessário uma boa conexão com a internet e possuir uma conta no Samsung Account. O uso desse assistente pode ficar limitado por conta de fatores como: reconhecimento de alguns sotaques e dialetos, gravação de mídia (vídeo, jogo, voz), durante uma chamada (incluindo chamadas realizadas), só é disponibilizado em dispositivos da Samsung considerando a disponibilidade do serviço (recursos, provedores de conteúdo - podem variar de acordo com país, operadora, versão do sistema operacional).

Uma novidade, recentemente lançada pela Samsung, foi a linha de geladeiras FamilyHub, equipada com tela *touch* da marca, possui câmeras embutidas que permitem a identificação do conteúdo da geladeira por meio do telefone do assistente Bixby. Com algumas atualizações que serão realizadas no assistente, será possível a detecção, rotulação e listagem automática de alguns dos alimentos presentes no interior da geladeira (TECHNOLOGY VISION, 2020).

Figura 4 – Logo Bixby



Fonte: Samsung (2020)

A Microsoft começou o desenvolvimento da Cortana em 2009, mas só o lançou em 2013 (Figura 5). Foi projetado inicialmente para os smartphones da marca, mas atualmente se faz presente nos computadores com sistema operacional Windows 10. Ele executa uma variedade de tarefas organizacionais como configuração de lembrete, agendamento de eventos no calendário, de acordo com a configuração estabelecida pelo usuário.

A Cortana recebe dados por meio de texto ou voz, dependendo do usuário, para que seja ativada nos dispositivos compatíveis como Android, iOS e Windows 10 (MICROSOFT SUPPORT, 2019). Ainda necessita de configurações de fala, marcação, digitação e ativação do rastreador de localização, para que seja eficaz na execução das tarefas, porém o usuário é quem determina o acesso do assistente às configurações. A Microsoft fornece atualizações do assistente independentemente das atualizações do Windows 10 (MICROSOFT SUPPORT, 2019).

A Microsoft vem realizando parceria com a Amazon para que a Cortana possa interagir com a Alexa nos dispositivos suportados, no caso, o Amazon Echo e os computadores com Windows 10, onde a Cortana terá acesso aos recursos exclusivos da Alexa e vice-versa. Outra parceria realizada pela Microsoft, é com a Harman Kardon, que anunciaram recentemente a criação da coluna Invoke, que integra suporte ao assistente Cortana, com suporte a dispositivos Android, iOS e Windows 10 (BLOGS DO WINDOWS, 2017a).

Além das parcerias realizadas, a Microsoft traz o Cortana Skills Kit, que permite que os desenvolvedores criem facilmente experiências inteligentes e personalizadas para a Cortana, possibilitando a criação de habilidades da Cortana, os desenvolvedores podem criar experiências personalizadas e altamente relevantes com base nas preferências e no contexto do usuário (BLOGS DO WINDOWS, 2017c).

Figura 5 – Cortana



Fonte: Blogs do Windows (2017b)

O Google Assistant (Figura 6) é o assistente da Google usado no Google Home (Figura 6), um dispositivo inteligente, e em smartphones com Android 7+ instalados (MAIOLINO, 2017). Ele atua como um gerenciador de tarefas auxiliando em tarefas diversas como criar lembretes, enviar mensagens, recomendar lugares, além de retornar mensagens em tempo real, e atuar como um assistente doméstico controlando dispositivos domésticos inteligentes compatíveis (GOOGLE, 2020).

O Google Assistant pode ser ativado ao pressionar o botão home do smartphone, com a necessidade de conexão com a internet para que buscas e comandos por voz sejam executados. Quando o acesso à internet não se torna possível, o AVI guarda a informação recebida, para quando a conexão for reestabelecida a busca seja feita e o usuário obtenha um retorno (GOOGLE, 2020). Ele está disponível nos sistemas Android e iPhone e, por meio de algumas investigações com esse assistente, foram detectados alguns problemas relacionados à forma com que o assistente entendia os comandos que os usuários proferiam, pois ele entendia melhor pessoas do sexo feminino (LIMA *et al.*, 2019a).

Esse AVI pode ser integrado a várias ferramentas que permitem aos usuários criar suas próprias habilidades usando as ferramentas Tasker (TASKER, 2020) e IFTTT (IFTTT, 2020), sendo disponível em ampla plataforma de hardware (smartphones, caixas de som, etc.), e na maioria das plataformas de smartphones, além de poder ser instalado como um aplicativo no iPhone (MAIOLINO, 2017).

Figura 6 – Google Assistant



Fonte: Google (2020)

O laboratório de pesquisa Articulab (2017) da Universidade de Carnegie Mellon, projetou o Sara (Figura 7), um robô que realiza a análise de comportamentos, sentimentos e ações dos usuários, por isso é considerado um robô socialmente consciente, sendo capacitado para a criação de uma interação diferenciada entre ele e o usuário, personalizando a interação por meio do desempenho avançado de tarefas que considera informações sobre o relacionamento estabelecido durante a interação.

Segundo o laboratório Articulab (2017) o Sara considera aspectos socioemocionais, classificados como importantes para melhor colaboração em conjunto com o usuário ao longo do tempo. Seu funcionamento é baseado no uso de inteligência artificial de ponta, com consciência social, que somatiza a ele a capacidade de detecção de comportamentos sociais nas conversas argumentando sobre como responder às intenções por trás desses comportamentos específicos.

O relacionamento que se tenta criar entre Sara e o usuário é um fator que conta para a infraestrutura social e para a criação de estratégias de conversação, despertando assim, os interesses e preferências do usuário. O AVI funciona ainda em plataformas da universidade no qual foi criado, necessitando de um sistema de câmeras para leitura de movimentos dos usuários, para que seja feita a leitura do comportamento do usuário (ARTICULAB, 2017).

Figura 7 – Sara



Fonte: Articulab (2017)

A Siri (Figura 8), pertencente a Apple, foi comprada e lançada em 2010, no sistema operacional dos smartphones da Apple. O smartphone possui um botão home que é um meio de comunicação entre o assistente e o usuário, além da ativação por voz por meio de uma frase ativadora, exibindo assim uma melhor integração com os recursos internos do smartphone (MAIOLINO, 2017).

Por meio dos comandos de voz dados pelo usuário, o assistente faz requisições a uma rede de servidores remotos hospedados na internet para o armazenamento e processamento de dados, em vez de um servidor local ou computador pessoal. Dessa forma, permite a realização de tarefas mais específicas como cálculos matemáticos ou até mesmo uma reserva em um restaurante (APPLE, 2020).

Alguns dos problemas enfrentados pelos usuários ao usar a Siri, são: se a qualidade da internet no dispositivo for de baixa qualidade, ou até mesmo não existir conexão com a internet, a realização de comandos para o assistente não irá acontecer. Outro problema é ser compatível apenas com dispositivos da marca Apple (MAIOLINO, 2017).

Figura 8 – Logotipo da Siri - versão atual



Fonte: Apple (2020)

O Quadro 1 a seguir mostra alguns critérios que resumem os AVIs descritos anteriormente.

Quadro 1 – Quadro Comparativo de AVIs

AVI / Empresa	Sistema(s) Compatível(eis)	Funcionalidades
Alexa (Amazon)	Android, Windows 10, smartphones Apple	Controla dispositivos inteligentes, Permite desenvolver Skills, Executa aplicativos, Permite configurar programas diários
Bixby (Samsung)	Android (apenas alguns modelos de smartphones Samsung)	Auxilia de forma personalizada, Realiza compras por meio da busca de imagens, Oferece detalhes sobre locais próximos com a tecnologia de reconhecimento de imagens
Cortana (Microsoft)	Android, iOS e Windows 10	Executa tarefas organizacionais (configura lembretes, agenda eventos no calendário, etc.), Oferece o Cortana Skills Kit
Google Assistant (Google)	Android 5.1, ou superiores	Cria lembretes, Envia mensagens, Recomenda lugares, Além de retornar mensagens em tempo real, Atua como um assistente doméstico controlando dispositivos domésticos inteligentes compatíveis
Sara (Universidade de Carnegie Mellon)	Computadores da universidade	Responde a comandos de acordo com a leitura de movimentos dos usuários, Realiza tarefas de acordo com o que os usuários pedem considerando seus sentimentos
Siri (Apple)	iOS (apenas modelos de smartphones Apple)	Realiza tarefas específicas como cálculos matemáticos, Reserva em um restaurante, Executa músicas específicas, Executa aplicativos, Controla outros dispositivos

Fonte: Elaborado pelo autor (2020)

## 2.2 Avaliação de Sistemas

A existência de milhares de sistemas, os diversos tipos de funcionalidades e as experiências vividas pelos usuários aumentam a diversidade dos sistemas e com essa diversidade uma possível evolução. A partir disso, deve-se ressaltar a importância da avaliação de sistemas como identificação das necessidades dos usuários ou até mesmo verificar o entendimento dos projetistas sobre essas necessidades, detecção de problemas de interação ou de interface, além do alcance de objetivos quantificáveis em métricas de usabilidade (PRATES; BARBOSA, 2003).

Segundo Rogers *et al.* (2013) a avaliação do sistema permite que problemas sejam detectados e corrigidos antes do sistema entrar em uso e cause algum tipo de prejuízo tanto para o cliente, quanto para os usuários daquele sistema. Para a detecção e correção desses problemas pela avaliação, é necessário levar em consideração o tipo de técnica usada na avaliação do sistema em questão, considerando o ambiente, o envolvimento do usuário e o nível de controle que o avaliador tem sobre o ambiente em que a avaliação ocorre.

O uso de um sistema interativo, é baseado nos seguintes critérios de qualidade de uso que são: usabilidade, experiência do usuário, acessibilidade e comunicabilidade (BARBOSA; SILVA, 2010). Nesta seção, são enfatizados os critérios de usabilidade e de experiência do usuário, visto que são os critérios mais usados e serão os adotados para a avaliação realizada neste trabalho.

A usabilidade é o critério de qualidade que se relaciona com a facilidade de aprendizagem do usuário em relação as funcionalidades do sistema, além de como a interface é usada e a satisfação do usuário em decorrência desse uso (Nielsen, 1993 apud BARBOSA; SILVA, 2010). O ambiente em que o usuário utiliza o sistema também é considerado, pois as características do usuário em relação ao ambiente podem modificar a forma de uso do sistema (BARBOSA; SILVA, 2010). Nielsen (1993 apud BARBOSA; SILVA, 2010, p. 29) define o critério de usabilidade como sendo:

Um conjunto de fatores que qualificam quão bem uma pessoa pode interagir com um sistema interativo. Esses critérios estão relacionados com a facilidade e o esforço necessários para os usuários aprenderem e utilizarem um sistema. Os fatores de usabilidade por ele considerados são: facilidade de aprendizado, facilidade de recordação, eficiência, segurança no uso e satisfação do usuário (Nielsen, 1993 apud BARBOSA; SILVA, 2010, p.29).

Os fatores de usabilidade estabelecidos por Nielsen (1993) determinam fatores que se relacionam tanto aos usuários, quanto aos sistemas, auxiliando em como classificar se o sistema é usável ou não usável.

A experiência do usuário é o critério de qualidade que expõe preocupação com os sentimentos e emoções dos usuários (Sharp *et al.*, 2007 apud BARBOSA; SILVA, 2010). Os sistemas interativos passaram a estar presentes em muitas atividades humanas e em diversos locais, por isso a necessidade de considerar a forma como o uso de um sistema interativo afeta os sentimentos e emoções dos usuários (BARBOSA; SILVA, 2010).

### 2.2.1 Tipos de Avaliação

Barbosa e Silva (2010) apresentam *métodos de inspeção*, *métodos de observação* e os *métodos de investigação* como principais categorias que abrangem os métodos de IHC que serão descritos a seguir:

#### 2.2.1.1 Métodos de Inspeção

Métodos que tentam identificar problemas que os usuários podem vir a ter quando interagirem com o sistema, permitindo a avaliação da conformidade com um padrão ou guia de estilo, além da possibilidade de comparação de designs alternativos (BARBOSA; SILVA, 2010). Geralmente esse método não envolve os usuários, então os *avaliadores* tentam se colocar no lugar de um usuário com determinado perfil, com certo conhecimento e experiência em algumas atividade (BARBOSA; SILVA, 2010). Alertando para o fato de que o avaliador pode se concentrar mais em alguns aspectos de usabilidade do que em outros (Nielsen, 1993). Esse método abrange os seguintes tipos de avaliação:

- **Avaliação Heurística:** método de avaliação de IHC criado para encontrar problemas de usabilidade durante um processo de design iterativo (Nielsen e Molich, 1990; Nielsen, 1993; Nielsen, 1994a apud BARBOSA; SILVA, 2010). Esse método orienta os avaliadores a inspecionar sistematicamente a interface em busca de problemas que prejudiquem a usabilidade (BARBOSA; SILVA, 2010). Tem como base um conjunto de diretrizes de usabilidade, que descrevem características desejáveis da interação e da interface, chamadas de heurísticas, e para a resolução de problemas, é necessário uma categorização em pontual (problema em um único local na interface), ocasional (problema em dois ou mais locais na interface) e sistemático (problema na estrutura geral da interface), para uma melhor avaliação (Nielsen, 1993 apud BARBOSA; SILVA, 2010).
- **Percurso Cognitivo:** método de avaliação de IHC com objetivo de avaliar a facilidade de aprendizado de um sistema interativo, através da exploração da sua interface (Wharton *et al.*, 1994 apud BARBOSA; SILVA, 2010). É motivado pela preferência de muitas pessoas em "aprender fazendo" em vez de depender de treinamentos, considerando principalmente a correspondência entre o modelo conceitual dos usuários e a imagem do sistema, guiando a inspeção da interface pelas tarefas do usuário (BARBOSA; SILVA, 2010).

- **Inspeção Semiótica:** método fundamentado na engenharia semiótica, teoria centrada na comunicação caracterizada como um caso particular de comunicação humana mediada por sistemas computacionais (de Souza, 2005a apud BARBOSA; SILVA, 2010), que avalia a comunicabilidade de uma solução de IHC por meio da inspeção (BARBOSA; SILVA, 2010). O avaliador orienta sua avaliação por meio de signos colocados pelo designer na interface, e segundo Peirce (1992 - 1998, vol. 2, p.326 apud BARBOSA; SILVA, 2010), os signos podem ser definidos como algo que representa alguma coisa para alguém, podendo ser imagens, diagramas, token, etc.. Esses signos tem como objetivo avaliar a qualidade da emissão da metacomunicação do designer codificada na interface, por essa razão é um método que não necessita do envolvimento do usuário (BARBOSA; SILVA, 2010).

### 2.2.1.2 *Métodos de Observação*

Métodos que fornecem dados sobre situações em que os usuários realizam suas atividades, com ou sem apoio de sistemas interativos. Os registros de dados observados, identificam problemas reais que os usuários enfrentam durante a experiência do uso do sistema sendo avaliado. A partir desses registros, é possível coleta de dados mais amplos e ricos sobre a atuação do usuário, obtendo o registro e a análise desses dados que permitem a identificação de problemas reais que os participantes enfrentam, e não apenas problemas potenciais previstos pelo avaliador como em uma investigação por inspeção (BARBOSA; SILVA, 2010). A seguir, são explicados brevemente alguns exemplos deste tipo de avaliação:

- **Teste de Usabilidade:** visa avaliar a usabilidade de um sistema interativo a partir de experiências de uso dos seus usuários-alvo (Rubin, 1994; Rubin e Chisnell, 2008 apud BARBOSA; SILVA, 2010). Os objetivos da avaliação determinam quais critérios de usabilidade devem ser medidos, assim um grupo de usuários é convidado a realizar um conjunto de tarefas usando o sistema em um ambiente real ou controlado como um laboratório, por exemplo. Durante o teste são registrados vários dados sobre o desempenho dos participantes na realização das tarefas e suas opiniões e sentimentos decorrentes de suas experiências de uso (BARBOSA; SILVA, 2010).
- **Avaliação de Comunicabilidade:** visa apreciar a qualidade da comunicação da metamensagem do designer para os usuários (Prates *et al.*, 2000a; de Souza, 2005a; Prates e Barbosa, 2007; de Souza e Leitão, 2009 apud BARBOSA; SILVA, 2010). Tem como fundamentação teórica a engenharia semiótica e avalia a qualidade da recepção dessa metacomunicação. Após a aplicação do método as ações dos usuários são observadas e registradas principalmente em vídeos de interação, tendo como foco os prováveis caminhos de interpretação dos usuários, suas intenções de comunicação e principalmente, as rupturas de comunicação que ocorram durante a interação (BARBOSA; SILVA, 2010).

- Prototipação em Papel: método que avalia a usabilidade de um design de IHC representado em papel, por meio de simulações de uso com a participação de potenciais usuários (Snyder, 2003 apud BARBOSA; SILVA, 2010). Durante a simulação, os usuários falam, fazem gestos ou escrevem para manifestar como desejam interagir com o sistema, sendo possível a avaliar facilmente soluções parciais (de baixa fidelidade) (BARBOSA; SILVA, 2010).
- Teste remoto de usabilidade: tipo de teste de usabilidade conduzido em um ambiente monitorado por especialistas, trazendo comodidade tanto para o participante, quanto para o avaliador (LUIZ AGNER, 2020). Dessa forma, a análise das ações dos usuários em um ambiente mais próximo do natural possibilita que os usuários se sintam mais à vontade, reduzindo a carga horária dos testes e também os custos (HARTSON *et al.*, 1996).

### 2.2.1.3 Métodos de Investigação

Métodos que permitem que o avaliador tenha acesso, interprete e analise concepções, opiniões, expectativas e comportamentos do usuário relacionados com sistemas interativos, permitindo investigar alternativas de design, como eles se apropriam da tecnologia existente e quais são suas expectativas para futuras interações com tecnologias atuais e novas. Geralmente, são métodos usados em etapas iniciais do processo de design, não exigindo que os usuários utilizem um sistema interativo durante a coleta de dados (BARBOSA; SILVA, 2010). Esse método abrange os seguintes tipos de avaliação:

- Entrevista: é uma das técnicas mais utilizadas de coleta de dados e levantamento de requisitos, tratando-se de uma conversa guiada por um roteiro de perguntas ou tópicos na qual um entrevistador busca obter informação de um entrevistado (Seidman, 1998 apud BARBOSA; SILVA, 2010). Dessa forma, é possível uma grande coleta de informações dos usuários de forma precisa, uma vez que a entrevista acontece de forma individual e flexível, por meio de perguntas fechadas ou abertas (Lazar *et al.*, 2010; Sharp *et al.*, 2007) apud BARBOSA; SILVA, 2010). No entanto, é necessário um treinamento dos entrevistadores e se existirem muitas pessoas a serem entrevistadas pode levar bastante tempo para a coleta de dados (BARBOSA; SILVA, 2010).
- Questionário: avaliação feita por meio de um formulário impresso ou on-line com perguntas que os usuários e demais participantes devem responder, a fim de fornecer os dados necessários em uma pesquisa, análise ou avaliação. Assim, é possível uma coleta de dados de um grande número de pessoas, dispersas geograficamente ou não, compondo amostras muito maiores do que com outros métodos. A única ressalva é direcionada ao avaliador, que precisa ser bem experiente para evitar que as perguntas induzam as respostas (BARBOSA; SILVA, 2010).

- Grupo de Foco: reunião de diversas pessoas reunidas por uma ou duas horas em uma espécie de discussão ou entrevista coletiva, guiada por um moderador experiente (BARBOSA; SILVA, 2010). Permite coletar informações sobre quem se tenha pouca informação, podendo ser promovida tanto para gerar ideias, quanto para solucionar problemas, desafios, etc., que podem surgir apenas num contexto social (Lazar *et al.*, 2010; Sharp *et al.*, 2007; Courage e Baxter, 2005 apud BARBOSA; SILVA, 2010). Para que esses grupos focais sejam formados, é necessário um recrutamento de usuários suficientes, o que requer muitos recursos.
- Estudo de Campo: a expressão "estudo de campo" inclui uma categoria ampla de atividades relacionadas com usabilidade que podem incluir investigação contextual, entrevistas no ambiente do usuário e observações simples (BARBOSA; SILVA, 2010). O investigador realiza visitas a usuários finais no seu próprio ambiente e os observa enquanto desempenha uma atividade, sendo a duração dessas visitas bem variáveis, objetivando o entendimento do comportamento do usuário final no contexto do seu próprio ambiente de atuação. Avaliações como essas exigem um nível de esforço mais alto para preparar as visitas, conduzir e analisar os dados (BARBOSA; SILVA, 2010).
- Investigação Contextual: esse tipo de avaliação que objetiva revelar os aspectos da prática do trabalho, onde o avaliador realiza a avaliação considerando o ambiente que o usuário trabalha. Por esse motivo, o avaliador vai até o local de trabalho do usuário e conversa com ele sobre seu trabalho, procurando obter dados mais detalhados sobre a estrutura de trabalho na prática que se tornam habituais e invisíveis (BARBOSA; SILVA, 2010). Essa avaliação adota o modelo mestre-aprendiz, reconhecendo que existe uma dificuldade em descrever práticas de trabalho por parte do usuário. Logo, o entrevistador assume o papel de aprendiz e o usuário assume o papel de mestre, explicando todo o seu trabalho para o entrevistador, gerando um compartilhamento do conhecimento de forma simples e natural (BARBOSA; SILVA, 2010).
- Metodologia para Avaliação da Interação a partir da Linguagem Textual do Usuário (MALTU): esse tipo de avaliação tem o objetivo de avaliar a usabilidade e experiência do usuário de sistemas por meio de um conjunto de Postagens Relacionadas ao Uso (PRUs) (MENDES, 2015).

A seguir é apresentado um Quadro 2 comparativo entre os métodos avaliativos descritos anteriormente, destacando pontos positivos e negativos de cada método.

Quadro 2 – Quadro Comparativo dos Métodos de IHC

Métodos de IHC	Técnicas	Pontos Positivos	Pontos Negativos
Métodos de Inspeção	Avaliação Heurística	Deteção rápida e clara de problemas. Descarte de possíveis confusões dos usuários	Como o teste é realizado em laboratório existe a falta de contexto
	Percurso Cognitivo	Obtenção de resultados objetivos relacionados a aprendizagem do usuário relacionado a tela	Falta de contexto
	Inspeção Semiótica	O avaliador guia a avaliação livre de possíveis informações que destoam do objetivo da avaliação	Poucas informações vindas dos usuários
Métodos de Observação	Teste de Usabilidade	Obtenção de uma grande número de respostas com mais diversidade, encorajando o usuário	Respostas com pouca variedade de diversidade, o que pode afetar os resultados
	Avaliação de Comunicabilidade	Abrange os prováveis caminhos de interpretação dos usuários	Poucas informações vindas dos usuários
	Prototipação em Papel	Fácil realização, pois possui baixo custo e execução	Pouca fidelidade em relação ao sistema original
	Teste Remoto de Usabilidade	O participante fica mais solto, gerando espontaneidade na realização das tarefas	Problemas durante a execução dos testes
Método de Investigação	Entrevista	Encoraja o contato entre os desenvolvedores e os usuários	A avaliação pode ficar cansativa e muito extensa
	Questionário	Alcance de muitas pessoas com poucos recursos	A taxa de resposta é sempre baixa, ou podem não ser o que o usuário deseja
	Grupo de Foco	Destaca áreas de consenso e conflito	Possibilidade de personalidades predominantes
	Estudo de Campo	Obtenção de resultados com enriquecimento de detalhes	As visitas podem ser desgastantes
	Investigação Contextual	Gera compartilhamento natural e simples do conhecimento	As visitas podem ser desgastantes
	MALTU	Obtenção de um grande número de informações com poucos recursos	Pouco contato com os usuários

Fonte: Elaborado pelo autor (2021)

Diante do Quadro 2, que realiza uma comparação entre os Métodos de IHC, foi possível entender os pontos positivos e negativos. A partir disso, considerando o contexto de pandemia de Covid-19 em que estamos inseridos, os métodos selecionados foram a Metodologia MALTU, pois proporciona a coleta de um grande número de opiniões espontâneas dos usuários, não existindo a necessidade de contato com os usuários. E, o Teste Remoto de Usabilidade que proporciona a obtenção de um resultado mais detalhado de uso, não existindo a necessidade de estar com o usuário de forma presencial.

### 2.3 Métodos de Avaliação Utilizados neste Trabalho

Nesta seção são explicados, com mais detalhes, os métodos utilizados neste trabalho.

#### 2.3.1 Metodologia MALTU

Metodologia proposta por Mendes (2015), que propõe uma avaliação textual por meio de Postagens Relacionadas ao Uso (PRUs) a partir de comentários realizados pelos usuários sobre o sistema, relatando algo sobre a usabilidade e a experiência de uso. Essa metodologia segue 5 etapas que são:

1. Definição do contexto da avaliação: é definido o contexto em que o usuário utiliza o sistema, o domínio do sistema e os objetivos da avaliação, considerando três elementos

presentes na definição do contexto, que são: usuários (indivíduos que interagem com o sistema), plataforma (ambiente computacional), ambiente (propriedades do ambiente real);

2. Extração de PRUs: obtenção de PRUs do sistema que está sendo avaliado, podendo ocorrer de duas formas: manual (busca de PRUs a partir de padrões fornecidos em campos de busca do sistema avaliado, bem como a partir de base de dados com postagens do usuário) ou automática (extração com o auxílio de ferramentas).
3. Classificação das PRUs: as PRUs podem ser classificadas em diferentes categorias de classificação sugeridas, que são detalhadas a seguir:
  - Classificação das PRUs: as PRUs podem ser classificadas em diferentes categorias de classificação sugeridas, que são:
    - Por tipo: esta classificação se divide em:
      - a) Crítica: postagens que contém reclamação, erro, problema ou comentário negativo em relação ao sistema;
      - b) Elogio: postagens que contém elogio ou comentário positivo sobre o sistema;
      - c) Dúvida: postagens que contém dúvida ou questionamento sobre o sistema ou suas funcionalidades;
      - d) Comparação: postagens que contém comparação com outro sistema;
      - e) Sugestão: postagens que contém sugestão sobre uma mudança no sistema;
      - f) Ajuda: postagens que contém passos para realizar uma atividade no sistema.
    - Por intenção do usuário: classificação das postagens de acordo com a intenção do usuário sobre o sistema.
      - a) Visceral: postagens caracterizadas por serem geralmente impulsivas, emotivas, com foco em criticar ou elogiar o sistema sem apresentar detalhes sobre ele, sendo mais indicada para a classificação de metas UX;
      - b) Comportamental: postagens mais centradas no uso do sistema, possuindo uma menor intensidade de sentimento, com foco em relatar o problema de forma precisa ou questionar soluções;
      - c) Reflexiva: postagens que possuem intensidade média de sentimento e são mais racionais, indicando valor agregado, motivação ou mesmo uma reflexão sobre o uso do sistema.
    - Por análise de sentimentos: classificação de acordo com os sentimentos expostos nas postagens.
      - a) Por polaridade: postagens que podem ser positivas (demonstram sentimento positivo), neutras (não demonstram sentimento algum), ou negativas (demonstram sentimento negativo); e
      - b) Por intensidade: classificação que classifica o quanto de sentimento (positivo ou negativo) é colocado na postagem.

- Por funcionalidade: classificação que tem um nível detalhamento maior sobre as funcionalidades.
  - Por critério de qualidade de uso: classificação que envolve determinar os critérios de qualidade de uso usabilidade e/ou UX e suas metas como, eficácia, segurança a serem avaliadas.
  - Por artefato: classificação que consiste em identificar o dispositivo (celulares, computadores) que o usuário está utilizando no momento em que realiza a postagem.
4. Interpretação dos resultados: os resultados obtidos são interpretados e avaliados por meio de medidas que são utilizadas pelo avaliador de acordo com a necessidade do avaliador, que consistem na frequência dos resultados de classificação das PRUs e nos relacionamentos entre eles, como:
- A frequência de cada tipo de PRU;
  - A frequência da classificação por intenção do usuário;
  - A frequência por análise da polaridade e intensidade dos sentimentos;
  - A frequência da causa (funcionalidade);
  - A frequência de cada meta de qualidade de uso avaliada;
  - O relacionamento entre a frequência da causa (funcionalidade) com o artefato;
  - Relacionamento de todos os resultados do objetivo de avaliação.
5. Relato dos resultados obtidos: produção de um relatório contendo um compilado de tudo que foi feito durante a avaliação com o uso do modelo, como: contexto de uso do sistema, objetivos e escopo da avaliação, informação da forma de extração e classificação, informações sobre os participantes da avaliação, informação do artefato utilizado, tabelas e gráficos com o resultado das classificações obtidas, uma lista das causas relacionadas às metas avaliadas.

### 2.3.2 *Teste Remoto de Usabilidade*

Nessa avaliação o teste remoto de usabilidade é conduzido em um ambiente monitorado por especialistas. Dessa forma, a análise das ações dos usuários em um ambiente mais próximo do natural possibilita que os usuários se sintam mais à vontade, reduzindo a carga horária dos testes e também os custos (HARTSON *et al.*, 1996). Esse tipo de teste se encaixa dentro do grupo de avaliação remota, que possui algumas ramificações como (HARTSON *et al.*, 1996):

- Avaliação Remota Instrumentada: avaliação em que o avaliador escolhe um aplicativo usa na interface que será avaliado que podem ser instrumentados com código de medição integrado para coletar e retornar um diário ou registro de dados que ocorrem como resultado natural do uso no ambiente de trabalho normal dos usuários (HARTSON *et al.*, 1996). O avaliador usa técnicas de reconhecimento de padrões para deduzir onde ocorrem os

problemas de usabilidade (HARTSON *et al.*, 1996).

- Avaliação Remota Semi-Instrumentada: avaliação em que o avaliador usa a coleta seletiva de dados acionada diretamente pelos usuários que executam tarefas em seu contexto normal de trabalho, ou seja, os usuários podem realizar as tarefas com um treinamento mínimo antes do teste (HARTSON *et al.*, 1996). Geralmente os usuários são treinados para identificar eventos de uso que podem ser negativos ou positivos significativos no desempenho ou na satisfação de suas tarefas (HARTSON *et al.*, 1996). Essa avaliação costuma ter custo-benefício e é capaz de gerar conclusões baseadas em dados qualitativos obtidos da avaliação que determinam quanto o sistema é usável ou não (HARTSON *et al.*, 1996).

### 3 TRABALHOS RELACIONADOS

Nesta seção são apresentados trabalhos que realizaram avaliações em AVIs.

Porcheron *et al.* (2018) realizaram durante um mês, experimentos com usuários por meio de implantações do alto-falante domiciliar Amazon Echo, para estudo da interação sistema-usuário, com o objetivo de explicar como os AVIs são rotineiramente incluídos pelos usuários em conversas cotidianas. Após a implantação dessa tecnologia, e de gravadores em alguns lugares das cinco residências, foram coletados uma grande variedade de áudios, mas somente aqueles áudios em que os usuários incluíam o AVI. A partir da transcrição dos áudios coletados, foi possível analisar as partes em que os usuários iniciavam a inclusão dos AVIs nas situações diárias, além de poder entender como esses usuários desenvolviam um relacionamento de acordo com os retornos dados pelo AVI. A partir disso, foi possível concluir que, as interações entre o sistema e os usuários eram alcançadas de acordo com a disponibilidade de tempo e com as solicitações para a realização de tarefas dos usuários, como forma de incluir o AVI no cotidiano domiciliar.

Ballati *et al.* (2018) compararam três AVIs (Google Assistant, Siri e Cortana) para uma avaliação de recursos em relação a pessoas italianas (nativos) que possuem discurso disártrico induzido pela Esclerose Lateral Amiotrófica (ELA), que impacta na comunicação de forma negativa, como fala arrastada. Esse trabalho procurou definir se pessoas com disartria moderada poderiam ser entendidas pelos AVIs e qual assistente fornecia respostas coerentes quando a fala fosse correta ou parcialmente errada. Para o experimento, foram selecionadas oito pessoas do Hospital Molinette, em específico pacientes com ELA que apresentavam disartria, para coleta das falas por meio de gravações realizadas com o auxílio de um smartphone. Ao todo foram feitas 34 gravações, que foram analisadas e transcritas, obtendo a acurácia do entendimento das falas e da adequação das respostas dos AVIs. Com isso, conclui-se que os três AVIs selecionados possuíam diferenças consideráveis quanto às transcrições das perguntas e aos tipos de respostas dadas. Para os aspectos definidos inicialmente para a avaliação dos AVIs, pessoas com disartria moderada podem ser facilmente compreendidas pelo Google Assistant, tendo como base a quantidade de perguntas transcritas de forma correta (174) que se mostrou superior aos outros dois AVIs, e as taxas de respostas adequadas (54,02%) e incorretas (1,15%). A Siri chegou a ultrapassar o Google Assistant em respostas adequadas, mas obteve a maior taxa de respostas incorretas (9,30%) e por último a Cortana que possuía a maior taxa para respostas padrão (75,93%), deixando evidente que só obtinha bom desempenho para perguntas padronizadas.

Lima *et al.* (2019a) apresentam uma análise empírica da presença de vies na interação por áudio que detectou problemas a partir da comparação entre o Google Assistant e Siri. Os autores fizeram testes com um grupo de 20 pessoas, 10 para o uso do Google Assistant e as outras 10 para o uso da Siri. Os testes com usuários foram realizados por meio da simulação de situações em que os usuários usavam o AVI para a realização de tarefas, as situações foram

criadas pelos avaliadores e os usuários escolhiam ao acaso. Foram utilizadas métricas para avaliação da qualidade do reconhecimento de frases proferidas aos assistentes, percebendo, dessa forma, uma diferença na comparação dos textos lidos entre os sexos, notando-se que o Google Assistant entendeu melhor as frases faladas por pessoas do sexo feminino, diferença essa que não foi relevante para Siri.

Lima *et al.* (2019b) apresentaram um estudo sobre a diferença de desempenho dos AVIs (Google Assistant e Siri) em relação as diferentes falas dos usuários de diferentes regiões e gêneros, podendo gerar um desestímulo à adoção desta tecnologia por esses grupos com falas diferenciadas. Para a realização do experimento foram selecionados por meio de um questionário anônimo, 20 participantes (7 mulheres e 13 homens) onde 10 foram destinados ao uso da Siri e 10 para o Google Assistant. Durante o experimento, as tarefas de leitura (os avaliadores elaboraram cartas com perguntas para que os participantes perguntassem ao AVI) e de comando (os participantes elaboram perguntas para que fossem lidas para o AVI) foram gravadas para que os áudios fossem analisados. Após a coleta dos áudios, foi realizada uma análise de acordo com: a) desempenho (onde cada um dos 20 participantes leu 4 frases, 24 das frases lidas foram lidas pela segunda vez, gerando um total de 115 leituras realizadas), b) transcrição (a causa esperada para a leitura inexata de parte das escritas se referem às pronúncias incorretas feitas pelos participantes) e c) aspectos linguísticos (cacoécia: na primeira tarefa 8 dos 20 participantes cometeram cacoécia, algo que não ocorreu na segunda tarefa. / disfluência: tanto na primeira tarefa, quanto na segunda ocorreram problemas de disfluência em 16 das 83 interações). Após a realização das análises, foi possível concluir que, participantes das regiões Sul (S) e Sudeste (SE) falam com menos cacoécia, o que gerou uma interação mais fluida, por outro lado os participantes da região Norte (N) e Nordeste (NE) tendem a falar com cacoécia e disfluência gerando interações menos fluidas com o AVI.

MOTTA e QUARESMA (2019) realizaram experimentos com usuários brasileiros de smartphones, com o objetivo de entender quais as principais barreiras dos AVIs que impactam negativamente o seu uso. Por meio de estudos baseados na literatura foram definidos os seguintes fatores: percepção de utilidade, percepção de usabilidade, atitude em relação ao uso e risco de privacidade. Esses fatores foram usados como parâmetro para auxiliar na análise dos dados que foram obtidos pela aplicação de um questionário online. A aplicação do questionário teve como foco alcançar o maior número de pessoas e variedade de respostas possível. Por conseguinte, foram obtidas um total de 532 respostas, destas, apenas 257 foram consideradas no resultado, pois eram respostas dos usuários que relataram não utilizar AVIs ou ter frequência de uso esporádica. A partir dessa análise, chegou-se a conclusão de que as principais barreiras no uso de assistentes são as questões de atitude, como: a) preguiça em usar os assistentes; b) o esquecimento da existência desses sistemas; c) o costume pela interface visual e; d) preocupação com outras pessoas ouvindo a interação por voz (vergonha por falar com o celular no momento da interação sistema-usuário). De forma contrária ao que mostra a literatura, a preocupação com a privacidade de dados (medo de perda de informações pessoais) e a usabilidade não se

mostraram como grande barreira no uso dos AVIs.

Larsen *et al.* (2020) investigaram sobre o uso de AVIs da Dinamarca, tendo como foco o entendimento de como os AVIs são utilizados em situações naturais, nas quais o usuário está ocupado com sua tarefa principal, que no caso, é dirigir um carro, e em que grau o AVI é bem sucedido na redução de distrações dos motoristas. Para a realização do experimento, o AVI escolhido foi a Siri, já que é o mais popular da Dinamarca na época em que o estudo foi feito e o dinamarquês foi a língua selecionada para o experimento. Foram selecionados 24 participantes que atendiam aos requisitos estabelecidos, como carteira de motorista, e antes de iniciarem os testes, responderam a um questionário e realizaram um pequeno treinamento. Os testes foram realizados em dois tipos de ambiente, para análise dos tipos de interação, onde 16 participantes foram destinados a executar tarefas dentro do carro, 8 para o uso da Siri e 8 usando o celular manualmente. O outro tipo de interação ocorreu em laboratório, usando um equipamento de realidade virtual, onde os outros 8 participantes foram destinados tanto ao uso da Siri, quanto de o uso do celular manualmente usando a Siri, após a conclusão do experimento houve uma entrevista pós-teste. A partir da análise dos resultados foram definidos alguns critérios de usabilidade que viessem a influenciar nos resultados de usabilidade da Siri, como: a) eficiência (análise dos tempos de conclusão das tarefas e contagem de etapas), b) efetividade (análise dos vídeos e gravações de monitoramento de conclusão ou abandono das tarefas), c) aprendizagem (tarefas passadas aos participantes de forma aleatória para redução de influência, fadiga ou efeitos de aprendizado), d) satisfação (análise de expressões faciais nos vídeos de monitoramento e na entrevista pós-teste), e) distração (análise das experiências durante todo o teste como forma de definir a experiência menos perturbadora). Dessa forma, foi possível concluir que a Siri ainda não é a alternativa mais segura quando comparada com o uso manual do celular na direção, pois, quando os resultados dos tipos de interação foram comparados, a usabilidade da Siri foi menor que a usabilidade na interação por conta de problemas na usabilidade, como esforço mental para o uso de execução de tarefas.

O Quadro 3 relata os trabalhos relacionados a este trabalho.

Quadro 3 – Quadro Comparativo dos Trabalhos Relacionados

Trabalho / Autor	Objetivo	Método	Critério Avaliado	Usuários	Resultados
Interfaces de Voz na Vida Cotidiana (PORCHERON et al., 2018)	Investigar como o AVI Amazon Echo é rotineiramente incluído pelos usuários nas conversas cotidianas (tabular)	Estudo de Campo (implantação do AVI e de gravadores)	Usabilidade (eficácia, eficiência e utilidade) Comunicabilidade	Ingleses	A interação entre o sistema e o usuário era alcançada por meio de situações com limitação de tempo em que os usuários realizavam solicitações para realização de tarefas para a inclusão do AVI no cotidiano doméstico.
Assessing Virtual Assistant Capabilities with Italian Dysarthric Speech (BALLATI et al., 2018)	Investigar se pessoas com disartria moderada podem ser entendidas pelos AVIs (Google Assistant, Siri e Cortana) selecionados pelo trabalho e qual desses AVIs fornece uma resposta coerente quando a fala reconhecida estiver correta ou parcialmente errada	Estudo de campo (gravações das falas dos participantes)	Acessibilidade Comunicabilidade	Italianos (nativos)	O Google Assistente foi o AVI que melhor entendeu os participantes e que respondeu de forma precisa às perguntas.
Análise Empírica do Viés em Assistentes Pessoais Baseados em Voz (LIMA et al., 2019)	Analisar a diferença na interação sistema-usuário por áudios a partir da comparação entre AVIs (Google Assistant e Siri)	Testes com usuários Questionário	Usabilidade	Brasileiros	Diferença na comparação dos textos lidos entre os sexos, onde o Google Assistant entendeu melhor frases faladas por pessoas do sexo feminino e a Siri.
Análise de Discriminação de Assistentes de Voz Inteligentes (LIMA et al., 2019)	Identificar uma diferença na qualidade do desempenho dos AVIs (Google Assistant e Siri) em relação às formas específicas de falas de diferentes grupos de pessoas	Estudo de Campo (gravações das falas dos participantes)	Usabilidade (eficiência, eficácia)	Brasileiros	Os participantes das regiões Norte e Nordeste tendem a falar com mais cacofonia e disfluência que os participantes das regiões Sul e Sudeste, gerando interações menos fluidas com os AVIs. As principais barreiras no uso de assistentes são as questões de atitude, como preguiça em usar os assistentes, o esquecimento da existência desses sistemas, o costume pela interface visual e a preocupação com outras pessoas ouvindo a interação por voz.
Barreiras no Uso de Assistentes de Voz por Usuários Brasileiros de Smartphones (MOTTA e QUARESMA, 2019)	Entender quais as principais barreiras no uso dos assistentes por voz em geral que impactam negativamente em seu uso	Questionário	Usabilidade	Brasileiros	A usabilidade da Siri foi menor quando comparada com a interação manual, por conta de problemas na usabilidade da Siri.
Mãos livres, mas não os olhos: Uma avaliação da Siri Enquanto Dirige (LARSEN et al., 2020)	Entender como os AVIs são utilizados a Siri é utilizada em situações naturais como dirigir um carro e em que grau esse AVI é bem sucedido em relação a redução de distrações de motoristas	Estudo de Campo	Usabilidade (utilidade, satisfação do usuário)	Dinamarqueses	A interação do usuário com a assistente ainda é problemática, por conta de interações não naturais e feedbacks não satisfatórios
Uma Avaliação da Usabilidade e Experiência do Usuário em um Assistente por Voz Inteligente (Este trabalho)	Apresentar resultados de uma avaliação da usabilidade e experiência do usuário em AVIs no contexto da linguagem	Testes com Usuários Questionários Avaliação Textual	Usabilidade Experiência do Usuário	Brasileiros	

Fonte: Elaborado pelo autor (2020)

## 4 APLICAÇÃO DA METODOLOGIA MALTU

Nesta seção é apresentado o procedimento metodológico da aplicação da Metodologia MALTU (Figura 9), e logo em seguida uma descrição de cada fase da aplicação da metodologia.



Fonte: Elaborado pelo autor (2020)

### 1. Definição do Contexto de Uso e do AVI

O AVI avaliado nesta fase do trabalho foi o Google Assistant, um sistema de comandos por voz que atua como um assistente pessoal, gerenciando tarefas de acordo com as necessidades dos usuários, sendo considerado o assistente pessoal mais ativo no Brasil em dispositivos móveis (STARTSE, 2019). A análise das postagens mais recentes, os usuários se caracterizam por utilizarem o sistema operacional Android 5.0 ou superior, e dispositivos móveis do tipo celular e tablet. O objetivo da definição do contexto descrito anteriormente, é conseguir identificar problemas na interação e na interface, avaliar a satisfação dos usuários com o sistema, e identificar a sugestão de novas funcionalidades para o sistema.

### 2. Extração das Postagens

Nesta fase, a extração de postagens foi realizada de forma automática, com o auxílio da ferramenta <sup>1</sup> UUX-Posts. A base de dados utilizada foi a loja de aplicativos Google Play, onde o AVI é encontrado. A extração foi realizada no dia 20/05/2020, obtendo um total de 3394 postagens relacionados ao Google Assistant.

- **Sentencição**

Processo realizado após o processo de extração das postagens, que consiste na divisão de cada frase em trechos de acordo com a exclusão de conjunções adversativas presentes na seção de anexos deste trabalho (Anexo A), com o objetivo de conseguir classificações únicas para postagens com significados opostos, como por exemplo: "É um app maravilhoso, mas cadê o deezer nas opções de músicas."; nesta PRU, temos um elogio e uma crítica, ou seja, um trecho com uma polaridade positiva e

<sup>1</sup> UUX-Posts <<http://uuxposts.russas.ufc.br/>>. Acessado em 19 set. 2020.

outra negativa. Com a aplicação desse processo, foi possível obter um total de 4938 trechos de postagens para serem classificadas em PRUs e não-PRUs. Nos quadros a seguir são mostrados alguns exemplos de postagens antes (Quadro 4) e depois (Quadro 5) do processo de sentença.

Quadro 4 – Quadro de Postagens Antes do Processo de Sentença

Data	Postagem
9 de maio de 2020,	"Gosto mas acho que pode ser possível melhorar o envio feedback constante";;
19 de junho de 2019,	"É um app maravilhoso, mas cadê o deezer nas opções de músicas.";;
17 de março de 2019,	"Poderia não precisar do comando de voz ok google para cada solicitação! mas eu amo a assistente!";;

Fonte: Elaborado pelo autor (2020)

Quadro 5 – Quadro de Postagens Depois do Processo de Sentença

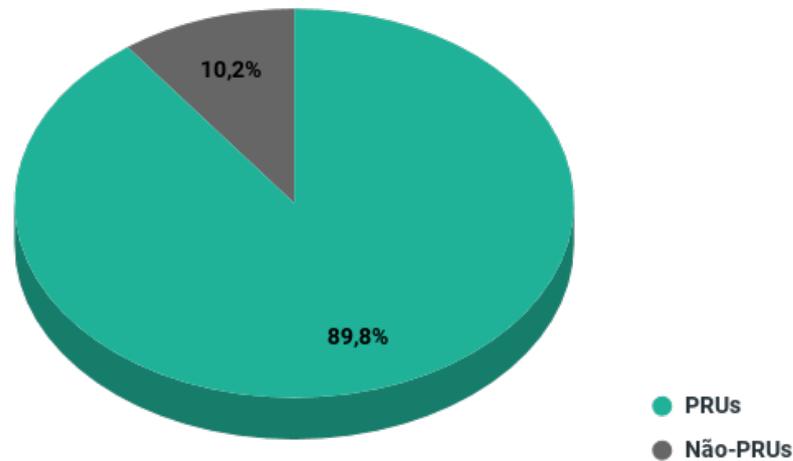
Data	Postagem	Classificação
9 de maio de 2020	Gosto [.]	PRU
9 de maio de 2020	acho que pode ser possível melhorar o envio feedback constante[.]	PRU
19 de junho de 2019	É um app maravilhoso[.]	PRU
19 de junho de 2019	cadê o deezer nas opções de músicas.";;	PRU
17 de março de 2019	"Poderia não precisar do comando de voz ok google para cada solicitação![.]	PRU
17 de março de 2019	eu amo a assistente!";;	PRU

Fonte: Elaborado pelo autor (2020).

### 3. Classificação das Postagens

Nesta fase, a autora deste trabalho juntamente com outra avaliadora, também participante do projeto, deram início as classificações manuais no dia 17/07/2020, com término no dia 24/07/2020. Com o total de 4938 trechos de postagens, foram obtidas 4436 PRUs (89,8%), como exemplo, "o reconhecimento por voz não funciona direito", e 502 Não-PRUs (10,2%), como exemplo, "Por favor, arrumem isso!", com o objetivo de eliminar trechos de postagens não relacionadas ao uso. Na Figura 10 é mostrado um Gráfico com as porcentagens de cada tipo.

Figura 10 – Gráfico de PRUs e Não-PRUs



Fonte: Elaborado pelo autor (2020)

#### 4. Classificação das PRUs em Categorias

Nesta fase, foi realizada a classificação dos 4436 trechos de postagens de forma manual, iniciada no dia 13/09/2020 com término no dia 22/09/2020. Essa classificação consistiu em categorizar cada PRU a partir das categorias estipuladas pela MALTU (MENDES, 2015) que são exemplificados nos quadros a seguir. No Quadro 6 são apresentados exemplos de classificação por tipo. No Quadro 7 são apresentados exemplos de classificação por intenção do usuário. No Quadro 8 são apresentados exemplos de classificação por informação de sentimento. No Quadro 9 são apresentados exemplos de classificação por critérios de qualidade de uso em facetas de usabilidade. No Quadro 10 são apresentados exemplos de classificação por critérios de qualidade de uso em facetas de UX. No Quadro 11 são apresentados exemplos de classificação por tipos de funcionalidades de acordo com o AVI. No Quadro 12 são apresentados exemplos de classificação por tipo de plataforma.

Quadro 6 – Exemplos de Classificação por Tipo

Tipo de Mensagem	Exemplos de Postagens
Crítica	“O reconhecimento por voz não funciona direito”.
Elogio	“Gosto do aplicativo e da funcionalidade que ele oferece”.
Ajuda	Para ativar nos celulares one vision, você tem que desativar o modo economia de energia, espero ter ajudado alguém”.
Dúvida	“Tem como funcionar com a tela desligada?”.
Comparação	“Pisa na Siri e na Bixby”.
Sugestão	“Vocês deveriam criar um nome a ela, invés de se Google assistente seriam útil”.

Fonte: Elaborado pelo autor (2021)

Quadro 7 – Exemplos de Classificação por Intenção do Usuário

Intenção do Usuário	Exemplos de Postagens
Visceral	“Muito bom a google assistente é minha amiga”.
Comportamental	“O assistente ativa mesmo quando não falamos ok google”.
Reflexiva	Poderiam colocar novamente a opção de tradução de tela, que existia há uns 3/4 anos atrás, seria ótimo”.

Fonte: Elaborado pelo autor (2021)

Quadro 8 – Exemplos de Classificação por Informação do Sentimento - Polaridade

Informação do Sentimento - Polaridade	Exemplos de Postagens
Positivo	"Gosto do aplicativo e da funcionalidade que ele oferece".
Neutro	"Poxa não tem cortana tem que se contentar com isso aí".
Negativo	"Lumus/Nox não estão funcionando."

Fonte: Elaborado pelo autor (2021)

Quadro 9 – Exemplos de Classificação por Qualidade de Uso - Usabilidade

Qualidade de Uso - Usabilidade	Exemplos de Postagens
Eficácia	"O reconhecimento por voz não funciona direito".
Eficiência	"Gosto do aplicativo e da funcionalidade que ele oferece."
Satisfação	"É muito bom mesmo e bom que para pesquisar kpop e etc.."
Segurança	"O assistente ativa mesmo quando não falamos Ok Google."
Utilidade	"É incrível eu adorei o unico defeito é que não funciona com a tela bloqueada."
Memorabilidade	"E acho que deveria dar pra salvar nossos comandos favoritos pois as vezes é difícil lembrar."
Aprendizado	"Como faz pra entra no Google assistente quero muito alguém me ajuda."

Fonte: Elaborado pelo autor (2021)

Quadro 10 – Exemplos de Classificação por Qualidade de Uso - UX

Qualidade de Uso - UX	Exemplos de Postagens
Satisfação	"Ótimo."
Afeto	"Google assistente é a pessoa mais gentil do mundo eu falei : google vc minha amiga? Ela falou : vc é minha melhor amiga."
Confiança	"Melhor assistente e confiável que uso."
Estética	"Fica horrível para de mexer por causa que fica uma borda verde ondeeu tenho que clicar daí eu sempre tenho que mudar repete tudo que eu falo."
Frustração	"Não entende muita coisa."
Motivação	"Eu não conheço Boa tarde para você Leila flor beijos."
Suporte	"Não tem suporte nem atualização pra explicar o que está dando problema."

Fonte: Elaborado pelo autor (2021)

Quadro 11 – Exemplos de Classificação por Funcionalidade

Funcionalidade	Exemplos de Postagens
Reconhecimento por Voz	"O reconhecimento por voz não funciona direito. "
Pesquisa	"No Android 9 só faz pesquisa. "
Fazer ligação	"Adorei o aplicativo, pois ele ajuda muito nas tarefas telefônicas basta dizer: ok google e fazer com comandos de voz ligações me ajuda muito na questão que é fácil e rápido!"

Fonte: Elaborado pelo autor (2021)

Quadro 12 – Exemplos de Classificação por Plataforma

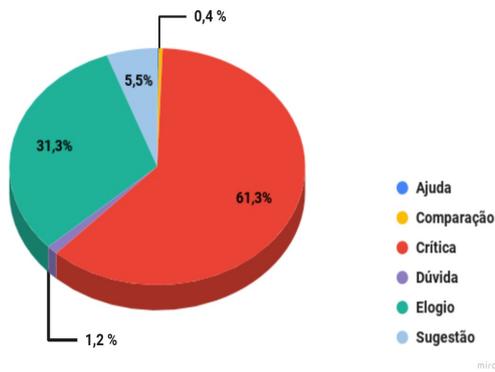
Plataforma	Exemplos de Postagens
Celular	"Para ativar nos celulares one vision, vc tem que desativar o modo econômica de energia, espero ter ajudado alguém."
Tablet	"Funciona no meu tablet Samsung (Android 5) , funciona no zenfone (Android 8.1), não funciona nem com reza no Android 9."

Fonte: Elaborado pelo autor (2021)

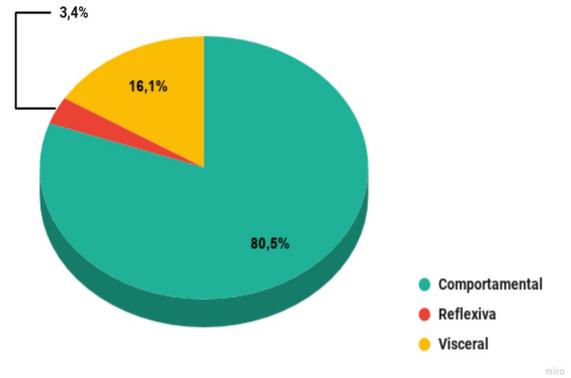
## 5. Interpretação dos Resultados

Os resultados apresentados nesta fase, foram gerados a partir da classificação de duas planilhas (foram realizadas por duas avaliadoras) validadas por dois participantes do projeto, que são orientadores dos demais membros do projeto, dividindo tarefas e guiando nas atividades. Os Gráficos e Tabelas apresentados a seguir, demonstram as relações entre as classificações de cada categoria, exibindo uma interpretação do sistema avaliado. A Figura 11(a) mostra o Gráfico que representa os tipos de mensagens enviadas pelo usuário, o qual o tipo crítica obteve o maior percentual (61,3%). A Figura 11(b) é apresentado o Gráfico que representa a classificação da Intenção de sentimento, onde o tipo comportamental obteve o maior percentual (80,5%). Na Figura 11(c) é apresentado

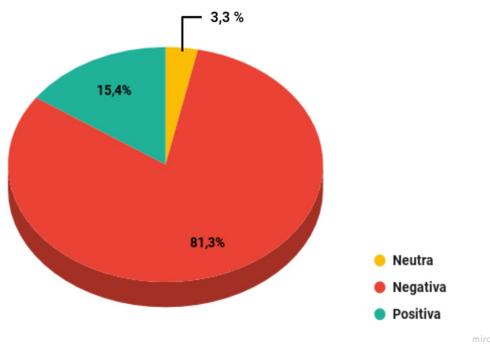
o Gráfico que representa a classificação da Análise do sentimento, onde o tipo negativa obteve o maior percentual (69,6%). Na Figura 11(d) é apresentado o gráfico que representa a classificação em Qualidade de uso por usabilidade, onde a faceta eficiência obteve o maior percentual (71,8%). Por fim, na Figura 11(e) é mostrado um gráfico que representa a classificação em qualidade de uso por UX, onde a faceta de frustração obteve o maior percentual com (68,5%).



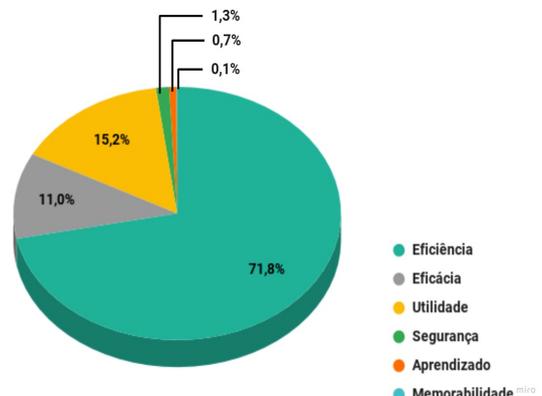
(a) Gráfico de Classificação por Tipo



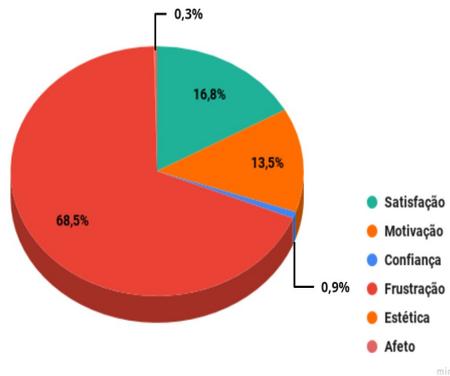
(b) Gráfico de Intenção do Usuário



(c) Gráfico de Análise de Sentimento



(d) Gráfico de Qualidade de Uso - Usabilidade



(e) Gráfico de Qualidade de Uso - UX

Figura 11 – Gráficos da Avaliação Textual

Fonte: Elaborado pelo autor (2021)

No Quadro 13 são apresentadas funcionalidades coletadas das PRUs classificadas por tipo. No Quadro 14 são apresentadas funcionalidades com os percentuais do tipo crítico em algumas facetas de usabilidade. No Quadro 15 são apresentadas funcionalidades do tipo crítico em facetas de UX.

Quadro 13 – Classificação por Tipo x Funcionalidades

Classificação por Tipo	Funcionalidades
Crítica	idioma, voice match, reconhecer por voz, reconhecer voz, responder, comando de voz, ativar sozinho, entrar, abrir sozinho, ativar, abrir, configurar voz
Elogio	conversar, pesquisar, responder, piadas, contar piadas, ativar, tradutor, reconhecer voz, idioma, fazer pesquisa
Sugestão	mudar nome, funcionar offline, dar nome, mudar comando, tocar música, colocar nome, mudar voz, conversar, funcionar com tela bloqueada

Fonte: Elaborado pelo autor (2021)

Quadro 14 – Critérios de Qualidade de Uso - Usabilidade x Funcionalidades

Porcentagem	Faceta de Usabilidade	Funcionalidades
71,8 %	Eficiência	idioma, entrar / reconhecer por voz, voice match, comando de voz, responder, ativar sozinho, abrir sozinho, ativar
15,2 %	Utilidade	ativar sozinho, comando de voz
11 %	Eficácia	abrir, configurar voz
1,3 %	Segurança	aparecer, ativar microfone
0,7 %	Aprendizado	acessar por voz, apagar histórico

Fonte: Elaborado pelo autor (2021)

Quadro 15 – Critérios de Qualidade de Uso - UX x Funcionalidades

Porcentagem	Faceta de UX	Funcionalidades
68,5%	Frustração	voice match, responder, reconhecer voz, reconhecer comando
16,5 %	Satisfação	ativar, clicar no microfone, reconhecer música
13,5 %	Motivação	abrir com tela bloqueada, acessar rede, alterar voz

Fonte: Elaborado pelo autor (2021)

E por fim, na Tabela 1 são apresentadas algumas funcionalidades sugeridas pelos usuários nas PRUs.

Tabela 1 – Sugestões de Funcionalidades

Tipo de PRU = sugestão
Mudar nome, Funcionar offline, Dar nome, Colocar nome, Mudar comando, Escolher nome

Fonte: Elaborado pelo autor (2021)

## 6. Relato dos Resultados

A avaliação textual aqui apresentada e detalhada, possui um relatório de avaliação, presente no Anexo A, que contém mais detalhes de todo o processo da metodologia. Diante dos Gráficos e Tabelas apresentados na Interpretação dos Resultados, é possível chegar a

conclusão de que diante do alto percentual de críticas apresentada no Gráfico 11(a), e das funcionalidades mais criticadas, entre elas a que mais se destaca é relacionada à interação, o que acaba influenciando na usabilidade, que no caso é o idioma. Baseado nisso foi possível perceber que as críticas feitas ao idioma estão diretamente relacionadas aos comandos por voz dados ao AVI, e por conta da falta de feedback, ou de comandos entendidos de forma errônea o tipo comportamental tem um alto percentual (Gráfico 11(b)). Como consequência disso, os usuários acabam desinstalando o aplicativo do assistente.

## 5 TESTE REMOTO

Nesta seção é apresentado o procedimento metodológico da aplicação da Avaliação Remota (Figura 12) e nele, já a execução da avaliação.



Fonte: Elaborado pelo autor (2021)

### 1. Definição do Contexto da Avaliação

Nesta fase, a ferramenta escolhida e o número de usuários, foram definidos pensando no cenário de pandemia de Covid-19, e por esse motivo a ferramenta Zoom <sup>1</sup> é a que melhor se enquadra, pois, com ela é possível realizar a gravação das reuniões. No momento anterior à reunião foi realizado um momento para tentar descontrair o usuário e deixar claro para ele que o teste feito era uma avaliação do Google Assistant. E assim, a medida que os usuários iam realizando as tarefas dentro dos cenários descritos iam apresentando seus resultados aos avaliadores. Já o número de usuários foi escolhido de acordo com as práticas realizadas na literatura produzidas por Nielsen (2000), Virzi (1992) e Lewis (1994), que afirmam que 5 participantes rendem 80% das descobertas de usabilidade (BEVAN *et al.*, 2003). A seguir são apresentadas informações relacionadas a dados pessoais dos participantes (sexo-feminino ou masculino, idade, escolaridade- grau de formação e tempo de atuação-quantos anos possui em sua área de trabalho) no Quadro 16.

Quadro 16 – Quadro de Perfil dos Usuários

#	Sexo	Idade	Escolaridade	Tempo de atuação
P1	F	34 anos	Graduada	12 anos
P2	F	52 anos	Graduada	26 anos
P3	F	42 anos	Graduada	5 anos
P4	F	50 anos	Graduada	5 anos
P5	F	46 anos	Graduada	10 anos

Fonte: Elaborado pelo autor (2021)

<sup>1</sup> Zoom <<https://zoom.us/pt-pt/meetings.html>>

## 2. Definição das Funcionalidades

Nesta fase, foram selecionadas funcionalidades do tipo crítica que mais vezes foram citadas entre os usuários, e podem ser vistas no Quadro 13. Estas serviram para o desenvolvimento de cenários de uso do Google Assistant, como forma de estabelecer um foco do contexto de execução dos testes. No Quadro 17 estão descritas essas funcionalidades selecionadas.

Quadro 17 – Quadro de Descrição das Funcionalidades

Funcionalidades	Descrição
Idioma	Suporte apenas para português brasileiro e inglês e não para português de Portugal
Reconhecer por Voz	O assistente não reconhece a voz cadastrada no dispositivo, e por isso, não executa comandos
Abrir / Ativar / Entrar	O usuário tenta dar comandos de acesso a outros aplicativos do dispositivo e não consegue
Voice Match	Problema na funcionalidade que permite vincular a voz aos aplicativos do dispositivo
Ativar e/ou Abrir sozinho	O comando de ativação não é falado, mas mesmo assim o assistente ativa e/ou abre sozinho
Responder	Não existência de feedback ao comando dado
Comando de voz	Execução de comandos de voz realizados de forma errônea ou não execução dos comandos dados
Configurar voz	O usuário tenta configurar a voz ao dispositivo e não consegue

Fonte: Elaborado pelo autor (2021)

No Quadro 18 são apresentados as porcentagens e os valores quantitativos de cada funcionalidade criticada e descrita anteriormente, bem como trechos de PRUs como exemplos.

Quadro 18 – Quadro de Funcionalidades mais Criticadas

Porcentagem / Valor	Funcionalidade	Exemplos de Trechos de Postagem
23,3% / 91	Idioma	"O Ok Google não funciona em meu celular, fala que não está disponível para minha linguagem, já mudei pra quase todas as línguas e não funciona! "
19,9% / 78	Reconhecer por voz	"Péssimo para ouvir, tente fazê-lo tocar Someone New e ele só vai ouvir Samuel que não tem nada a ver, está péssimo para reconhecer."
15,6% / 61	Abrir / Ativar / Entrar em outros aplicativos	"Não abre o que peço, peço pra abrir musicas do youtube ele vai pro rádio, mesmo eu falando."
11,5% / 45	Voice match	"A única coisa que crítico no aplicativo é que eu não acesso ao voice match, não consigo realizar procedimento de vincular a voz pelo aplicativo e pelo meu dispositivo."
11,5% / 45	Ativar e/ou Abrir sozinho	"Há um pequeno erro que incomoda no qual o aplicativo acaba abrindo sozinho sem minha permissão."
8,4 % / 33	Responder	"Quero saber por que falo com o Google e ele não me responde."
6,6 % / 26	Comando de voz	"Ela não reage corretamente a comandos, por exemplo Desligar Wi-fi, ela não desliga."
3,1 % / 12	Configurar voz	"Simplesmente sumiu todas minhas configurações e não grava mais minha voz."

Fonte: Elaborado pelo autor (2021)

Com a definição das funcionalidades apresentadas, foram criados storyboards, ou seja as funcionalidades serviram como base para a criação desses cenários de uso, com o objetivo de guiar as avaliadoras no momento de execução dos testes (Figura 13). No primeiro cenário o usuário dá o comando de ativação "Ok Google!" e com isso, dá o comando de desligamento do Wi-fi. No segundo cenário o usuário dá o comando de ativação "Ok Google!" e com isso, dá o comando para o assistente abrir outro aplicativo, no caso, o whatsapp. A seguir são apresentados no Quadro 19.

Quadro 19 – Cenários x Funcionalidades

#	Cenários	Funcionalidades
1	Desligar Wi-Fi	idioma, voice match, reconhecer por voz, abrir/ativar/entrar, responder, comando de voz, configurar voz
2	Abrir um aplicativo - WhatsApp	idioma, abrir/ativar/entrar, responder, comando de voz

Fonte: Elaborado pelo autor (2021)

Figura 13 – Storyboards dos Cenários



(a) Cenário 1

(b) Cenário 2

Fonte: Elaborado pelo autor (2021)

### 3. Planejamento dos Testes com Usuários

Nesta fase, tive a participação de duas avaliadoras, uma avaliadora para guiar os participantes (a autora deste trabalho), e outra atuando como observadora (participante do projeto DICAVI), realizando anotações sobre as reações e sugestões dos participantes, sendo as duas alunas do curso de Engenharia de Software na Universidade Federal do Ceará do campus de Russas. Foram criados um termo de consentimento (Anexo B), para esclarecer alguns pontos sobre como o teste seria realizado e sobre a utilização dos dados dos participantes, e um roteiro de testes (Anexo C) com a descrição de cada momento, as ferramentas necessárias, possíveis problemas de execução que poderiam acontecer, objetivando prevenir as avaliadoras de qualquer situação que impedisse ou atrapalhasse a

execução dos testes . O número de tentativas da execução de cada cenário foi deixado a escolha do participante, objetivando não causar nenhum tipo de desconforto ou exaustão. Durante a execução dos testes foi possível considerar que nenhuma das participantes desistiu de realizar os testes e conseguiram executar os comandos fora do aplicativo da assistente, por esse motivo a existência do aplicativo do assistente instalado no celular se fez necessária. Cada participante levou de 3 a 6 minutos para a execução dos dois cenários. Após a execução dos cenários foi realizada uma entrevista aberta com o objetivo de captar as sugestões ou alguma dificuldade que o usuário queira relatar.

#### 4. Aplicação dos Testes com Usuários

Nesta fase, após a execução do roteiro de testes, foi dito que existiam dois cenários de uso para a execução, e a explicação de cada cenário foi realizada de forma separada e no momento anterior da execução, para que o usuário não se confundisse ao executar os cenários. Foi possível observar que as participantes levaram de 3 a 6 minutos para a execução dos cenários de uso. No Quadro 20 são apresentados categorias com os tempos de cada momento, as categorias do quadro são: a) tempo para entender os cenários (explicação e tira-dúvidas dos cenários 1 e 2); b) tempo de execução do cenário 1 (tempo de execução do cenário 1); c) tempo para o cenário 2 (tempo de execução do cenário 2); d) número de tentativas do cenário 1 (número de vezes que os usuários tentaram executar o cenário de uso 1); e) número de tentativas do cenário 2 (número de vezes que os usuários tentaram executar o cenário de uso 2); f) tempo total de execução do teste (adição dos tempos de entendimento do cenário, bem como o tempo de execução do cenário 1 e do cenário 2).

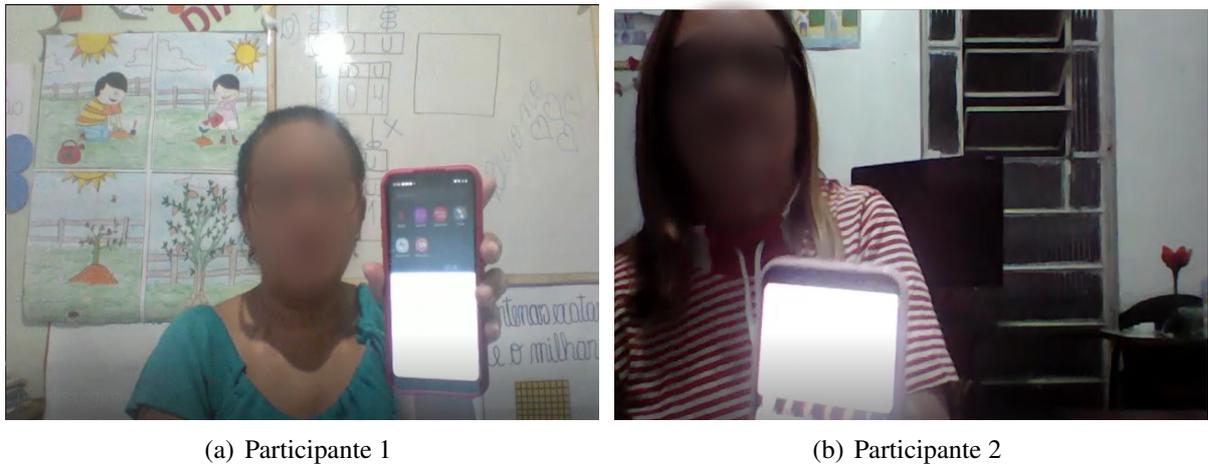
Quadro 20 – Quadro com os Tempos de Execução dos Testes

#	Tempo para entender os cenários	Tempo para executar o cenário 1	Número de tentativas de execução do cenário 1	Tempo para executar o cenário 2	Número de tentativas de execução do cenário 2	Tempo de execução total do teste
P1	2 min	01 min 44 seg	2 tentativas	1 min	2 tentativas	04 min 44 seg
P2	01 min 24 seg	01 min 35 seg	3 tentativas	1 min	3 tentativas	03 min 59 seg
P3	01 min 42 seg	45 seg	1 tentativa	04 min 01 seg	4 tentativas	06 min 28 seg
P4	2 min	01 min 49 seg	3 tentativas	01 min 13 seg	1 tentativa	05 min 02 seg
P5	2 min	01 min 23 seg	5 tentativas	47 seg	1 tentativa	04 min 10 seg

Fonte: Elaborado pelo autor (2021)

No Teste Remoto proposto, as participantes são professoras de uma escola particular de Teresina, cidade do estado Piauí. Foram contactadas, por meio da autora deste trabalho que ainda mantém uma relação com o local que estudou. As participantes se enquadram, pois possuem as ferramentas específicas (notebook e celular) e porque já tinham experiência com reuniões remotas por estarem dando aula durante o momento da pandemia de Covid-19. Durante a execução de algumas, por conta da curiosidade em aprender novos comandos, foi possível até mesmo executar outros cenários de uso. Sendo assim possível capturar as ideias e sugestões durante a entrevista aberta, além das observações em relação as interações de uso relatadas pelas avaliadoras. A seguir são apresentadas imagens das participantes durante o momento de execução dos testes (Figura 14).

Figura 14 – Imagens Durante a Execução dos Testes



(a) Participante 1

(b) Participante 2

Fonte: Elaborado pelo autor (2021)

## 5. Interpretação dos Resultados

Nesta fase, as gravações foram analisadas e com isso foi possível relatar elogios, críticas e sugestões, como as PRUs categorizadas por tipo na metodologia MALTU, bem como relatos de dificuldades durante o uso. Esses relatos estão descritos a seguir na ordem em que os participantes foram executando os testes. As descrições qualitativas estão presentes no Quadro 21 - participante 1, no Quadro 22 - participante 2, no Quadro 23 - participante 3, no Quadro 24 - participante 4.

Quadro 21 – Quadro Descritivo do Teste - Participante 1

Participante 1	
Observação (ões) do aplicativo do assistente	O assistente não respondeu aos comandos do usuário fora do aplicativo do assistente
Observação (ões) da participante	Cenário 2 - A participante falou whatsapp e a assistente abriu a loja de aplicativos Google Play Store para que o participante baixasse o aplicativo, que já estava presente no aparelho
Elogio (s)	"Eu tô amando!"
Crítica (s)	"Mas eu sempre tenho que ir lá!" "Se não fosse isso tava ótimo!"
Sugestão (ões)	Sugeriu que a assistente estivesse vinculada dentro do whatsapp e que a assistente ativasse sem a necessidade de ir até o aplicativo da assistente
Dificuldade(s)	Não se aplica

Fonte: Elaborado pelo autor (2021)

Quadro 22 – Quadro Descritivo do Teste - Participante 2

Participante 2	
Observação (ões) do aplicativo do assistente	O assistente não respondeu aos comandos do usuário fora do aplicativo do assistente
Observação (ões) da participante	Não aplica
Elogio (s)	"Muito bom o aplicativo, pois ele facilita muito!" "Ele facilita muito quando a gente quer usar o aparelho!"
Crítica (s)	"Eu tive que buscar o microfone para poder falar, se não precisasse do microfone seria melhor."
Sugestão (ões)	Sugeriu não ter que ir até o aplicativo do assistente para usá-la
Dificuldade(s)	Não se aplica

Fonte: Elaborado pelo autor (2021)

Quadro 23 – Quadro Descritivo do Teste - Participante 3

Participante 3	
Observação (ões) do aplicativo do assistente	O assistente não respondeu aos comandos do usuário fora do aplicativo do assistente e além de ativar sozinho sem a participante falar deu a definição da palavra ligar
Observação (ões) da participante	Cenário 2 - A participante deu o comando "ligar o whatsapp" e a assistente não abriu o aplicativo whatsapp, mas perguntou a participante para quem ela queria ligar
Elogio (s)	"Muito bom o aplicativo, pois ele facilita muito!" "Ele facilita muito quando a gente quer usar o aparelho!"
Crítica (s)	Não se aplica
Sugestão (ões)	"Engraçado que ela pode ser um dicionário não é?!"
Dificuldade(s)	A participante tentou usar o aplicativo offline

Fonte: Elaborado pelo autor (2021)

Quadro 24 – Quadro Descritivo do Teste - Participante 4

Participante 4	
Observação (ões) do aplicativo do assistente	O assistente não respondeu aos comandos do usuário fora do aplicativo do assistente, além de ativar sem ser solicitado
Observação (ões) da participante	Cenário 1 - A participante relatou que após duas tentativas de execução do comando, a assistente não respondeu ao comando dela, e após essas tentativas respondi que se confundiu
Elogio (s)	"Quando não entendia, ela pedia desculpas, ela é tão educada né?!"
Crítica (s)	Não se aplica
Sugestão (ões)	Sugeriu não ter que ir até o aplicativo da assistente para usá-la, e também que o aplicativo fosse livre de propagandas, pois elas confundem
Dificuldade(s)	A participante ficou confusa ao entrar no aplicativo da assistente por conta das inúmeras propagandas

Fonte: Elaborado pelo autor (2021)

Quadro 25 – Quadro Descritivo do Teste - Participante 5

Participante 5	
Observação (ões) do aplicativo do assistente	O assistente não respondeu aos comandos do usuário fora do aplicativo do assistente
Observação (ões) da participante	Não se aplica
Elogio (s)	Não se aplica
Crítica (s)	"Pra eu poder dar o comando, tenho que apertar no microfone!" "Ele não reconheceu, tenho que ir no aplicativo dele!"
Sugestão (ões)	Sugeriu não ter que ir ao aplicativo do assistente para usá-la
Dificuldade(s)	A participante encontrou dificuldade ao dar comandos a assistente com a tela do celular desligada

Fonte: Elaborado pelo autor (2021)

## 6. Relato dos Resultados

Por fim, com a avaliação remota aqui detalhada e apresentada foi possível obter relatos como "Ela não compreendeu o que eu falei", "Ela não abre se eu não for no aplicativo", e com as análises realizadas foi possível entender que estão relacionados à problemas que ocorrem durante o uso, nos comandos por voz, na compreensão do idioma que acabaram consolidando dados obtidos na avaliação textual. Os resultados aprese foram obtidos a partir, de um grupo específico de usuários.

## 6 DISCUSSÃO

Nesta seção, são apresentados alguns pontos a serem discutidos sobre as descobertas realizadas nas avaliações para a construção do trabalho. Os pontos a serem discutidos são: o sistema por voz avaliado, a aplicação da metodologia MALTU, a aplicação de Testes Remotos com usuários, validação dos problemas, novos problemas encontrados e detalhamento dos problemas.

O assistente avaliado Google Assistant foi escolhido para a avaliação com o objetivo de conseguir uma amostragem considerável para os resultados, tendo em vista que o assistente é nativo em celulares com o sistema operacional Android e, por esse motivo, é esperado que o aplicativo receba muitos comentários de avaliação de seus usuários.

Na aplicação da metodologia MALTU foram consideradas as Postagens Relacionadas ao Uso (PRUs) mais recentes, pois assim é possível detectar problemas que estão ocorrendo em uma época de destaque dos AVIs. Estas postagens estão compreendidas no período de 13 de setembro de 2018 à 18 de maio de 2020, o que é algo a favor pela possibilidade de detecção de problemas recentes como dito antes, mas que também tem algo contra, pois pode haver a possibilidade de restringir as postagens aos problemas de um período específico. Outra limitação, é que a classificação das postagens em PRUs e não-PRUs reduz a quantidade de postagens que são consideradas para os resultados, e com isso, necessita de um grande número de postagens com o objetivo de obter amostragens maiores. Mas, para o seguinte trabalho foi possível obter uma amostragem que foi satisfatória para a avaliação.

Na aplicação dos Testes Remotos foram considerados os problemas descobertos na aplicação da metodologia MALTU, que serviram como base para montar os cenários de uso mais abrangentes. Diante disso, foram encontradas algumas limitações durante o processo de planejamento pois, não foi possível criar mais cenários de uso por conta do período de tempo em que foram realizados os testes. Outra limitação foi a quantidade de usuários definidos, pois com isso não foi possível obter uma diversidade de usuários. Mas, essas limitações não foram consideradas problemas, pois os cenários foram criados com base nas funcionalidades mais criticadas entre os usuários na avaliação textual. Já em relação a quantidade de usuários, segundo Nielsen (2000), Virzi (1992) e Lewis (1994) 5 participantes rendem 80% das descobertas de usabilidade (BEVAN *et al.*, 2003), e por esse motivo, a quantidade de usuários foi satisfatória para os testes.

Com a realização das duas avaliações, houve consequentemente uma validação dos problemas encontrados pois, foram muitos relatos semelhantes, principalmente em relação as críticas e sugestões obtidas na avaliação textual e nos testes remotos.

Durante a execução dos testes remotos foi percebida a existência de novos problemas que ocorreram durante o uso, algo que não foi relatado na avaliação textual, como: a) a necessidade do usuário ir até o aplicativo do assistente para a execução dos comandos; b) confusão do usuário em relação a quantidade de propagandas que surpreendem o usuário ao

entrar no aplicativo, gerando conflito de interesses; c) o usuário solicita entrar em um aplicativo e o assistente oferece a instalação de um aplicativo que já se encontra no celular.

Por fim, com a execução dos testes remotos foi possível obter esses dados mais detalhados do momento de uso do usuário, bem como os comentários que surgiram antes, durante e depois do uso, enriquecendo os resultados do trabalho.

## 7 CONCLUSÃO

Este trabalho teve como finalidade apresentar o resultado de avaliações que pudessem expressar problemas de usabilidade e de experiência de uso do usuário no Google Assistant. Para isso, foram executados dois tipos de avaliação focadas no momento de uso do AVI, uma avaliação textual baseada em comentários de usuários em uma loja de aplicativos, a outra uma aplicação de testes remotos com cenários de uso das principais funcionalidades criticadas. Diante disso foi possível obter os seguintes resultados: a) detecção de problemas provenientes do momento de uso por meio aplicação da Metodologia MALTU baseados em comentários realizados pelos usuários do assistente, como exemplo, "O reconhecimento por voz não funciona direito"; b) confirmação dos problemas de uso antes detectados na aplicação da metodologia MALTU, testados em cenários de uso vistos como problemáticos, como exemplo, "Ele não reconheceu, tenho que ir no aplicativo dele!".

Os resultados obtidos por meio dessas duas avaliações, apontam para o fato de que problemas de uso relacionados a forma de falar (comandos de voz) - usuário: "Ligar whatsapp"; assistente: "Pra quem deseja ligar?!"; ao idioma - usuário: "ok google!, entrar whatsapp"; assistente: responde em inglês, e ao feedback - usuário: "ok google!, Ligar whatsapp"; assistente: definição da palavra ligação, podem causar stress ao usuário, ou até mesmo criar algum trauma relacionado à tecnologia como visto em alguns comentários extraídos na fase de avaliação textual, e em alguns relatos de usuários durante os testes remotos, relatando um trauma em relação ao notebook que usava, como exemplo.

Conclui-se com este trabalho que a consideração das funcionalidades que mais criticadas foi um resultado importante obtido com a avaliação textual, para a aplicação dos testes. Com isso, foi possível ter uma validação dos problemas encontrados, pois funcionalidades como exemplo o reconhecimento por voz, ativar sozinho e idioma com altos percentuais de falha durante o momento de uso, causam um certo desconforto ao usuário quando este precisa instalar o aplicativo do assistente para que o assistente seja ativado, ou entenda o que é dito, como relatado por alguns usuários no teste remoto como exemplo "Se não fosse isso, estava ótimo", se referindo a ida até o aplicativo do assistente.

### 7.1 Trabalhos Futuros

Como trabalhos futuros, pretende-se realizar testes remotos com um maior número de usuários, abrangendo diferentes grupos e até mesmo regiões, para que outros pontos como o desempenho do assistente em relação ao tempo de resposta, impacto da quantidade de informações em uma tela pode causar nos usuários também possam ser contemplados.

## REFERÊNCIAS

ALENCAR, A. J.; SCHMITZ, E. A.; CRUZ, L. T. **Assistentes Virtuais Inteligentes: Conceitos e estratégias**. [S.l.]: Brasport, 2013.

ALINE PRADO. **Viagens na Era da Assistência: os consumidores querem ajuda. Cadê sua marca nessa hora?** 2018. Disponível em: <<https://www.thinkwithgoogle.com/intl/pt-br/tendencias-de-consumo/viagens-na-era-da-assistencia-os-consumidores-querem-ajuda-cade-sua-marca-nessa-hora/>>. Acesso em: 29 mar. 2020.

AMAZON. **Alexa Skills de Bradesco, Turma da Mônica e iFood; experiências de casa inteligente com Positivo, Philips Hue, D-Link; dispositivos com Alexa integrada de Intelbras, Bose, Sony e LG; e muito mais - agora vindo para o Brasil**. 2019. Disponível em: <<https://www.amazon.com.br/b?ie=UTF8&node=19927229011>>. Acesso em: 13 abr. 2020.

APPLE. **Apple**. 2020. Disponível em: <[https://www.apple.com/br/?afid=p2387Cs6584Q6YW-dc\\_mtid\\_1870765e38482\\_pcrd\\_381950750831\\_pgrid\\_17049243298\\_&cid=aos-br-kwgo-brand--slid--product-](https://www.apple.com/br/?afid=p2387Cs6584Q6YW-dc_mtid_1870765e38482_pcrd_381950750831_pgrid_17049243298_&cid=aos-br-kwgo-brand--slid--product-)>. Acesso em: 23 abr. 2020.

ARTICULAB. **SARA: o assistente de robô socialmente consciente**. 2017. Disponível em: <<http://articulab.hcii.cs.cmu.edu/projects/sara/>>. Acesso em: 14 abr. 2020.

BALLATI, F.; CORNO, F.; RUSSIS, L. D. Assessing virtual assistant capabilities with italian dysarthric speech. In: **Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility**. [S.l.: s.n.], 2018. p. 93–101.

BARBOSA, S.; SILVA, B. **Interação humano-computador**. [S.l.]: Elsevier Brasil, 2010.

BEVAN, N.; BARNUM, C.; COCKTON, G.; NIELSEN, J.; SPOOL, J.; WIXON, D. The "magic number 5" is it enough for web testing? In: **CHI'03 extended abstracts on Human factors in computing systems**. [S.l.: s.n.], 2003. p. 698–699.

BLOGS DO WINDOWS. **Harman Kardon Invoke com Cortana: som cativante encontra assistente digital pessoal**. 2017. Disponível em: <<https://blogs.windows.com/windowsexperience/2017/05/08/harman-kardon-invoke-featuring-cortana-captivating-sound-meets-personal-digital-assistant/#X3pCdLRpIBQKugFP.97/>>. Acesso em: 14 abr. 2020.

BLOGS DO WINDOWS. **Hey Cortana, abra o Alexa: a colaboração inédita da Microsoft e da Amazon**. 2017. Disponível em: <<https://blogs.microsoft.com/blog/2017/08/30/hey-cortana-open-alex-microsoft-amazons-first-kind-collaboration/>>. Acesso em: 14 abr. 2020.

BLOGS DO WINDOWS. **O Cortana Skills Kit capacita os desenvolvedores a criar experiências inteligentes para milhões de usuários**. 2017. Disponível em: <<https://blogs.windows.com/windowsdeveloper/2017/05/10/cortana-skills-kit-empowers-developers-build-intelligent-experiences-millions-users/>>. Acesso em: 14 abr. 2020.

BUDIU, R.; LAUBHEIMER, P. Intelligent assistants have poor usability: A user study of alexa, google assistant, and siri. **Nielsen Norman Group, Fremont, CA, USA**, 2018.

FELIPE PAYÃO. **Samsung Bixby: conheça a assistente pessoal que pode revolucionar o celular**. 2017. Disponível em: <<https://www.tecmundo.com.br/samsung-galaxy-s8/115384-samsung-bixby-conheca-assistente-pessoal-busca-revolucionar-mercado.htm>>. Acesso em: 13 abr. 2020.

GOOGLE. **Google Assistant**. 2020. Disponível em: <[https://assistant.google.com/intl/pt\\_br/](https://assistant.google.com/intl/pt_br/)>. Acesso em: 28 abr. 2020.

HARTSON, H. R.; CASTILLO, J. C.; KELSO, J.; NEALE, W. C. Remote evaluation: the network as an extension of the usability laboratory. In: **Proceedings of the SIGCHI conference on human factors in computing systems**. [S.l.: s.n.], 1996. p. 228–235.

IBM. **Shoebox - IBM Archives**. 2011. Disponível em: <[https://mediacenter.ibm.com/media/0\\_4m2ynnkk?mhsr=ibmsearch\\_a&mhq=shoebox](https://mediacenter.ibm.com/media/0_4m2ynnkk?mhsr=ibmsearch_a&mhq=shoebox)>. Acesso em: 13 abr. 2020.

IFTTT. **IFTTT**. 2020. Disponível em: <<https://ifttt.com/>>. Acesso em: 20 mai. 2020.

IMPRESA MERCADO E CONSUMO. **Cerca de 70 por cento das pessoas substituirão ida às lojas por assistentes por voz**. 2019. Disponível em: <<https://www.mercadoeconsumo.com.br/2019/09/25/cerca-de-70-das-pessoas-substituirao-ida-as-lojas-por-assistentes-de-voz/>>. Acesso em: 10 abr. 2020.

LARSEN, H. H.; SCHEEL, A. N.; BOGERS, T.; LARSEN, B. Hands-free but not eyes-free: A usability evaluation of siri while driving. In: **Proceedings of the 2020 Conference on Human Information Interaction and Retrieval**. [S.l.: s.n.], 2020. p. 63–72.

LEWIS, J. R. Sample sizes for usability studies: Additional considerations. **Human factors**, SAGE Publications Sage CA: Los Angeles, CA, v. 36, n. 2, p. 368–378, 1994.

LIMA, L.; FURTADO, V.; FURTADO, E.; ALMEIDA, V. Análise empírica do viés em assistentes pessoais baseados em voz. In: **Anais complementares da Conferência da World Wide Web 2019**. Nova York, NY, EUA: Association for Computing Machinery, 2019. (WWW '19). ISBN 9781450366755. Disponível em: <<https://doi.org/10.1145/3308560.3317597>>.

LIMA, L.; FURTADO, V.; FURTADO, E. S.; ALMEIDA, V.; SILVA, T. H. d. Discrimination analysis of intelligent voice assistants. In: **Proceedings of the 18th Brazilian Symposium on Human Factors in Computing Systems**. [S.l.: s.n.], 2019. p. 1–7.

LUIZ AGNER. **Testes Remotos de Usabilidade - Mini seminários na PUC-Rio**. 2020. Disponível em: <<https://agner.com.br/2014/03/21/mini-seminarios-testes-de-usabilidade-na-puc-rio/>>. Acesso em: 20 mai. 2020.

MAIOLINO, T. **Maximus: automatizando tarefas por voz**. Niterói, 2017.

MENDES, M. S. **Maltu—um modelo para avaliação da interação em sistemas sociais a partir da linguagem textual do usuário**. 2015.

MICROSOFT SUPPORT. **Cortana e a privacidade**. 2019. Disponível em: <<https://support.microsoft.com/pt-br/help/4468233/cortana-and-privacy-microsoft-privacy>>. Acesso em: 14 abr. 2020.

MOTTA, I.; QUARESMA, M. Barreiras no uso de assistentes de voz por usuários brasileiros de smartphone. 2019.

NIELSEN, J. **Why you only need to test with 5 users**. [S.l.]: Useit. com Alertbox, 2000.

PORCHERON, M.; FISCHER, J. E.; REEVES, S.; SHARPLES, S. Interfaces de voz na vida cotidiana. In: **Anais da Conferência CHI de 2018 sobre fatores humanos em sistemas de computação**. Nova York, NY, EUA: Association for Computing Machinery, 2018. (CHI '18). ISBN 9781450356206. Disponível em: <<https://doi.org/10.1145/3173574.3174214>>.

PRATES, R. O.; BARBOSA, S. D. J. Avaliação de interfaces de usuário—conceitos e métodos. In: **Jornada de Atualização em Informática do Congresso da Sociedade Brasileira de Computação, Capítulo**. [S.l.: s.n.], 2003. v. 6, p. 28.

ROGERS, Y.; SHARP, H.; PREECE, J. **Design de Interação**. [S.l.]: Bookman Editora, 2013.

RY CRIST. **O Bixby da Samsung identificará em breve os alimentos em sua geladeira do Family Hub**. 2019. Disponível em: <<https://www.cnet.com/news/samsungs-bixby-will-soon-id-the-foods-in-your-family-hub-fridge/>>. Acesso em: 13 abr. 2020.

SAMSUNG. **Bixby**. 2020. Disponível em: <<https://www.samsung.com/br/apps/bixby/>>. Acesso em: 13 abr. 2020.

STARTSE. **Como as empresas estão usando os assistentes por voz?** 2019. Disponível em: <<https://www.startse.com/noticia/nova-economia/empresas-usam-tecnologia-voz>>. Acesso em: 21 set. 2020.

TASKER. **Tasker**. 2020. Disponível em: <<http://www.tasker.com.br/>>. Acesso em: 20 mai. 2020.

TECHNOLOGY VISION. **As pessoas estão mudando, por que você não está?** 2020. Disponível em: <[https://www.accenture.com/\\_acnmedia/Accenture/Redesign-Assets/DotCom/Documents/Local/2/Accenture-TechVision2020-GEO-Sum.pdf](https://www.accenture.com/_acnmedia/Accenture/Redesign-Assets/DotCom/Documents/Local/2/Accenture-TechVision2020-GEO-Sum.pdf)>. Acesso em: 13 abr. 2020.

THIAGO LAVADO. **Em 10 anos no Brasil, o Android foi de 2 smartphones para o sistema operacional dominante do mercado**. 2019. Disponível em: <<https://glo.bo/2RulUxb>>. Acesso em: 27 mar. 2020.

VIRZI, R. A. Refining the test phase of usability evaluation: How many subjects is enough? **Human factors**, SAGE Publications Sage CA: Los Angeles, CA, v. 34, n. 4, p. 457–468, 1992.

WEST, M.; KRAUT, R.; CHEW, H. E. Eu coraria se pudesse: fechar o gênero divide as habilidades digitais por meio da educação. UNESCO, 2019.

## **ANEXO A – RELATÓRIO DE AVALIAÇÃO**

Esse anexo apresenta o relatório de avaliação do aplicativo Google Assistant, com dados validados pela professora orientadora deste trabalho e por um aluno de mestrado.



UNIVERSIDADE  
FEDERAL DO CEARÁ  
Campus Russas

# RELATÓRIO DE AVALIAÇÃO DO APLICATIVO GOOGLE ASSISTANT

Avaliadores: Elis Ionara V. do N. e Silva

Sabrina de O. Rebouças

Validadores: Marília Soares Mendes

Thiago H. O. Silva



Russas, 2020

## 1. Definição do contexto de avaliação

- a. **Sistema avaliado:** Google Assistant.
- b. **Descrição do sistema:** o Google Assistant é um Assistente por Voz Inteligente (AVI) que atua como um assistente pessoal, gerenciando tarefas de acordo com a necessidade de cada usuário (sendo para casa, trabalho, estudo ou lazer). Esse AVI apresenta como principais funcionalidades a realização de pesquisas, execução de outros aplicativos e controle de dispositivos compatíveis quando solicitado, além da criação de rotinas.
- c. **Usuários:** usuários diversos, sendo estes usuários do sistema Android 5.0 ou superiores.
- d. **Plataforma:** disponível na loja online Google Play, nos celulares.
- e. **Critérios de qualidade de uso:** usabilidade e eXperiência do Usuário (UUX).
- f. **Objetivos da avaliação:**
  - i. Identificar problemas na interação e na interface;
  - ii. Avaliar a satisfação dos usuários com o sistemas;
  - iii. Identificar a sugestão de novas funcionalidades.

## 2. Extração das postagens

- a. **Base de dados:** loja de aplicativos Google Play.
- b. **Descrição da base de dados:** o Google Play é uma plataforma de distribuição digital que foi criada e é gerenciada pela Google, oferecendo aplicativos, jogos, filmes, músicas, livros e programas de televisão podendo ser gratuitas ou pagas de forma remota. Essa plataforma é a loja oficial de aplicativos voltados para o sistema operacional Android.
- c. **Justificativa em utilizar a base de dados escolhida:**
  - i. O Google Play disponibiliza uma área para que os usuários coloquem sugestões e classifiquem o aplicativo que baixaram por meio da forma de classificação por estrelas (star rating).
  - ii. Por conta da variedade de aplicações disponíveis, possui uma grande variedade de publicações.

iii. As postagens realizadas são públicas e sem influência, por se sentirem mais à vontade para expressar suas experiências sobre o produto adquirido.

d. **Forma de extração:** automática.

e. **Ferramenta de extração:** UUX-Posts <sup>1</sup>.

f. **Padrão de extração:** postagens mais recentes.

g. **Data da extração:** 20 de maio de 2020.

h. **Quantidade de postagens extraídas:** 3394 postagens.

i. **Fonte:** Google Play.

### 3. Classificação das Postagens

#### a. Fases da classificação:

- i. **Sentencição:** nesse processo as 3394 postagens foram divididas em trechos de acordo com a exclusão de conjunções adversativas (Quadro 1), com o objetivo de conseguir frases claras e objetivas. Após a aplicação desse processo foram obtidos 4938 trechos de postagens. A seguir serão apresentadas postagens antes (Quadro 2) e depois (Quadro 3) do processo de sentencição.

Quadro 1 - Conjunções Adversativas

Conjunções Adversativas	
Todavia: 0 ocorrências	Entretanto: 1 ocorrência
Não obstante: 0 ocorrências	No entanto: 2 ocorrências
Apesar disso: 0 ocorrências	Ainda assim: 5 ocorrências
Senao/Senão: 0 ocorrências	Porem: 21 ocorrências
Se nao: 2 ocorrências	Porém: 67 ocorrências
Se não: 43 ocorrências	Mesmo assim: 39 ocorrências
Contudo: 1 ocorrência	Mas: 710 ocorrências

Fonte: Elaborado pelos autores (2021)

Quadro 2 - Quadro de Postagens Antes da Sentenciação

<b>Data</b>	<b>Postagem</b>
9 de maio de 2020,	"Gosto mas acho que pode ser possível melhorar o envio feedback constante";;
19 de junho de 2019,	"É um app maravilhoso, mas cadê o deezer nas opções de músicas.";;
17 de março de 2019,	"Poderia não precisar do comando de voz ok google para cada solicitação! mas eu amo a assistente!";;

Fonte: Elaborado pelos autores (2020)

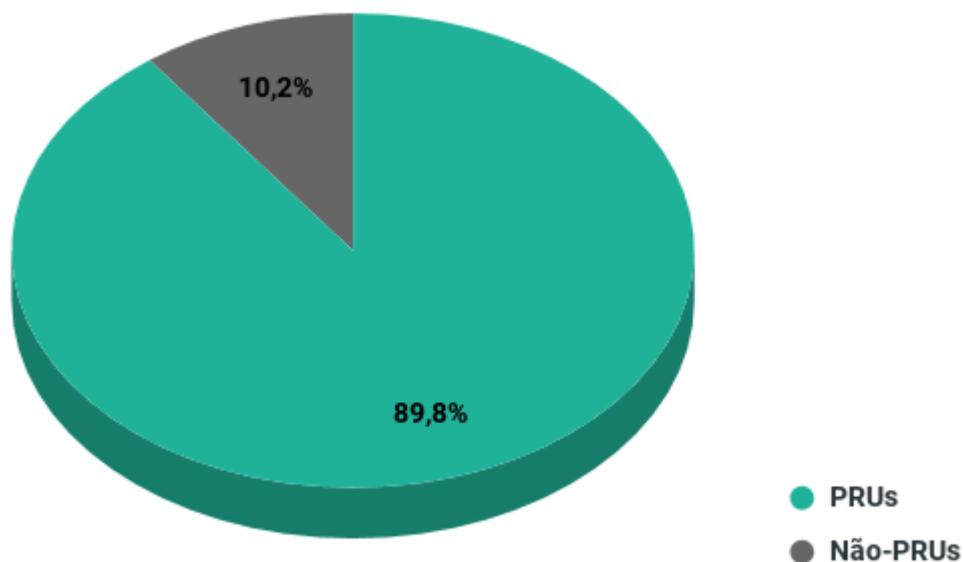
Quadro 3 - Quadro de Postagens Depois da Sentenciação

<b>Data</b>	<b>Postagem</b>	<b>Classificação</b>
9 de maio de 2020,	Gosto [.]	PRU
9 de maio de 2020,	acho que pode ser possível melhorar o envio feedback constante[.]	PRU
19 de junho de 2019,	É um app maravilhoso[.]	PRU
19 de junho de 2019,	cadê o deezer nas opções de músicas.	PRU
17 de março de 2019	Poderia não precisar do comando de voz ok google para cada solicitação![.]	Não-PRU
17 de março de 2019	Eu amo a assistente!	PRU

Fonte: Elaborado pelos autores (2020)

- ii. **Classificação em PRUs e Não-PRUs:** a classificação manual das postagens em PRUs e Não-PRUs foi iniciada no dia 17/07/2020 com término no dia 24/07/2020. Os 4938 trechos de postagens foram classificados em PRUs (4436 trechos de postagens), como exemplo, “o reconhecimento por voz não funciona direito”, e Não-PRUs (502 postagens), como exemplo, “se não fosse isso tenho cer...”, com o objetivo de eliminar os trechos de postagem não relacionados ao uso. As porcentagens estão representadas na Figura 1.

Figura 1 - Gráfico da Classificação em PRUs e Não-PRUs



Fonte: Elaborado pelos autores (2021).

**iii. Classificação das PRUs:** a classificação dos 4436 trechos de postagens realizada de forma manual, foi iniciada no dia 13/09/2020 com término no dia 22/09/2020. Essa classificação consistiu em categorizar cada PRU por: a) tipo de mensagem; b) intenção do usuário; c) polaridade do sentimento; d) intensidade do sentimento; e) critério de qualidade de uso; f) funcionalidade; e g) plataforma.

**a) Tipo de Mensagem:** tipo de classificação referente à investigação de que tipo de mensagem está sendo enviada pelo usuário sobre o sistema em uso, podendo ser dos seguintes tipos:

**i) crítica:** contém reclamação, erro, problema ou comentário negativo relacionado ao sistema, como exemplo, “o reconhecimento por voz não funciona direito”.

**ii) elogio:** contém comentários positivos relacionados ao sistema, como exemplo, “gosto do aplicativo e da funcionalidade que ele oferece”.

**iii) ajuda:** contém um passo-a-passo de alguma atividade do sistema, como exemplo, “para ativar nos celulares one vision, vc tem que desativar o modo econômica de energia,espero ter ajudado alguém”.

**iv) dúvida:** contém dúvida sobre o sistema ou suas funcionalidades, como exemplo, “tem como funcionar com a tela desligada?”.

**v) comparação:** contém uma comparação com outro sistema, como exemplo, “pisa na Siri e na Bixby”.

**vi) sugestão:** contém alguma sugestão para mudanças no sistema, como exemplo, “vocês deveria criar um nome a ela, invés de se Google assistente seriam útil”.

**b) Intenção do usuário:** tipo de classificação referente à intenção do usuário com o sistema, podendo ser dos seguintes tipos:

**i) Visceral:** não contém detalhes, nem referências a funcionalidades do sistema, com alta intensidade de sentimento, como exemplo, “muito bom a google assistente é minha amiga”.

**ii) Comportamental:** contém detalhes de uso, problemas ou funcionalidades do sistema, com pouca ou nenhuma intensidade de sentimento, como exemplo, “o assistente ativa mesmo quando não falamos ok google”.

**iii) Reflexiva:** contém referência a alguma funcionalidade do sistema, mas sem apresentar muitos detalhes de uso, como exemplo, “Poderiam colocar novamente a opção de tradução de tela, que existia há uns 3/4 anos atrás, seria ótimo”.

**c) Informação de sentimento:** tipo de classificação que analisa a intensidade do sentimento das seguintes formas:

**i) Polaridade:** contém demonstrações de sentimento positivo, como exemplo, “gosto do aplicativo e da funcionalidade que ele oferece” ; sentimento neutro, como exemplo, “Eu gostei muito

claro que anida algumas coisas mais e demais” e sentimento negativo, como exemplo, “Lumus/Nox não estão funcionando”.

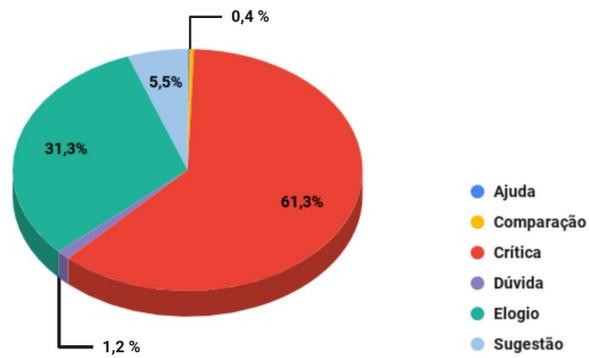
**ii) Intensidade:** contém demonstração de sentimento positivo, e de sentimento negativo.

- d) Critério de qualidade em uso:** tipo de classificação que determina o critério de qualidade em uso. A metodologia Maltu utiliza os seguintes critérios: i) usabilidade e/ou ii) UX. A Maltu utiliza as seguintes facetas para usabilidade: eficácia, eficiência, satisfação, segurança, utilidade, memorabilidade e aprendizado. Já para UX as facetas utilizadas são: satisfação, afeto, confiança, estética, frustração, motivação e suporte.
- e) Funcionalidade:** tipo de classificação que contém detalhamento sobre o uso do sistema, possibilitando dessa forma a classificação das funcionalidades pertencentes ao sistema informadas pelo usuário, ou até mesmo problemas relacionados ao sistema. Algumas dessas funcionalidades podem ser: ativação por voz do aplicativo, comandos por voz, ou até mesmo o aplicativo como um todo (usuário fala do aplicativo por inteiro). Como exemplo dessa classificação temos, “o reconhecimento por voz não funciona direito”, que é a funcionalidade de reconhecimento por voz.
- f) Plataforma:** classificação que determina o sistema operacional e dispositivo que o usuário utiliza durante o uso do sistema. No caso dessa avaliação, o sistema é Android 5.0 ou superior e os dispositivos considerados são celulares e tablets.

#### 4. Interpretação dos Resultados

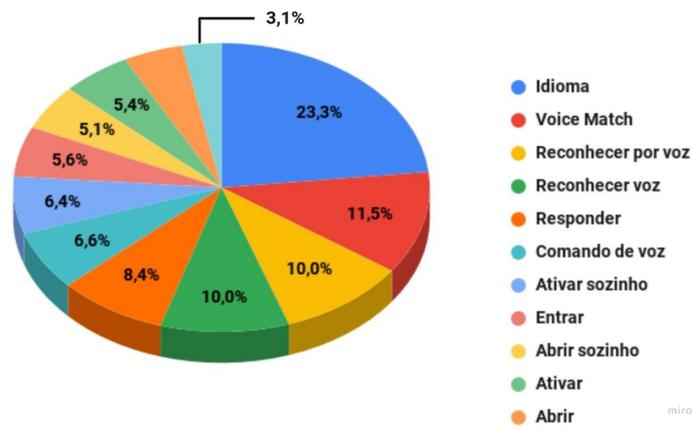
Nesta seção são apresentados gráficos e tabelas, relatando cada classificação descrita anteriormente, e a partir disso, proporcionar uma visão geral das principais características do sistema avaliado. A classificação por tipo é representada na Figura 2, onde o tipo crítica obteve o maior percentual (61,3%). Diante desse percentual de críticas, foi interessante listar as funcionalidades mais criticadas entre os usuários (Figura 3). Na Figura 4, é apresentado o Gráfico de Análise de Sentimentos onde o tipo negativa obteve o maior percentual (69,9%). Na Figura 5, é apresentado o Gráfico da Classificação em Qualidade de Uso por Usabilidade, onde a faceta eficiência obteve o maior percentual (71,8%). Por fim, na Figura 6 é apresentado o Gráfico da Classificação em Qualidade de Uso por UX, onde a faceta de frustração obteve o maior percentual (68,5%).

Figura 2 - Gráfico da Classificação por Tipo



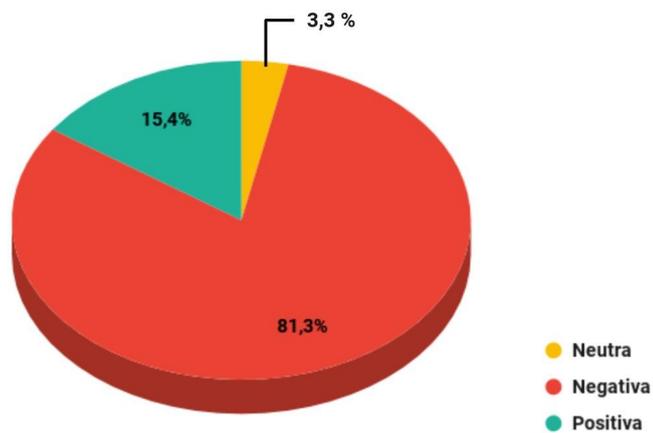
Fonte: Elaborado pelos autores (2021).

Figura 3 - Gráfico das Funcionalidades mais Criticadas



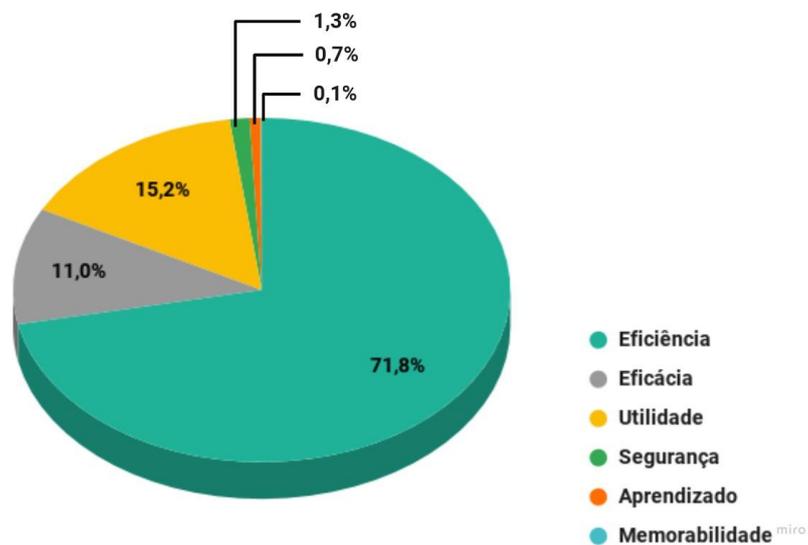
Fonte: Elaborado pelos autores (2021)

Figura 4 - Gráfico de Classificação por Análise do Sentimento



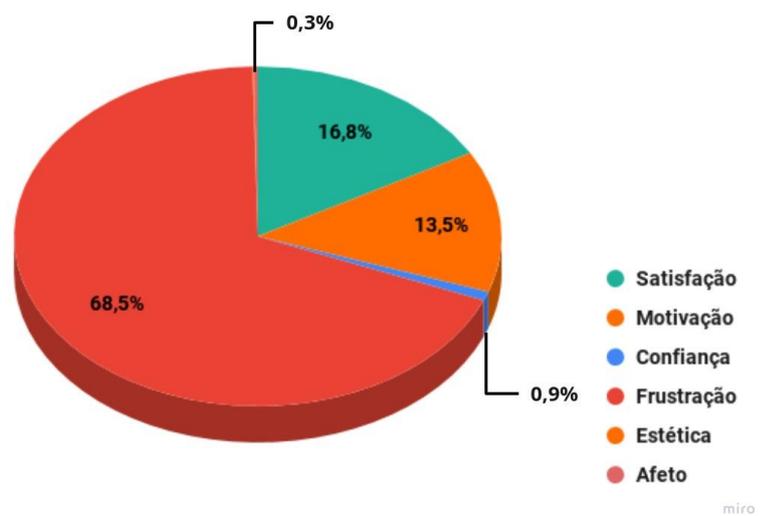
Fonte: Elaborado pelos autores (2021)

Figura 5 - Gráfico de Classificação por Qualidade em Uso (Usabilidade)



Fonte: Elaborado pelos autores (2021)

Figura 6 - Gráfico de Classificação por Qualidade em Uso (UX)



Fonte: Elaborado pelos autores (2021)

## **5. Relato dos Resultados**

Os gráficos mostrados nas imagens acima com suas respectivas descrições, possuem dados coletados de duas planilhas que foram validadas pela professora orientadora deste projeto e por um aluno de mestrado. Diante disso, foi possível chegar ao consenso das funcionalidades que mais se destacaram dentro de cada classificação.

Nas Figuras 2, 3, 4, 5 e 6 os critérios que mais se destacaram foram do tipo crítica, negativa, eficiência e frustração com as maiores porcentagens. Como o foco é detectar problemas de usabilidade, foram analisadas as funcionalidades mais criticadas pelos usuários e destas a funcionalidade idioma foi a que mais se destacou. Por fim, a conclusão é que chega é que a funcionalidade idioma foi criticada porque muitas vezes os usuários dão um comando e o AVI não entende por conta de algumas palavras do idioma do usuário.

## **ANEXO B – TERMO DE CONSENTIMENTO**

Esse anexo apresenta o termo de consentimento enviado aos usuários antes dos testes, para esclarecimento e autorização do uso dos dados.

# Termo de Consentimento e Esclarecimentos

VIA DO PARTICIPANTE

Sou \_\_\_\_\_ estudante da Universidade Federal do Ceará - UFC do Campus de Russas. Estou desenvolvendo um trabalho cujo tema é “**Avaliação da Usabilidade e da Experiência do Usuário no Google Assistant**”, sob a orientação da Prof<sup>a</sup>. Dr<sup>a</sup>. Marília Soares Mendes. Este estudo tem como objetivo identificar problemas na usabilidade e na experiência do usuário durante o uso do assistente pessoal cuja interação é realizada por comandos de voz. Por esse motivo, precisa ser submetido a um processo de teste. Logo, venho por meio deste convidá-lo(a) a participar deste estudo.

Assim, o senhor(a) receberá:

1. Instruções para a realização das tarefas e ao final,
2. Uma entrevista aberta para o relato de possíveis sugestões que o usuário queira dar após a execução do teste.

Você está sendo convidado(a) a participar, como voluntário(a). Será preservado o anonimato dos participantes. Sua participação é livre e exigirá disponibilidade de tempo para avaliação da aplicação. Durante a aplicação do teste será gravada a reunião por questões da validação da coleta de dados.

Asseguro que as informações obtidas serão utilizadas apenas para o estudo e, a qualquer momento, poderá receber informações sobre os procedimentos realizados. Sua participação também não implica em ônus ou gratificações financeiras. Você tem a liberdade de retirar seu consentimento sem qualquer prejuízo ou penalidade.

Sinta-se livre para fazer qualquer pergunta durante a leitura do termo de consentimento ou a qualquer momento da pesquisa pelo telefone (\_\_\_\_) \_\_\_\_\_. Informo que este termo será feito em duas vias, uma para o pesquisador e outra para o participante.

Eu, \_\_\_\_\_, de RG \_\_\_\_\_ declaro que depois de ser esclarecido pelo pesquisador e ter entendido o que me foi explicado, concordo em participar da pesquisa.

Russas, \_\_ de \_\_\_\_\_ de 2021

---

Assinatura do Pesquisador

---

Assinatura do Pesquisador Responsável

## **ANEXO C – ROTEIRO DE TESTE**

Esse anexo apresenta o roteiro de teste, para a execução dos testes com os usuários, que serviu como guia para as avaliadoras do teste.

# Roteiro de Teste

## 1. Detalhes do Usuário

Nome do Entrevistado:	
Idade:	
Ocupação:	
Data e Hora:	

## 2. Pré- Teste

### a. Reunião no Zoom:

- i. Envio do link da reunião;
- ii. Vai aparecer uma janela na parte superior da tela, onde você clica no botão de cancelar (azul);
- iii. Vai aparecer um botão de **iniciar a reunião** (azul);
- iv. Vai aparecer novamente uma janela na parte superior da tela, onde você irá clicar novamente em cancelar (azul);
- v. Vai aparecer um link escrito **entrar pelo navegador** (azul);
- vi. Após o clique ira aparecer uma janela e dois campos pedindo o nome do participante;
- vii. O botão de entrar ficará azul e vc clica nele;
- viii. Após estar dentro da reunião, é necessário **incluir o áudio**.

### b. Explicação e esclarecimento sobre pontos do teste;

- i. O navegador da Web vai estar aberto na sala de reunião;
- ii. Oi, *nome\_do\_usuario* meu nome é *Elis Ionara* e vou acompanhá-lo nesta sessão de teste;
- iii. Informe e relembre o porque do usuário estar ali;
- iv. Pedimos que o usuário tente usar esse AVI no qual estou realizando um estudo para que eu possa ver se ele funciona como o esperado;
- v. A sessão deve levar cerca de 20 a 30 minutos;
- vi. Quero deixar claro que estou testando o aplicativo, não você;
- vii. Você não precisa se preocupar em cometer erros, pois preciso de suas reações sinceras;
- viii. Se você tiver alguma dúvida, basta perguntar;

- ix. Se você precisar fazer uma pausa a qualquer momento é só me informar;
  - x. Com sua **permissão já consentida no termo de consentimento** vamos registrar o que acontece na tela do aplicativo do seu celular, bem como nossa conversa;
  - xi. A gravação será usada apenas para ajudar nas pesquisas de um estudo que vai contribuir para uma melhora no AVI, por meio da descoberta de incoerências que ele possa vir a ter;
  - xii. Pergunte novamente ao usuário, se ele tem alguma dúvida antes do início do teste;
- c. Quebrando o gelo...
- i. Conversa informal com o objetivo de fazer com que o usuário fique à vontade e relaxado para iniciar o teste (preferencialmente algo que sirva de gancho para o início do teste).

### 3. Teste

- a. Executando o teste
  - i. Com o entendimento da tarefa o usuário vai tentar iniciar e terminar as tarefas passadas, diante dos dois cenários.
    - 1. Cenário 1 - Desligar wi-fi;
    - 2. Cenário 2 - Abrir WhatsApp.
- b. Relato dos acontecimentos durante o teste.

### 4. Pós - Teste

- a. Pergunte ao usuário se ele tem alguma dúvida após o término do teste;
- b. Entrevista aberta: captar sugestões do usuário para o aplicativo, caso ele tenha alguma.
- c. Para a reunião;
- d. Agradecimento!