

UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO

**Reconhecimento de Voz Online e Offline:
Uma Análise de Viabilidade e Segurança em
Assistentes Virtuais**

Jean Felipe Henriques Coutinho

JUIZ DE FORA
AGOSTO, 2025

Reconhecimento de Voz Online e Offline: Uma Análise de Viabilidade e Segurança em Assistentes Virtuais

JEAN FELIPE HENRIQUES COUTINHO

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharelado em Ciência da Computação

Orientador: Eduardo Pagani Julio

JUIZ DE FORA
AGOSTO, 2025

RECONHECIMENTO DE VOZ ONLINE E OFFLINE: UMA ANÁLISE DE VIABILIDADE E SEGURANÇA EM ASSISTENTES VIRTUAIS

Jean Felipe Henriques Coutinho

MONOGRAFIA SUBMETIDA AO CORPO DOCENTE DO INSTITUTO DE CIÊNCIAS
EXATAS DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTE-
GRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE
BACHAREL EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Eduardo Pagani Julio
Doutor em Computação

Luciano Jerez Chaves
Doutor em Ciência da Computação

Eduardo Barrére
Doutor em Engenharia de Sistemas e Computação

JUIZ DE FORA
13 DE AGOSTO, 2025

Resumo

Este Trabalho de Conclusão de Curso investiga o uso de assistentes virtuais com reconhecimento de voz *offline*, destacando seus benefícios em termos de privacidade, segurança e autonomia operacional. Diferentemente das soluções baseadas na nuvem, que dependem de conexão constante com servidores externos, abordagens *offline* processam comandos localmente, reduzindo riscos de vazamento de dados e melhorando a responsividade do sistema. Ademais, a execução local fornece maior controle sobre personalização e integração com dispositivos, tornando a tecnologia mais acessível e eficiente para diferentes cenários de uso. Essa pesquisa examina esses elementos, contribuindo para o conhecimento das vantagens e desafios do reconhecimento de voz *offline* em assistentes virtuais.

Palavras-chave: Assistente, Virtual, Reconhecimento de Voz, Internet, *Offline*, Privacidade, Segurança.

Abstract

This Final Paper discusses the use of offline speech recognition-based virtual assistants with their benefits to privacy, security, and system independence in operations. Unlike cloud-based approaches that rely on ongoing server connections, offline techniques operate locally, ensuring less data leakage risk and more responsive systems. Moreover, local operation provides better control over customizations and integration with devices, making the technology more accessible and effective for various applications. This research examines those factors, which will assist in furthering knowledge about the advantages and limitations of offline speech recognition in virtual assistants.

Keywords: Virtual, Assistant, Offline, Internet, Privacy, Security, Voice Recognition.

Agradecimentos

Ao meu orientador, Prof. Eduardo Pagani Julio, pelo acompanhamento dedicado, pelas discussões enriquecedoras e pela orientação precisa que guiaram cada etapa deste trabalho.

Agradeço em especial a minha noiva que fez toda a diferença, pelo incentivo nos dias de desafio, pela paciência nos imprevistos e pelo carinho que tornou possível a realização deste trabalho.

Agradeço a todos os meus amigos que, provavelmente sem perceber, contribuíram de forma significativa para minha disposição e motivação para trabalhar neste projeto.

“A inteligência artificial e os assistentes virtuais estão transformando não apenas a maneira como interagimos com os computadores, mas também como vivemos nossas vidas”.

Sundar Pichai

Conteúdo

Lista de Figuras	7
Lista de Tabelas	8
Lista de Abreviações	9
1 Introdução	10
1.1 Apresentação do Tema	10
1.2 Contextualização	11
1.3 Descrição do Problema	11
1.4 Justificativa	12
1.5 Objetivo Geral	12
1.6 Objetivo Especifico	13
1.7 Material e Métodos	13
1.8 Organização do Trabalho	14
2 Fundamentação Teórica	16
2.1 Introdução	16
2.2 Inteligência Artificial	16
2.2.1 Aprendizado de Máquina	17
2.2.2 Processamento de Linguagem Natural	17
2.3 Assistentes Virtuais	18
2.4 Lei Geral de Proteção de Dados (LGPD)	20
2.4.1 Principais Contribuições	20
3 Revisão da Literatura	23
3.1 Um Assistente de Voz Embarcado	24
3.2 Assistente Virtual para Rede Doméstica Controlado por Voz	26
3.3 Assistente Jaco	27
3.4 Assistente Virtual em Python	29
3.5 Privacidade em Assistentes Digitais	30
3.6 Snips	31
3.7 Considerações Finais	32
4 Assistente Virtual <i>Offline</i> Desenvolvido	34
4.1 Visão Geral	34
4.2 Definições e Etapas de Desenvolvimento	35
4.3 Referencias Técnicas	37
4.4 Natureza dos Testes Realizados	39
4.5 Resultados	39
5 Considerações Finais	44
5.1 Conquistas e Limitações	44
5.2 Cumprimento dos Objetivos	45
5.3 Trabalhos Futuros	45

Lista de Figuras

4.1	Ilustração de fluxo metodológico representando as tomadas de decisão em cada etapa do desenvolvimento dos métodos do trabalho.	37
4.2	Ilustração do Assistente Virtual Desenvolvido.	38
4.3	Tráfego TLS entre o Copilot e os servidores da Microsoft (Wireshark). . . .	40
4.4	TLS ClientHello da Siri para <code>p159-contacts.icloud.com</code> contendo extensão SNI (Wireshark).	41
4.5	Ditado de Palavras Sem Contexto	42
4.6	Resultado do Ditado de Palavras Sem Contexto	42
4.7	Ditado de Palavras em Formato de Texto Corrido	43
4.8	Resultado Ditado de Palavras Texto Corrido	43

Lista de Tabelas

2.1	Visão Geral das Vulnerabilidades Críticas de Segurança em Assistentes. . .	19
3.1	Tabela Comparativa Entre os Trabalhos Relacionados	33

Lista de Abreviações

ASR	<i>Automatic Speech Recognition</i> (Reconhecimento de Fala Automática)
DCC	Departamento de Ciência da Computação
DNN	<i>Deep Neural Network</i> (Redes Neurais Profundas)
GDPR	<i>General Data Protection Regulation</i> (Regulamento Geral de Proteção de Dados)
GUI	<i>Graphical User Interface</i> (Interface Gráfica do Usuário)
HMM	<i>Hidden Markov Model</i> (Modelos Ocultos de Markov)
HTTPS	<i>Hypertext Transfer Protocol Secure</i> (Protocolo de Transferência de Hipertexto Seguro)
IA	Inteligência Artificial
IoT	<i>Internet of Things</i> (Internet das Coisas)
LGPD	Lei Geral de Proteção de Dados
MQTT	<i>Message Queuing Telemetry Transport</i> (Protocolo de Enfileiramento de Mensagens)
NLU	<i>Natural Language Understanding</i> (Entendimento de Linguagem Natural)
PLN	Processamento de Linguagem Natural
SLU	<i>Spoken Language Understanding</i> (Entendimento de Linguagem Falada)
SSL	<i>Secure Sockets Layer</i> (Camada de Sockets Segura)
STT	<i>Speech-To-Text</i> (Conversão de Fala em Texto)
TTS	<i>Text-To-Speech</i> (Conversão de Texto em Fala)
UFJF	Universidade Federal de Juiz de Fora
VCDA	<i>Voice-Controlled Digital Assistant</i> (Assistentes Digitais Controlados por Voz)
VM	<i>Virtual Machine</i> (Máquina Virtual)
WER	<i>Word Error Rate</i> (Taxa de Erro de Palavras)

1 Introdução

1.1 Apresentação do Tema

O avanço da Inteligência Artificial (IA) tem impulsionado significativamente o desenvolvimento de tecnologias baseadas em voz, tornando os assistentes virtuais ferramentas essenciais na interação entre humanos e máquinas (GUZMAN, 2019). Nesse contexto, os assistentes virtuais se destacam como exemplos proeminentes dessa tecnologia.

Esses sistemas utilizam o Processamento de Linguagem Natural (PLN), uma subárea da IA voltada à compreensão e interpretação da linguagem humana por computadores, o que permite que os assistentes reconheçam comandos e forneçam respostas com eficiência (CAMBRIA; WHITE, 2014).

O PLN é crucial para permitir que esses agentes virtuais compreendam, interpretem e respondam de maneira significativa às solicitações dos usuários, facilitando uma comunicação mais fluida e natural. Nomes como Alexa, Siri e Cortana, entre muitos outros, já se tornaram familiares para a maioria de nós, à medida que os assistentes virtuais se integram cada vez mais ao nosso cotidiano (HOY, 2018).

No entanto, esse aumento no uso desses dispositivos levanta uma questão crucial: a segurança. À medida que esses assistentes estão constantemente conectados à Internet e processam grandes volumes de dados pessoais, surgem preocupações sobre vulnerabilidades de segurança e privacidade (ALEPIS; PATSAKIS, 2017). Esses desafios se tornam ainda mais críticos à medida que a tecnologia se torna parte integrante de nossas vidas pessoais e profissionais.

Diante dessa realidade, a necessidade de uma análise crítica sobre a segurança e privacidade de assistentes virtuais torna-se evidente. A evolução contínua da tecnologia de voz e a crescente integração desses dispositivos no cotidiano exigem uma abordagem cuidadosa para equilibrar inovação e proteção dos dados dos usuários.

1.2 Contextualização

A popularização dos assistentes virtuais reflete uma mudança significativa na forma como interagimos com a tecnologia. Sua capacidade de compreender e processar comandos por meio da voz tem tornado a comunicação homem-máquina mais intuitiva, contribuindo para a automação de diversas atividades do dia a dia. No entanto, a crescente dependência de servidores na Internet para processar informações tem levantado debates sobre o equilíbrio entre conveniência e proteção dos dados pessoais.

O reconhecimento de voz *offline* surge como uma solução promissora para reforçar a privacidade e a segurança desses sistemas. Ao realizar o processamento diretamente no dispositivo do usuário, elimina-se a necessidade de transmissão constante de informações para servidores externos, reduzindo riscos de vazamento de dados e exposição a ataques cibernéticos. Além disso, essa abordagem pode aumentar a eficiência e a acessibilidade dos assistentes, tornando-os utilizáveis mesmo em ambientes com conectividade limitada ou inexistente.

Diante desse cenário, este trabalho se propõe a investigar os impactos do reconhecimento de voz *offline* em assistentes virtuais, considerando seus benefícios para a segurança digital e os desafios associados à sua implementação.

1.3 Descrição do Problema

O problema da segurança dos assistentes virtuais reside principalmente na sua constante conexão à Internet e na maneira como interagem com os dados dos usuários. Esses dispositivos coletam e processam uma quantidade significativa de informações pessoais e sensíveis, como comandos de voz, dados de localização e até mesmo detalhes de contas bancárias.

A preocupação central é que, devido a essa interconexão e à natureza dos dados manipulados, os assistentes virtuais estão suscetíveis a uma variedade de ameaças cibernéticas, como ataques de hackers, interceptação de dados e até mesmo uso indevido de informações por parte das próprias empresas desenvolvedoras dos assistentes.

Além disso, a segurança dos assistentes virtuais também pode ser comprometida

por meio de vulnerabilidades no software, falhas de projeto ou até mesmo por ações maliciosas de usuários autorizados.

Essa questão ganha ainda mais relevância com o aumento da dependência desses dispositivos para tarefas cotidianas, desde controlar dispositivos domésticos até realizar transações financeiras, tornando essencial abordar e resolver os desafios de segurança associados aos assistentes virtuais.

1.4 Justificativa

Com a crescente adoção de assistentes virtuais, torna-se fundamental assegurar a proteção das informações manipuladas por esses sistemas. A dependência de processamento em servidores externos gera preocupações quanto à privacidade, pois dados sensíveis são constantemente transmitidos e armazenados em infraestrutura remota, aumentando a exposição a vulnerabilidades.

Nesse contexto, o reconhecimento de voz *offline* surge como uma alternativa que pode reduzir esses riscos, ao permitir que as interações sejam processadas diretamente no dispositivo do usuário, sem a necessidade de envio de informações para a nuvem. Considerando esses aspectos, esta pesquisa propõe o desenvolvimento de um assistente virtual baseado nessa tecnologia, analisando seus benefícios para a segurança e sua viabilidade como solução para uma experiência mais privada e confiável.

1.5 Objetivo Geral

O objetivo central deste projeto é desenvolver um assistente virtual multiplataforma, capaz de operar nos três principais sistemas operacionais do mercado Windows, Linux e macOS, que permita o reconhecimento de voz em tempo real de maneira *offline*, e que possa interagir com o sistema operacional assim como a Siri da Apple ou Copilot da Microsoft, para realizar tarefas simples de maneira automatizada.

1.6 Objetivo Especifico

Estes são os objetivos específicos definidos para este trabalho:

- Realizar uma revisão bibliográfica sobre métodos de PLN e reconhecimento de voz *offline*.
- Desenvolver um sistema de assistente virtual que opere sem necessidade de conexão à Internet.
- Implementar funcionalidades básicas no assistente virtual, como resposta a consultas e execução de tarefas específicas.
- Conduzir testes e análises para avaliar a segurança do usuário ao interagir com o assistente virtual em um ambiente controlado.

1.7 Material e Métodos

O desenvolvimento do assistente virtual seguiu uma abordagem metodológica dividida em quatro etapas principais.

1. **Definição e Integração de Ferramentas:** Para a escolha das ferramentas utilizadas no assistente virtual, realizou-se um estudo comparativo entre alternativas disponíveis. Esse estudo avaliou parâmetros como a compatibilidade com operação totalmente *offline*, a facilidade de uso das bibliotecas e a qualidade da documentação técnica Ferraro et al. (2023) ressaltam que “é essencial avaliar o desempenho e a qualidade” das ferramentas de conversão da fala em texto (STT, *Speech-To-Text*) para selecionar a mais adequada a cada aplicação. Com base nessa análise, foram selecionadas as bibliotecas que melhor equilibram desempenho em modo *offline*, cobertura funcional e facilidade de integração ao projeto;
2. **Desenvolvimento da Interface Gráfica:** Foi implementada uma interface gráfica (GUI, *Graphical User Interface*) simples para capturar comandos do usuário, por meio de botões disponibilizados na interface ou por meio de voz, permitindo interação e controle dos processos do sistema operacional;

3. **Reconhecimento de Fala:** Por meio do microfone, captura-se o áudio do usuário, com a ferramenta de STT converte-se esse áudio em texto. O áudio é processado em tempo real, utilizando técnicas de tratamento de ruído (DUA; AKANKSHA; DUA, 2023) e ajuste de sensibilidade para melhorar a precisão da detecção;
4. **Execução de Comandos:** Após o reconhecimento do comando, a interface interpreta o texto e executa ações correspondentes, como abrir programas com *feedback* fornecido ao usuário através da síntese de voz;

1.8 Organização do Trabalho

No Capítulo 2 é apresentado o arcabouço teórico que fundamenta o estudo, são definidos os principais conceitos de Inteligência Artificial, Aprendizado de Máquina e Processamento de Linguagem Natural, descrevendo como essas áreas suportam o desenvolvimento de assistentes virtuais. Em seguida, discute-se a arquitetura básica de um assistente, suas vulnerabilidades de segurança e as exigências da Lei Geral de Proteção de Dados para sistemas de reconhecimento de voz.

No Capítulo 3, são analisados diversos trabalhos acadêmicos e soluções comerciais que abordam assistentes de voz, tanto *online* quanto *offline*. Cada trabalho é apresentado em detalhes, indicando seus objetivos, métodos e resultados, seguido de uma avaliação dos pontos fortes e fracos. Ao final, uma tabela comparativa resume as características de cada proposta, permitindo compreender melhor os desafios e as oportunidades no uso de reconhecimento de voz *offline*.

O Capítulo 4, descreve o projeto prático realizado neste trabalho. Apresenta as escolhas tecnológicas, o capítulo detalha também o fluxo metodológico de desenvolvimento e os testes conduzidos nos ambientes Windows e macOS. Detalha também o protocolo de monitoramento de tráfego (com Wireshark) para verificar a independência de comunicação com servidores remotos e garante que todo o processamento ocorra localmente.

Por fim, o Capítulo 5 reúne os principais achados deste trabalho, avaliando se o reconhecimento de voz *offline* atende aos requisitos de segurança e privacidade. São discutidas as contribuições do trabalho, suas limitações e sugestões para trabalhos futuros,

apontando caminhos para aprimorar a experiência e a confiabilidade de assistentes virtuais que operam sem conexão com a nuvem.

2 Fundamentação Teórica

2.1 Introdução

Neste capítulo são apresentados os conceitos que apoiam o entendimento deste trabalho, incluindo Inteligência Artificial, Processamento de Linguagem Natural, Assistentes Virtuais e a Lei Geral de Proteção de Dados. Esses tópicos fornecem a base teórica necessária para compreender o funcionamento dos assistentes virtuais, seus desafios em termos de segurança e privacidade, bem como o impacto da legislação na proteção dos dados dos usuários.

2.2 Inteligência Artificial

A IA é definida como um campo de pesquisa que tem sido aplicado em diversos setores, tornando-se um campo de estudo no século XXI passando a ser essencial em áreas como a engenharia, finanças e direito.

Estudos como os de Halal (2003), Masnikosa (1998) e Metaxiotis et al. (2003) destacam a vasta aplicação da IA, abrangendo desde a automação industrial até o desenvolvimento de sistemas autônomos e inteligentes para a resolução de problemas complexos. Além disso, a IA tem se fragmentado em diversos campos especializados, como algoritmos genéticos, sistemas especialistas, e aprendizado de máquina, cada um com objetivos e métodos específicos (OKE, 2004).

Um dos maiores desafios da IA, segundo Oke (2004), é a rápida expansão do campo, o que torna difícil acompanhar os estudos e os avanços tecnológicos. Além disso, a complexidade dos problemas que a IA tenta resolver, como o PLN e a representação do conhecimento, exige o desenvolvimento de novas técnicas e abordagens que permitam uma compreensão mais profunda e aplicável do campo.

Portanto, não podemos apenas abranger as aplicações práticas da IA, mas temos que incluir também os desafios técnicos e éticos que emergem à medida que novos avanços

são feitos, como a necessidade de uma maior compreensão e controle sobre os sistemas inteligentes que estão sendo desenvolvidos.

2.2.1 Aprendizado de Máquina

O aprendizado de máquina (*machine learning*) é um subcampo da inteligência artificial que se dedica ao desenvolvimento de algoritmos capazes de aprender a partir de dados, sem programação explícita. Em termos gerais, trata-se de treinar modelos estatísticos ou algorítmicos com exemplos (dados de treinamento) para que reconheçam padrões e façam previsões sobre novos dados. As principais técnicas incluem métodos supervisionados (como redes neurais e árvores de decisão), não supervisionados (como agrupamento, *clustering*) e híbridos conforme explica Brunialti et al. (2015).

2.2.2 Processamento de Linguagem Natural

O PLN é uma das subáreas da IA que se concentra em técnicas computacionais para análise e representação automática da linguagem humana. Segundo Cambria e White (2014), o PLN evoluiu significativamente, passando de um processamento lento baseado em cartões perfurados para a capacidade atual de processar milhões de páginas web em segundos. Essa evolução possibilitou uma maior interação entre humanos e máquinas, aproximando-se cada vez mais da compreensão de linguagem natural.

Historicamente, o PLN começou com foco na sintaxe, devido à sua aplicabilidade prática e ao suporte das técnicas de aprendizado de máquina. No entanto, ao longo do tempo, a comunidade científica reconheceu a necessidade de um processamento baseado na semântica para lidar com a interpretação de textos, extração de informações significativas e tarefas complexas, como a desambiguação de sentidos das palavras (CAMBRIA; WHITE, 2014). Esse avanço é evidenciado pelo surgimento de sistemas que tentam imitar a forma como o cérebro humano processa a linguagem.

2.3 Assistentes Virtuais

De acordo com Alencar, Schmitz e Cruz (2013), assistentes virtuais são programas de computador desenvolvidos para interagir com os usuários utilizando a linguagem natural. Esses sistemas são projetados para compreender e processar comandos e consultas em linguagem humana, permitindo uma comunicação intuitiva e eficiente. Eles podem realizar uma ampla gama de tarefas, desde responder a perguntas simples e fornecer informações até realizar ações mais complexas, como agendar compromissos ou controlar dispositivos domésticos inteligentes. Os assistentes virtuais empregam técnicas avançadas de PLN e aprendizado de máquina para melhorar continuamente suas capacidades de compreensão e resposta.

Vulnerabilidades

Os assistentes virtuais (como Alexa, Google Assistant, Siri etc.) apresentam vulnerabilidades que podem ser exploradas por atacantes. Em geral, muitos desses pontos fracos decorrem de comunicações inseguras e configurações de segurança inadequadas. Estudos mostram que vulnerabilidades críticas comuns incluem o uso de *Non-SSL URLs* (URLs sem criptografia HTTPS, versão segura do protocolo HTTP) e a ausência de validação de certificados SSL, permitindo ataques nos assistentes mencionados.

Essas falhas aparecem em múltiplas plataformas (por exemplo, Alexa, Cortana, Google Assistant) e representam riscos importantes de interceptação de dados. Outros vetores recorrentes são permissões excessivas e configurações inseguras de intenção (*Implicit Intent Usage*) e verificação de *hostname* (*Permissive Hostname Verifier*), que podem permitir escalonamento de privilégios ou execução de código malicioso no dispositivo do assistente.

A Tabela 2.1 resume essas falhas críticas de segurança, indicando quais assistentes são afetados em cada caso. Em resumo, Kalhor e Das (2023) aponta que a maior parte das vulnerabilidades em assistentes de voz está relacionada a lacunas de segurança na comunicação e criptografia e a práticas inseguras de desenvolvimento (por ex., permissões não restritas, execução insegura de comandos *SQL*). Essas vulnerabilidades recorrentes aumentam a superfície de ataque dos sistemas de voz, reforçando a necessidade de projeto de software seguro e controle rigoroso das permissões em aplicações de assistentes virtuais.

Tabela 2.1: Visão Geral das Vulnerabilidades Críticas de Segurança em Assistentes.

Vulnerabilidades Críticas	Assistentes Virtuais
Non-SSL URLs	Amazon, Cortana, Extreme, Hound, Kalliope, Mycroft
Implicit Intent Usage	Alexa, Cortana, Extreme, Hound, Mycroft
Missing Stack Canary in Shared Objects	Alexa, Cortana, Hound, Mycroft
Permissive HOSTNAME VERIFIER	Cortana, Extreme, Kalliope
SSL Certificate Validation Bypass	Cortana, Extreme, Kalliope
Raw SQL Queries in SQLite	Cortana, Kalliope
Runtime Command Execution	Cortana, Extreme
WebView Javascript Interface Vulnerability (CVE-2013-4710)	Cortana, Extreme
ContentProvider Exported	Alexa
Base64 Encoding Used as Encryption	Alexa
Misconfigured "intent-filter"	Cortana
ECB Mode AES Usage	Cortana
Clear Text Traffic Enabled	Cortana
Non-Standard Activity Launch Mode	Cortana
CBC Mode AES with Padding Oracle	Cortana
Standhogg 2.0 Vulnerability	Extreme
System-Level Permissions in Manifest	Extreme
Insecure SSL Pinning with Byte Array/Hard-Coded Cert Info	Kalliope
Fragment Vulnerability (CVE-2013-6271)	Mycroft
Debug Mode Enabled	Mycroft
Dangerous Sandbox Permissions	Mycroft

¹ Tabela adaptada de Kalhor e Das (2023)

Cabe ressaltar que a Siri, assistente virtual da Apple, não foi contemplada no estudo de Kalhor e Das (2023). A ausência dessa análise pode estar relacionada a restrições de acesso às informações técnicas da plataforma ou ao escopo metodológico adotado pelos autores. Dessa forma, não é possível afirmar, a partir desse levantamento específico, se a Siri compartilha das mesmas vulnerabilidades identificadas em outros assistentes, o que abre espaço para investigações futuras.

2.4 Lei Geral de Proteção de Dados (LGPD)

A Lei Geral de Proteção de Dados (LGPD)¹ foi sancionada em 14 de agosto de 2018, sob o número 13.709/2018. Seu objetivo principal é garantir a privacidade e a proteção dos dados pessoais, além de regular o tratamento de informações pessoais tanto por parte de empresas quanto por órgãos públicos.

2.4.1 Principais Contribuições

A LGPD estabeleceu regras claras sobre o uso de dados pessoais no Brasil. Algumas das suas principais contribuições são:

- **Consentimento:** O tratamento de dados pessoais só pode ser realizado mediante o consentimento do titular, que deve ser informado de maneira clara e explícita sobre a finalidade da coleta;
- **Finalidade:** Os dados devem ser utilizados apenas para fins específicos e legítimos, e o titular dos dados deve ser informado sobre o propósito do uso;
- **Transparência:** A LGPD obriga as empresas a manterem uma comunicação clara com os titulares dos dados sobre como suas informações estão sendo utilizadas;
- **Segurança:** As organizações devem adotar medidas técnicas e administrativas para proteger os dados pessoais de acessos não autorizados, destruição acidental, perda, ou qualquer forma de tratamento inadequado;

¹https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm

- **Responsabilidade:** As empresas que processam dados pessoais são responsáveis por garantir a conformidade com a LGPD e podem ser penalizadas em caso de infrações, incluindo multas de até 2% do faturamento anual, limitadas a R\$50 milhões por infração;
- **Direitos dos titulares:** Os titulares de dados têm uma série de direitos, incluindo o direito de acesso, retificação, eliminação, portabilidade dos dados, e oposição ao tratamento de dados em determinados casos;
- **Tratamento de dados sensíveis:** A LGPD impõe restrições adicionais ao tratamento de dados sensíveis, como informações sobre origem racial, convicções religiosas, e dados biométricos;
- **Transferência internacional de dados:** O compartilhamento de dados com outros países só pode ser feito se garantido um nível adequado de proteção ou mediante garantias específicas;
- **Criação da Autoridade Nacional de Proteção de Dados (ANPD):** A LGPD também criou a ANPD, que é responsável por fiscalizar e garantir o cumprimento da legislação, além de orientar empresas e entidades públicas sobre boas práticas na proteção de dados.

Com a entrada em vigor da LGPD, as organizações precisam reavaliar como coletam, tratam e armazenam dados pessoais. É necessário que as empresas adotem práticas de conformidade, mapeiem os dados sob sua custódia, revisem suas políticas de privacidade, e implementem controles de segurança mais rígidos para garantir a proteção das informações pessoais de seus clientes, usuários e colaboradores.

Além disso, o respeito aos direitos dos titulares de dados se tornou uma prioridade, e as organizações que não se adequarem às novas exigências legais correm o risco de sofrer sanções severas, tanto financeiras quanto de reputação.

Nesse contexto, a abordagem *offline* adotada neste trabalho contribui diretamente para a conformidade com a LGPD. Ao processar os dados de voz localmente no dispositivo, evita-se a transmissão contínua de informações pessoais para servidores externos, atendendo ao princípio de minimização de dados. Essa configuração local reduz

os riscos de interceptação ou uso indevido dos dados, reforçando a privacidade desde a concepção (*privacy by design*) conforme previsto na legislação. Em outras palavras, manter o processamento de fala em modo *offline* alinha-se às diretrizes da LGPD, pois limita a coleta de dados ao mínimo necessário e mantém os dados sensíveis sob controle do usuário. Dessa forma, o assistente virtual projetado tende a apresentar menor exposição de dados pessoais, facilitando o cumprimento das obrigações legais de proteção de dados e promovendo maior confiança dos usuários.

3 Revisão da Literatura

Nesta seção são apresentados seis trabalhos relacionados na literatura. O primeiro trata de um assistente de voz embarcado de Lazzaroni, Bellotti e Berta (2024), destinado ao uso em veículos, e que, assim como este trabalho, também processa voz localmente. O assistente embarcado mostra que é possível alcançar WER competitivo sem enviar áudio a servidores externos. A diferença é que, enquanto eles visam controlar funções de bordo (navegação, mídia, clima), este trabalho foca na integração com o sistema operacional embarcado.

No trabalho de Martins (2020) , é apresentado um assistente virtual para redes domésticas, com o objetivo de automatizar tarefas e permitir controle por voz, de forma semelhante ao assistente proposto neste trabalho. A solução de Martins (2020) integra os protocolos de fila MQTT(*Message Queuing Telemetry Transport*) e o sistema operacional para dispositivos embarcados *OpenWRT* para realizar diagnósticos de rede por comandos de voz, simplificando operações sem a necessidade de interfaces gráficas complexas. Porém, a proposta de Martins (2020) depende de conexão com a Internet, enquanto o assistente aqui apresentado funciona de forma totalmente *offline*.

O Assistente Jaco, descrito por Bermuth, Poeppel e Reif (2022), é um *framework* modular que funciona de forma *offline* em dispositivos Raspberry Pi, utilizando o Rasa (plataforma de código aberto para construção de *chatbots* e assistentes virtuais) e o PicTTS (sintetizador de texto para fala). Assim como o assistente desenvolvido neste trabalho, o Jaco prioriza a privacidade do usuário. A principal diferença está na complexidade: o Jaco requer a configuração de múltiplos módulos e do Rasa, além de ser projetado especificamente para o Raspberry Pi, enquanto o assistente apresentado neste trabalho possui uma estrutura mais simples e de fácil implantação.

O Assistente virtual em *Python* de Appalaraju et al. (2021) usa *SpeechRecognition* (biblioteca para realização de reconhecimento de fala) e Google API para um assistente básico que abre aplicações e busca informações *online*. O trabalho aqui apresentado é inspirado nessa simplicidade de código em *Python*, mas inova ao substituir chamadas a

APIs remotas por modelos *offline*, garantindo operação contínua mesmo sem rede, ao custo de não poder garantir a mesma precisão de serviços *online*.

O trabalho Privacidade em Assistentes Digitais de Acosta e Reinhardt (2022) analisam ameaças de privacidade e propõem mascaramento de voz e autenticação local, pautados pelo Regulamento Geral de Proteção de Dados (GDPR, *General Data Protection Regulation*) da União Europeia. Esses conceitos reforçam a abordagem do assistente apresentado neste trabalho: Ao manter todo o processamento de voz local, você adota o conceito “*privacy by design*” sugerido segundo (ACOSTA; REINHARDT, 2022), evitando tanto a coleta desnecessária de dados quanto a exposição a múltiplos intervenientes externos.

Por fim, o Snips, proposto por Coucke et al. (2018), é um assistente embarcado desenvolvido na linguagem de programação *Rust*, voltado para aplicações em Internet das Coisas (IoT). Assim como o assistente apresentado neste trabalho, o Snips também priorizava a privacidade por meio do processamento local dos dados. No entanto, o projeto foi descontinuado, o que limita seu uso.

3.1 Um Assistente de Voz Embarcado

O artigo de Lazzaroni, Bellotti e Berta (2024) discute o desenvolvimento de um assistente de voz integrado e *offline* para veículos inteligentes, buscando desenvolver um assistente capaz de operar totalmente *offline* em um ambiente embarcado, como veículos inteligentes, com desempenho comparável a soluções baseadas na nuvem.

A proposta visa mitigar questões de privacidade e latência associadas a assistentes de voz baseados na nuvem, utilizando técnicas de aprendizado por transferência (*Transfer Learning*). Segundo Weiss, Khoshgoftaar e Wang (2016), o aprendizado por transferência é usado para melhorar um aprendiz em um determinado domínio, transferindo informações de um domínio relacionado.

Essa abordagem é empregada para dar suporte ao reconhecimento automático de fala, adaptando modelos treinados com disponibilidade de grandes bases de dados de voz para uma língua com menor disponibilidade. O sistema visa oferecer controle por voz de funcionalidades do veículo, como aquecimento, navegação e controle de mídia, sem a

necessidade de conexão com a Internet, preservando a privacidade dos usuários.

O sistema é construído com módulos de reconhecimento de fala automática (ASR, *Automatic Speech Recognition*), entendimento de linguagem natural (NLU, *Natural Language Understanding*), e conversão de texto em fala (TTS, *Text-To-Speech*), todos otimizados para execução local em dispositivos embarcados com recursos limitados. A solução é comparada com modelos baseados em nuvem e sistemas embarcados existentes, destacando o seu desempenho superior em termos de latência e privacidade, sem perda significativa de precisão no reconhecimento de fala.

Os resultados experimentais mostram que o assistente de voz proposto alcança uma taxa de erro de palavras (WER, *Word Error Rate*) de 11,7%, um resultado comparável a soluções na nuvem, como Amazon Alexa e Google Assistant. Além disso, o tempo de resposta médio é inferior a um segundo, com baixo consumo de energia e memória, adequados para dispositivos embarcados. O sistema é capaz de operar de maneira robusta em ambientes com ruído, como veículos em movimento, e oferece uma extensibilidade fácil para outras línguas.

Pontos Positivos

- Privacidade: O sistema não depende de uma conexão com a Internet, eliminando preocupações com a transmissão de dados sensíveis para a nuvem;
- Baixa latência: A eliminação da necessidade de uma conexão à Internet reduz significativamente a latência nas respostas do assistente;
- Desempenho: O assistente apresenta desempenho comparável a soluções de ponta baseadas na nuvem em termos de reconhecimento de fala;
- Extensibilidade: A arquitetura do sistema é configurável e pode ser facilmente adaptada para suportar outras línguas.

Pontos Negativos

- Disponibilidade de Dados: A principal limitação é a falta de grandes bases de dados de voz para línguas minoritárias, como o italiano, o que impacta o treinamento do modelo;

- Desafios com *Transfer Learning*: Embora o *Transfer Learning* tenha sido eficaz, ainda existem limitações em adaptar completamente os modelos treinados em inglês para capturar com precisão todas as nuances fonéticas da língua utilizada.

3.2 Assistente Virtual para Rede Doméstica Controlado por Voz

No artigo de Martins (2020) foi priorizado a implementação de um assistente virtual para redes domésticas controlado por voz, que permita ao usuário, mesmo sem conhecimentos técnicos, gerenciar sua rede doméstica de maneira eficiente, diagnosticando e resolvendo problemas de conexão, além de permitir que utilizem funcionalidades como a criação de uma rede ou obter informações sobre o estado de sua rede.

O assistente virtual foi implementado em um ambiente de roteador que permite que múltiplos dispositivos se conectem à Internet e se comuniquem entre si, utilizando um único endereço IP público e pode diagnosticar e resolver problemas de rede através de comandos de voz, reduzindo a necessidade de suporte técnico tradicional.

O sistema foi desenvolvido usando tecnologias de assistentes virtuais e redes domésticas, integrando as plataformas OpenWRT e Amazon Alexa. A solução foi implementada em roteadores, que se comunicam com o assistente virtual via protocolo MQTT. O assistente foi configurado para receber comandos de voz e agir de acordo com eles, interagindo com o roteador para realizar ações como diagnóstico de conectividade e ajustes na rede.

Os resultados mostraram que o sistema trouxe melhorias significativas na experiência do usuário em comparação com os métodos tradicionais de suporte técnico. Durante os testes, usuários com diferentes níveis de conhecimento técnico conseguiram utilizar o assistente para diagnosticar problemas e realizar tarefas de gerenciamento de rede. O sistema, no entanto, apresentou algumas limitações na interação linguística, como a falta de suporte para diversas expressões e diálogos mais complexos.

Pontos Positivos

- Simplicidade na Interação: O sistema permitiu que usuários sem conhecimento

técnico pudessem realizar tarefas complexas de gerenciamento de rede com facilidade;

- Automação e Conforto: A integração de voz com o gerenciamento de redes trouxe mais conforto aos usuários, eliminando a necessidade de chamadas para suporte técnico;
- Flexibilidade: O sistema mostrou-se capaz de ser adaptado para outras funcionalidades no futuro, como a integração com outros dispositivos conectados na Internet das Coisas (IoT, *Internet of Things*).

Pontos Negativos

- Limitada Variedade de Comandos: O sistema ainda não consegue lidar com uma grande variedade de expressões verbais, o que limita a fluidez da interação;
- Dependência de Internet: Embora o assistente virtual opere em um ambiente doméstico, ele depende da Internet para realizar muitas de suas funções, o que pode ser um problema se a rede estiver completamente inativa.

3.3 Assistente Jaco

O trabalho de Bermuth, Poeppel e Reif (2022) aborda o desenvolvimento de um assistente de voz que funciona completamente *offline*, com foco em garantir a privacidade dos usuários. O assistente, chamado Jaco, foi projetado para rodar em dispositivos de baixa capacidade (recursos limitados), como o dispositivo Raspberry Pi, mantendo uma alta precisão no reconhecimento de fala. Ele se destaca por ser extensível, permitindo a adição de novas funcionalidades por meio de *skills*, além de ser compatível com vários idiomas, como alemão, inglês, espanhol e francês. O projeto visa resolver um problema comum em assistentes de voz tradicionais, como o Alexa da Amazon, que processam dados na nuvem e levantam preocupações sobre a segurança e a privacidade das informações dos usuários. O Jaco executa todo o processamento localmente, o que elimina a necessidade de enviar dados para servidores externos, tornando-o uma solução mais segura para o uso de assistentes de voz em ambientes domésticos e industriais.

A arquitetura do Jaco é modular, composta por um modulo satélite *Alpha*, outro modulo satélite *Beta*, todos distribuídos em diferentes ambientes e um nó central denominado *Master*, responsável por processar as requisições. O reconhecimento de fala é feito através de modelos pré-treinados, enquanto o entendimento de linguagem natural é realizado por meio de uma plataforma de inteligência artificial conversacional de código aberto denominada Rasa, com a geração de texto para fala feita pelo sintetizador de fala PicoTTS. A comunicação entre os módulos ocorre via protocolo MQTT, e a estrutura modular permite fácil substituição dos componentes, caso o desenvolvedor deseje usar outras soluções.

Pontos Positivos

- Privacidade: O Jaco roda completamente *offline*, garantindo que nenhum dado do usuário seja transmitido para a nuvem, protegendo as informações sensíveis e garantindo maior segurança;
- Extensibilidade: O assistente pode ser facilmente ampliado com novas funcionalidades, permitindo que desenvolvedores criem *skills* que acessam diretamente os recursos do dispositivo;
- Compatibilidade com vários idiomas: O sistema suporta múltiplas línguas e pode ser facilmente adaptado para novos idiomas com a adição de modelos de reconhecimento de fala específicos;
- Baixo consumo de recursos: Jaco foi otimizado para rodar em dispositivos com baixa capacidade de processamento, como o dispositivo Raspberry Pi, o que o torna acessível para uma ampla variedade de usuários e aplicações.

Pontos Negativos

- Complexidade na criação de *skills*: Embora o sistema permita a criação de *skills* personalizadas, o processo pode ser complexo para desenvolvedores iniciantes, especialmente quando comparado a outras soluções com interfaces mais simples;
- Reconhecimento limitado de nomes não usuais: Nos testes, Jaco apresentou dificuldades no reconhecimento de nomes de artistas ou músicas com ortografia incomum, o

que poderia ser melhorado com a implementação de mapas de pronúncia específicos.

3.4 Assistente Virtual em Python

Appalaraju et al. (2021) apresenta o desenvolvimento de um assistente pessoal virtual inteligente baseado em voz, implementado em Python, buscando criar um sistema que auxilie em atividades cotidianas por meio de comandos de voz, permitindo que o usuário foque em tarefas mais importantes.

O assistente foi capaz de realizar tarefas básicas como abrir arquivos, buscar informações, e fornecer respostas em formato de voz e texto, utilizar bibliotecas e ferramentas disponíveis no Python, integrando funções como reconhecimento de fala e síntese de voz além de reconhecer comandos de voz. O trabalho utilizou bibliotecas do Python como SpeechRecognition, Pyttsx3, e PyAudio para o reconhecimento de voz e conversão de texto para fala, com o apoio de APIs como a do Google para processar os comandos.

Os resultados do projeto desenvolvido por Appalaraju et al. (2021) indicam que o assistente virtual é funcional e eficiente em tarefas simples como abrir programas (ex.: Word); reproduzir vídeos no YouTube; buscar e exibir informações de filmes; datas e horários; enviar mensagens a um destinatário específico; responder com piadas e informações de entretenimento.

Pontos Positivos

- Simples e funcional: O sistema é básico, mas eficaz para as tarefas propostas;
- Integração com ferramentas existentes: O uso de bibliotecas Python disponíveis, como SpeechRecognition e Pyttsx3, facilita o desenvolvimento;
- Expansibilidade: O projeto pode ser facilmente expandido para incluir novas funcionalidades, como envio de e-mails e chamadas telefônicas, conforme a demanda.

Pontos Negativos

- Dependência de Internet: O sistema requer conexão com a Internet para utilizar as APIs de reconhecimento de voz, limitando seu uso em ambientes *offline*;

- Funcionalidades limitadas: Embora o projeto seja eficiente para tarefas simples, ele carece de recursos mais avançados, como aprendizado de máquina ou reconhecimento de padrões mais complexos.

3.5 Privacidade em Assistentes Digitais

Acosta e Reinhardt (2022) explora as questões de privacidade associadas aos Assistentes Digitais Controlados por Voz (VCDAs, *Voice-Controlled Digital Assistant*), onde o principal objetivo é analisar as ameaças à privacidade em assistentes de voz, mapear as soluções existentes e sugerir novas direções de pesquisa para melhorar a proteção dos dados dos usuários.

O trabalho revisa as ameaças à privacidade e propõe soluções baseadas nos princípios do Regulamento Geral de Proteção de Dados da União Europeia além disso, os autores propõem contramedidas para minimizar a coleta de dados sensíveis e garantir maior controle dos usuários sobre suas informações pessoais, soluções como processamento local de dados, ofuscação de voz e autenticação mais segura para melhorar a privacidade.

Pontos Positivos

- Análise Abrangente: O artigo fornece uma análise detalhada dos riscos de privacidade associados aos assistentes de voz, abrangendo tanto as interações intencionais quanto não intencionais;
- Soluções Propostas: Os autores apresentam várias soluções inovadoras, como o uso de mascaramento de voz e processamento local de dados, que podem melhorar significativamente a privacidade dos usuários;
- Apoio ao GDPR: O estudo foca em aplicar os princípios do GDPR, como minimização de dados e limitação de armazenamento, garantindo que as soluções propostas sigam regulamentos de proteção de dados já estabelecidos.

Pontos Negativos

- Falta de Implementação: Muitas das soluções sugeridas, como o processamento local de dados, ainda não são amplamente implementadas, o que limita sua aplicabilidade

prática no mercado atual;

- Complexidade para o Usuário: Algumas soluções propostas, como o mascaramento de voz e a substituição de palavras sensíveis, podem exigir um nível de configuração elevado, o que pode dificultar a adoção por usuários comuns;
- Privacidade dos Observadores: O artigo aborda que, em ambientes multiusuários, pessoas próximas ao dispositivo, como observadores, também têm suas conversas registradas sem consentimento, e as soluções para esses casos ainda são limitadas.

3.6 Snips

Coucke et al. (2018) apresenta a plataforma Snips, um sistema de entendimento de linguagem falada (SLU, *Spoken Language Understanding*) voltado para assistentes de voz embarcados em dispositivos IoT.

Seus objetivos são descrever a arquitetura de aprendizado de máquina por trás da plataforma de voz Snips, focando em desenvolver um sistema de ASR e NLU que funcione localmente em dispositivos de IoT, garantir a privacidade do usuário com um sistema que funcione sem o uso de servidores em nuvem, processando todos os dados localmente e treinar modelos de aprendizado de máquina que sejam leves, rápidos e robustos, adequados para dispositivos com recursos limitados, como processadores de baixo custo e pouca memória.

Os autores demonstram que o *Snips Voice Platform* é capaz de realizar inferências precisas e rápidas, com um tempo de processamento em tempo real em dispositivos como o Raspberry Pi 3. O sistema é composto por motores de ASR e NLU, que processam a fala do usuário, convertem-na em texto, identificam a intenção da fala e executam ações correspondentes. Além disso, a plataforma permite personalização local, sem comprometer a privacidade, e o treinamento de modelos de linguagem adequados a diferentes domínios, como assistentes para casas inteligentes.

Pontos Positivos

- Privacidade por design: O principal destaque do projeto é a forte ênfase na privacidade, com toda a execução acontecendo localmente no dispositivo, eliminando a

necessidade de coleta de dados na nuvem;

- Leveza e eficiência: A arquitetura é otimizada para funcionar em dispositivos com hardware limitado, como o Raspberry Pi, mantendo precisão e velocidade aceitáveis;
- Portabilidade: O sistema foi desenvolvido para ser portátil, suportando uma variedade de dispositivos com diferentes sistemas operacionais e arquiteturas de hardware, como Linux e Android.

Pontos Negativos

- Dependência de configurações de hardware: Embora a plataforma funcione bem em dispositivos como o Raspberry Pi, seu desempenho pode ser comprometido em dispositivos ainda mais limitados, ou quando se tenta executar assistentes mais complexos;
- Necessidade de personalização avançada: Embora ofereça privacidade, o processo de personalização para diferentes usuários ou domínios pode exigir maior intervenção técnica, o que pode dificultar a adoção por usuários sem conhecimento técnico;
- Complexidade de implementação: O uso de técnicas avançadas, como a combinação de redes neurais com HMM, e o código utilizando a linguagem de programação *Rust* para melhorar o desempenho, adiciona complexidade ao desenvolvimento e à manutenção do sistema;
- Descontinuado: O Snips era *open source* até janeiro de 2020 quando o projeto foi descontinuado.

3.7 Considerações Finais

Ao analisar os trabalhos relacionados apresentados nas seções anteriores foi possível perceber algumas similaridades e diferenças em seus conteúdos. A Tabela 3.1 trás um comparativo dessas similaridades e diferenças, pontuadas em critérios como, operação do assistente 100% *offline*, o trabalho possui processamento local de dados, zelou pela privacidade do

usuário, avaliou vulnerabilidades, há dependência de Internet, está em conformidade com LGPD / GDPR e se o assistente proposto foi implementado:

Tabela 3.1: Tabela Comparativa Entre os Trabalhos Relacionados

Critério	Lazzaroni	Martins	Bermuth	Appalaraju	Acosta & Reinhardt	Coucke	Este Trabalho
Operação 100% <i>offline</i>	✓	– ¹	✓	–	✓ ²	✓	✓
Processamento local de dados	✓	–	✓	–	✓	✓	✓
Privacidade	Média ³	Baixa	Alta	Média	Alta	Alta	Alta
Avaliação de vulnerabilidades	Média	Baixa	Alta	Baixa	Alta	Média	Alta
Dependência de Internet	–	✓	–	✓	–	–	–
Conformidade LGPD / GDPR	Implícita	–	Implícita	–	Explícita	Explícita	Explícita
Multiplataforma	–	–	–	–	–	✓	✓
Implementação funcional	–	–	–	–	–	–	✓

¹ Opera *offline* só para comandos locais, mas depende da nuvem para diagnósticos.

² Propostas de contramedidas sem implementação de sistema *offline*.

³ Processamento local sem análise explícita.

Na Tabela 3.1, observa-se que este trabalho é um dos poucos que atinge todos os critérios de segurança analisados: operação 100% *offline*, ausência total de dependência de Internet, conformidade explícita com princípios da LGPD, e avaliação consistente das vulnerabilidades envolvidas no uso de assistentes, garantindo que a arquitetura e as bibliotecas utilizadas (como Vosk e pyttsx3) estejam alinhadas com a minimização de dados e não retenção de histórico de voz.

4 Assistente Virtual *Offline* Desenvolvido

4.1 Visão Geral

Este capítulo apresenta o assistente virtual desenvolvido com foco em reconhecimento de voz *offline*, considerando aspectos de segurança e privacidade. A principal motivação para a criação deste assistente foi fornecer uma alternativa viável às soluções baseadas na nuvem, como o Copilot (*Windows*) e a Siri (*macOS*), que podem apresentar riscos relacionados ao envio de dados pessoais para servidores na Internet.

No caso do Copilot, assistente integrado ao *Windows* a partir da versão 11, a política de privacidade do *Copilot*² enfatiza que determinados dados como informações de diagnóstico e uso são coletados, armazenados e processados em servidores na nuvem com o intuito de melhorar os serviços, personalizar a experiência do usuário e garantir a segurança do sistema. Conforme consta na política de privacidade há um trecho que destaca: A Microsoft coleta dados de você, por meio de nossas interações com você e por meio de nossos produtos. Você fornece alguns desses dados diretamente, enquanto alguns deles obtemos ao coletar dados sobre suas interações, uso e experiências com nossos produtos. Os dados que coletamos dependem do contexto de suas interações com a Microsoft e as opções que você escolhe, incluindo as suas configurações de privacidade e os produtos e recursos que você usa. Também obtemos dados sobre você por meio de afiliadas, subsidiárias da Microsoft e terceiros.³ Essa abordagem, contudo, levanta preocupações quanto à exposição de informações pessoais, principalmente quando os dados passam por múltiplos pontos de armazenamento e processamento.

De forma semelhante, a Siri, assistente virtual da Apple integrada ao *macOS*, também realiza parte significativa do processamento por meio de servidores na nuvem. Segundo a política de privacidade da Apple⁴, embora a empresa afirme adotar medidas para proteger a privacidade dos usuários, está previsto em como a Apple compartilha os

²<https://learn.microsoft.com/pt-br/copilot/privacy-and-protections>

³<https://www.microsoft.com/pt-br/privacy/privacystatement>

⁴<https://www.apple.com/legal/privacy/br>

dados pessoais: A Apple pode compartilhar dados pessoais com empresas afiliadas da Apple, prestadores de serviço que agem em nosso nome, nossos parceiros, desenvolvedores e editores ou outras pessoas de acordo com suas orientações. Apesar de a Apple destacar que "não compartilha dados pessoais com terceiros para seus propósitos de marketing", ainda há a possibilidade de circulação de informações entre múltiplas entidades envolvidas no processamento, o que pode gerar preocupações quanto à confidencialidade e ao controle total dos dados por parte do usuário.

Dessa forma, o assistente virtual desenvolvido neste projeto opera completamente *offline*. Todo o processamento é realizado localmente, garantindo que os dados dos usuários permaneçam sob seu exclusivo controle e reduzindo significativamente os riscos associados ao armazenamento e à análise de informações na nuvem.

4.2 Definições e Etapas de Desenvolvimento

As bibliotecas escolhidas foram definidas através de testes de funcionalidade, destacando-se por sua capacidade de operar eficientemente em modo *offline*. A biblioteca Vosk⁵ foi selecionada para implementação prática do reconhecimento de voz *offline*, que combina um modelo acústico híbrido de redes neurais profundas (DNN, *Deep Neural Network*) com um decodificador baseado em Modelos Ocultos de Markov (HMM, *Hidden Markov Model*). O Vosk oferece um modelo pré-treinado em português brasileiro, capaz de identificar fonemas. Com essa arquitetura, o assistente consegue processar comandos de voz inteiramente no dispositivo, sem depender de servidores externos, garantindo maior privacidade dos dados. A biblioteca pyttsx3⁶, foi selecionada para síntese de voz, destacando-se pela sua documentação extensa. Além disso, foram integradas ferramentas como a biblioteca subprocess para manipulação de processos do sistema e pyautogui⁷ para interação com a interface gráfica do usuário, a biblioteca pyautogui ao longo do projeto, se mostrou ineficiente para acentuação de palavras e acabou sendo necessária sua exclusão do projeto.

A Figura 4.1 sintetiza todo o percurso de construção do assistente virtual *offline* em um único fluxo iterativo, começando pela seleção e integração das ferramentas mais

⁵<https://alphacephei.com/vosk>

⁶<https://pypi.org/project/pyttsx3/>

⁷<https://pyautogui.readthedocs.io/en/latest/>

adequadas ao funcionamento sem dependência de Internet. Nessa primeira fase, avaliou-se cada biblioteca quanto à capacidade de operação *offline*, à documentação disponível e à facilidade de integração; apenas quando esses critérios eram plenamente atendidos é que o processo avançava, senão retornava-se à escolha de uma alternativa.

Em seguida, desenvolveu-se a interface gráfica mínima, composta por botões e acionamento por voz, garantindo que as interações do usuário fossem corretamente capturadas e que o sistema fornecesse o *feedback* esperado, diante de qualquer falha de usabilidade ou de retorno de resposta, refinou-se imediatamente a lógica de captura de eventos.

A terceira etapa concentrou-se na implementação do reconhecimento de fala em tempo real, utilizando o Vosk com filtros de ruído e ajustes de sensibilidade: somente quando as palavras efetuavam algum dos comandos identificamos como aceitável, caso contrário revisavam-se os parâmetros de pré-processamento ou experimentavam-se modelos alternativos.

Por fim, as transcrições validadas passaram a ser mapeadas em comandos do sistema operacional, aberturas de programas, controle de janelas, executados via subprocessos e utilizando a biblioteca PyAutoGUI, com confirmação em síntese de voz pelo pyttsx3. Eventuais falhas nessa etapa levavam a ajustes no mapeamento do texto em ação ou no tratamento de exceções, até que todas as ações ocorressem corretamente (exceto em casos de não reconhecimento por parte do Vosk).

Em cada uma dessas fases, o ponto de decisão ilustrado pelo losango no fluxograma assegurou a conformidade com os requisitos de funcionalidade, desempenho, usabilidade e segurança antes de prosseguir, garantindo que o produto final fosse um assistente robusto, privado e efetivamente *offline*.

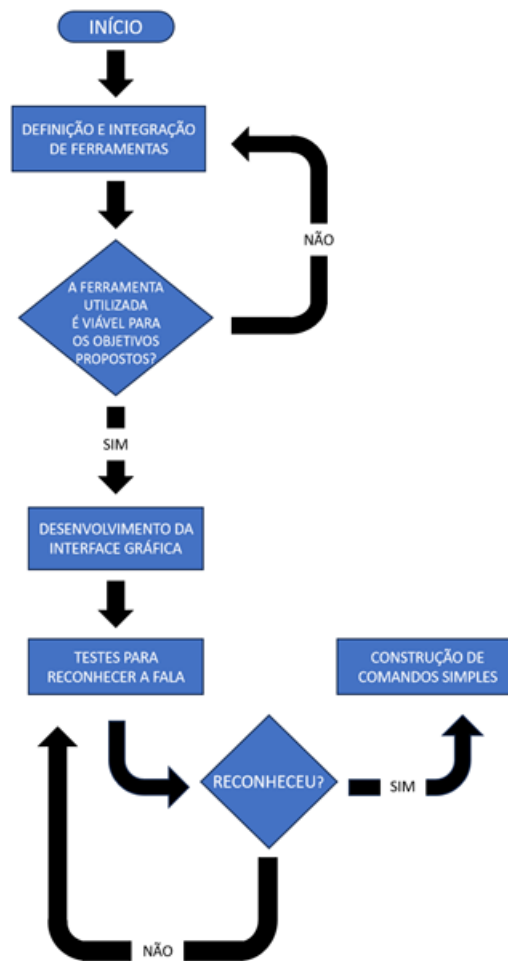


Figura 4.1: Ilustração de fluxo metodológico representando as tomadas de decisão em cada etapa do desenvolvimento dos métodos do trabalho.

4.3 Referencias Técnicas

O assistente virtual desenvolvido emprega a biblioteca Vosk para realizar o reconhecimento de fala inteiramente em modo *offline*, valendo-se de um modelo acústico híbrido que combina redes neurais profundas com um decodificador baseado em Modelos Ocultos de Markov, o que permite converter o áudio capturado pelo microfone em texto sem depender de servidores externos, garantindo rapidez e privacidade dos dados do usuário.

Para o pré-processamento do sinal de áudio, foi utilizado o PyAudio, que fornece uma interface estável para capturar o som do microfone em tempo real, permitindo ainda

a aplicação de filtros de ruído e ajustes de sensibilidade antes de alimentar o Vosk. Após o texto ser produzido, a biblioteca `pyttsx3` é responsável pela síntese de voz, transformando as respostas do sistema em áudio sintetizado localmente, sem qualquer chamada a APIs *online*, o que reforça o caráter *offline* da solução e oferece *feedback* auditivo imediato ao usuário.

Para a execução dos comandos reconhecidos, utiliza-se a biblioteca padrão *subprocess* do *Python*, que invoca processos do sistema operacional de forma segura e flexível, permitindo abrir aplicativos e executar *scripts* conforme as instruções de voz. A interação com a interface gráfica do usuário se dá por meio de botões que podem executar os comandos ou clicar na opção falar para que capture o áudio do usuário, assim como presente na figura 4.2 abaixo,

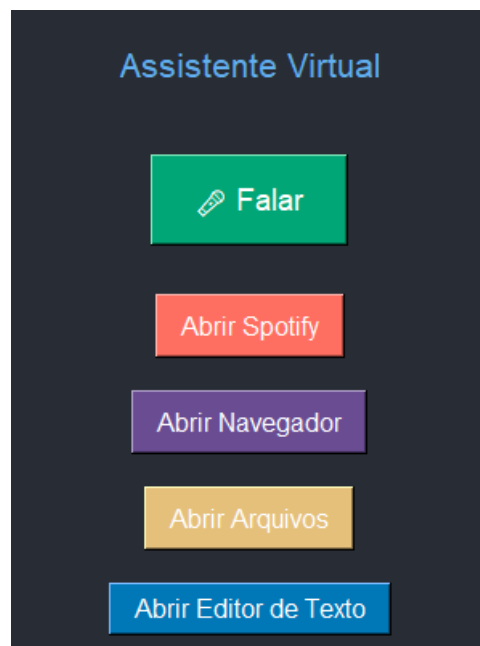


Figura 4.2: Ilustração do Assistente Virtual Desenvolvido.

Os cliques e movimentação de janelas, ficaram a cargo da biblioteca `PyAutoGUI`, que simula eventos de mouse e teclado de maneira programática, garantindo que comandos como “texto” do ditado sejam cumpridos sem intervenção manual, embora tenha sido necessário ajustar o tratamento de acentuação nos textos transcritos para manter a naturalidade das interações. Dessa forma, a escolha e integração dessas ferramentas técnicas asseguram que o assistente opere de forma totalmente *offline*, mantida a privacidade e a

autonomia do usuário.

4.4 Natureza dos Testes Realizados

Para avaliar o comportamento dos assistentes virtuais, foi elaborado um protocolo de testes executado em ambientes controlados. Inicialmente, configurou-se o uso de uma máquina virtual (*Virtual Machine*) com sistema operacional Windows para o Copilot e foi utilizado um MacBookAir para o teste com a assistente Siri, no macOS e o assistente desenvolvido neste trabalho em ambos os sistemas operacionais.

Em seguida, os assistentes foram submetidos a uma série de comandos básicos, como a abertura de aplicativos e a resposta a consultas simples, a fim de aferir a funcionalidade e o desempenho de cada sistema em condições de isolamento. Durante essa execução, utilizou-se o software Wireshark para monitorar e registrar quaisquer tentativas de comunicação com servidores na Internet. O objetivo dessa fase era identificar, de forma precisa, se os assistentes buscavam transmitir dados, evidenciando, assim, seu grau de aderência aos critérios de privacidade estabelecidos.

Por fim, os dados coletados foram analisados de forma integrada, possibilitando a mensuração do tempo de resposta, a eficácia do processamento dos comandos e a independência dos assistentes em relação à comunicação com servidores remotos. Essa metodologia proporcionou subsídios para a comparação dos sistemas, fundamentando a análise dos aspectos de segurança e privacidade.

4.5 Resultados

Para cada assistente virtual (Copilot no Windows e Siri no macOS) abriu-se o assistente, solicitou-se uma informação simples e em seguida o fechou-se, repetindo o processo 10 vezes. As figuras 4.3 e 4.4 a seguir mostram, respectivamente, o tráfego TLS gerado pelo Copilot em comunicação com servidores da Microsoft, nela, identifica-se claramente diversas sessões TLS estabelecidas entre o cliente local e servidores da Microsoft, identificáveis pelos endereços IP remotos e pelas portas padrão 443. É possível observar o *handshake* inicial (*ClientHello*, *ServerHello*), seguido da troca de chaves e da transmissão de dados

cifrados.

Já o TLS *ClientHello* enviado pela Siri ao servidor iCloud pode-se observar que além do *handshake*, destacam-se extensões como SNI (Server Name Indication), que revelam o nome do *host* ao qual o assistente se conecta, neste caso, um subdomínio da *iCloud* usado para sincronização de contatos. O detalhamento dessas extensões evidencia que a Siri não apenas transmite o áudio para reconhecimento, mas também sincroniza informações de perfil e dados de uso, o que amplia o escopo de dados trafegados. Com base nessa captura, concluímos que a Siri depende fortemente de infraestrutura de nuvem, não apenas para entender comandos de voz, mas também para manter e atualizar dados pessoais do usuário, reforçando a vantagem de soluções locais para cenários em que a privacidade é prioritária.

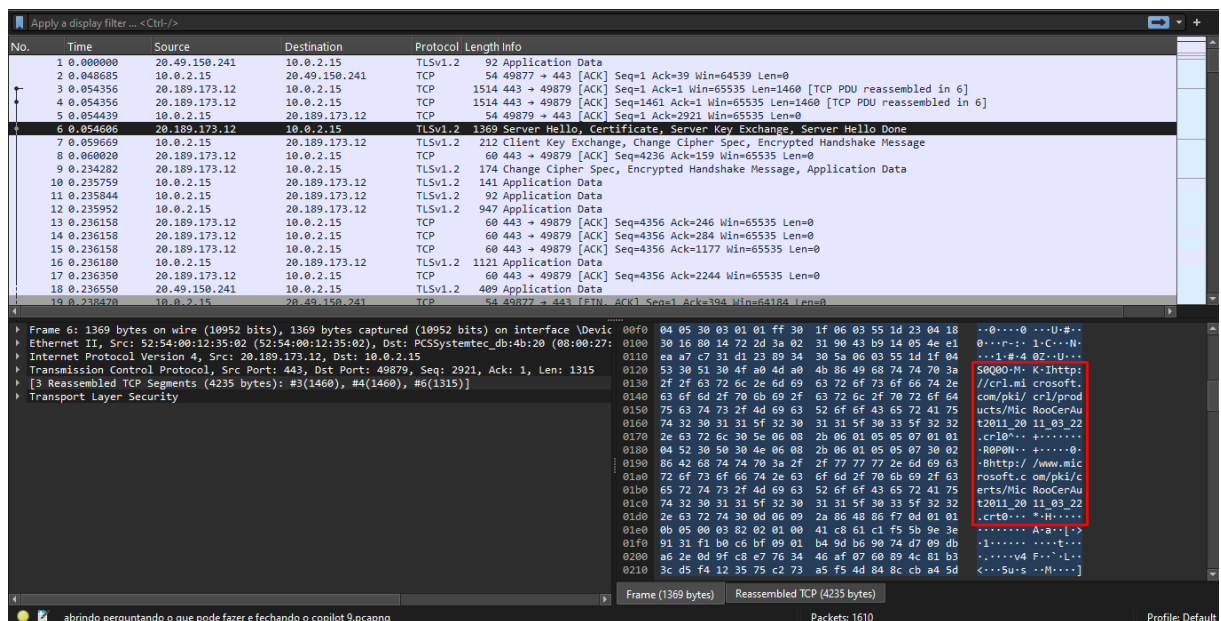


Figura 4.3: Tráfego TLS entre o Copilot e os servidores da Microsoft (Wireshark).

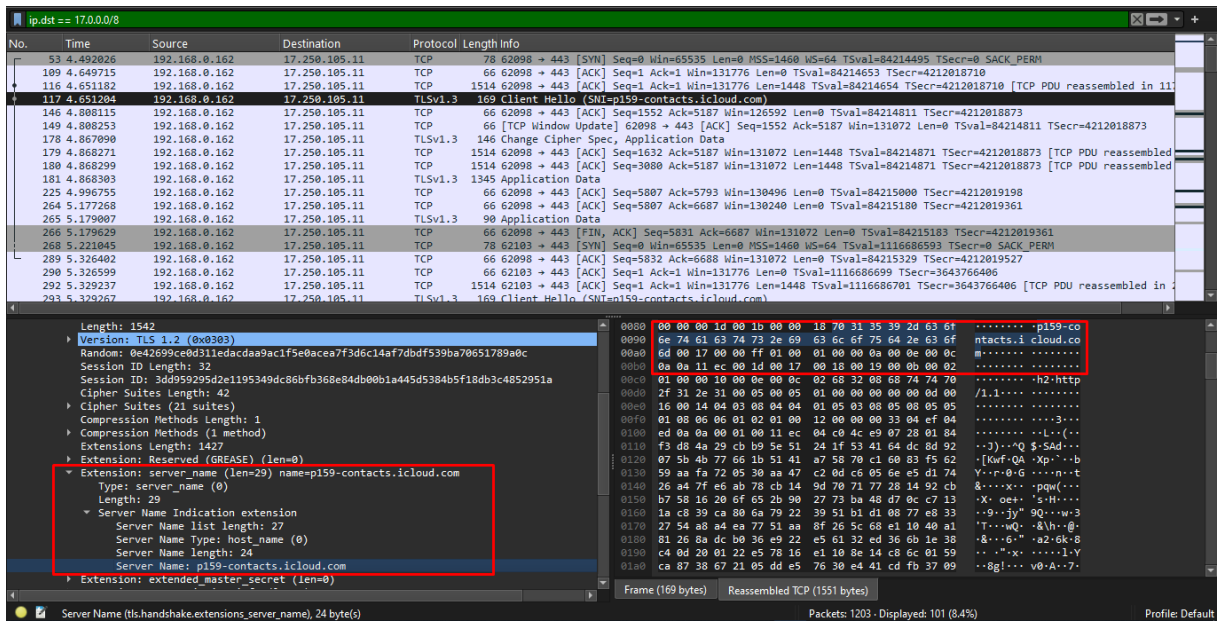


Figura 4.4: TLS ClientHello da Siri para `p159-contacts.icloud.com` contendo extensão SNI (Wireshark).

Durante a realização dos testes, o assistente não gerou nenhum tipo de pacote de rede, o que impossibilita a apresentação de capturas de tráfego (prints) como evidência. Entretanto, essa ausência de tráfego confirma justamente o objetivo principal do trabalho: ao não transmitir dados para servidores externos, o assistente reforça a privacidade do usuário. Dessa forma, pode-se afirmar que o reconhecimento de fala em modo offline atua como um mitigador relevante do problema de privacidade em assistentes virtuais, já que todo o processamento ocorre localmente no dispositivo. Todos os dados de testes e código do assistente podem ser encontrados no *github* do trabalho ⁸

Adicionalmente, foram conduzidos testes de ditado para avaliar o desempenho do reconhecimento de fala do assistente desenvolvido. No primeiro cenário, em que foram ditadas palavras isoladas, observou-se um índice de aproximadamente 33% de erros, com maior dificuldade no reconhecimento de fonemas como “Z”, “H” e “S”. Esse resultado indica que o sistema ainda carece de ajustes para lidar com palavras soltas em ambientes de maior variabilidade sonora.

⁸<https://github.com/Jeansfeliipe/TccAssistenteVirtual>

bala	hoje	ralo
boto	jaca	roupa
bule	jiló	rodo
boca	jipe	rato
bico	jogo	rima
boné	joia	robô
cana	juba	saia
café	limão	sala
caju	lixo	sofá
cama	luva	sola
copo	lobo	sete
cola	lata	sino
dado	laje	tio
dama	mala	teto
data	mãe	tatu
dedo	mamão	totó
doze	mico	time
dia	mofo	toca
foto	mola	vaca
fubá	nada	vela
fumo	neto	vida
fila	nome	vovó
faca	nota	vila
feio	nuca	veia
galo	nove	xale
gato	pano	xixi
gula	pato	xodó
gota	papai	zona
galã	pavão	zaga
gude	pipa	zelo
hino	peixe	zoo

Figura 4.5: Ditado de Palavras Sem Contexto

Esses resultados, detalhados na figura a seguir, reforçam a limitação do sistema no processamento de unidades de fala descontextualizadas, evidenciando a necessidade de ajustes para esse tipo de cenário.

bala boto boyle boca o bico olé cana a ser azul cama copo cola um dado dama data dedo doze dia foto bar fumo fila faca sério galo gato gorila conta galão onde e no hoje já cá genoa jipe e jogo jóia juba limão lixo luva lobo lata lage mala mãe mamão mico mofo mala nada neto nome nota nunca nove ano pato papai pavão pipa peixe ralo roupa quando rato bom sala surfar cela sete sino tio teto tatu total time toca faca vela vida vovó vila venha chalé xixi shadow joana usada exemplo já lógico

Figura 4.6: Resultado do Ditado de Palavras Sem Contexto

Por outro lado, quando o teste foi repetido utilizando um texto corrido, o desempenho apresentou melhora significativa. De um total de 69 palavras ditadas, apenas

2 foram transcritas incorretamente, evidenciando que o sistema se mostra mais eficaz na compreensão de frases contínuas do que na interpretação de termos isolados.

Na quinta, o senhor Antoninho começa a tratar dos animais logo pela manhã. Primeiro, dá milho às galinhas e aos patinhos, trata da porca e dos porquinhos, dos coelhos malhados, das ovelhas e das vacas. Que grande barulheira fazem os animais durante a hora da refeição!

Os netos do senhor Antoninho adoram acompanhar o rebanho pelo campo, observar gafanhotos, aranhas, joaninhas e abelhas e cheirar os perfumes da Natureza.

Figura 4.7: Ditado de Palavras em Formato de Texto Corrido

Nos testes com ditado em formato de texto corrido, o desempenho foi consideravelmente superior em relação às palavras isoladas. O reconhecimento mostrou-se mais estável, com menor ocorrência de erros, especialmente devido à utilização do contexto linguístico, que auxilia o modelo a prever e corrigir ambiguidades durante a transcrição. A figura a seguir apresenta em detalhe esses resultados, evidenciando a maior eficiência do sistema na interpretação de frases contínuas.

na quinta o senhor antonio começa a tratar dos animais logo pela manhã primeiro dar milho às galinhas e aos patinhos trata da porca e dos porquinhos dois coelhos malhados das ovelhas e das vacas que grande barulheira fazem os animais durante a hora da refeição os netos do senhor antoninho adoram acompanhar o rebanho pelo campo observar gafanhotos aranhas joaninhas e abelhas e cheirar os perfumes da natureza

Figura 4.8: Resultado Ditado de Palavras Texto Corrido

Esses experimentos confirmam que, embora haja espaço para aprimoramento no reconhecimento de palavras individuais, o assistente já apresenta um funcionamento consistente para interações mais naturais em formato de frases, reforçando sua viabilidade prática em cenários de uso cotidiano.

5 Considerações Finais

Este trabalho teve como objetivo avaliar se o reconhecimento de voz *offline* pode ser uma solução para mitigar os problemas relacionados ao uso contínuo da voz em assistentes virtuais, em especial no que diz respeito à privacidade e à proteção dos dados. Os resultados obtidos confirmam que a abordagem *offline* elimina a necessidade de transmissão de informações para a nuvem, uma vez que, durante os testes realizados, não foi identificado nenhum tráfego de rede associado ao funcionamento do assistente.

Essa característica garante que todos os dados de voz sejam processados localmente, reduzindo significativamente a superfície de ataque e os riscos de interceptação ou uso indevido de informações pessoais. Assim, pode-se concluir que o reconhecimento de fala *offline* constitui um mitigador relevante dos problemas de privacidade em assistentes virtuais, mostrando-se uma alternativa viável para cenários que exigem maior controle e segurança sobre os dados do usuário.

5.1 Conquistas e Limitações

O desenvolvimento deste trabalho, trouxe consigo algumas conquistas e aprendizagens bastante valiosas. Primeiro, demonstrar na prática que é possível oferecer um assistente de voz totalmente *offline*, o que reforça um ponto importante hoje na tecnologia: privacidade, sem abrir mão da usabilidade de um assistente. Além disso, ao adaptar ferramentas como o Vosk e o PyAudio para funcionar estável em Windows e MacOS, há o ganho de experiência em engenharia de software multiplataforma, lidando com problemas reais de compatibilidade e usabilidade.

Além dos pontos positivos, também houveram problemas, durante a fase inicial de testes, utilizamos a biblioteca pyautogui para controlar o assistente, mas logo percebemos que todas as palavras ditas perdiam a acentuação, por exemplo, “ação” era transcrito como “acao”, o que comprometeu tanto a naturalidade das interações quanto a cobertura de cenários mais complexos. Além disso, embora o projeto tivesse sido idealizado para

ser multiplataforma (Windows e Linux), passamos vários meses avaliando diferentes assistentes no ambiente Linux para comparar com o assistente desenvolvido, e nos deparamos com documentações imprecisas ou interfaces gráficas que não eram funcionais, atrasando significativamente o desenvolvimento nessa plataforma. Por fim, durante os testes de reconhecimento de voz, identificamos uma ambiguidade recorrente: o modelo confundia a palavra “fim” com “sim”, acionando comandos indevidos e exigindo ajustes finos tanto no limiar de confiança quanto no pós-processamento das transcrições.

5.2 Cumprimento dos Objetivos

O desenvolvimento do assistente virtual atingiu os objetivos propostos, consolidando-se como uma solução multiplataforma, capaz de operar de forma consistente em *Windows*, *Linux* e *macOS*. Essa compatibilidade amplia a aplicabilidade da ferramenta, permitindo que usuários de diferentes ambientes usufruam da mesma experiência de privacidade, segurança e usabilidade. Além disso, ficou evidente que, ao contrário das soluções comerciais testadas (*Copilot* no *Windows* e *Siri* no *macOS*), que durante os testes encaminharam pacotes de rede a servidores externos, o assistente virtual desenvolvido não gerou nenhum tráfego de rede, assegurando o isolamento completo do processamento de voz.

5.3 Trabalhos Futuros

Como trabalhos futuros, sugere-se um conjunto estruturado de iniciativas para aprimorar o assistente.

Inicialmente, recomenda-se o aperfeiçoamento do reconhecimento de voz por meio do treinamento de modelos acústicos específicos, voltados à redução da Taxa de Erro de Palavras (WER) em ambientes ruidosos, bem como a implementação de técnicas de adaptação de locutor e a busca por bases de dados maiores para garantir maior precisão.

Em seguida, propõe-se a incorporação de novas funcionalidades, como controle de dispositivos IoT, envio de e-mails, gerenciamento de calendário local.

A terceira frente envolve o desenvolvimento de uma interface gráfica mais refinada, capaz de fornecer em tempo real, um retorno ao usuário, do texto reconhecido, ajustes de

sensibilidade do microfone e opções de seleção de idioma.

Além disso, sugere-se a realização de testes de usabilidade abrangentes com usuários reais, seguidos de avaliações específicas de acessibilidade junto a pessoas com deficiências visual, auditiva ou motora, de modo a validar e refinar cada recurso.

Por fim, destaca-se a importância de elaborar documentação e tutoriais em formatos alternativos, como PDF acessível, vídeos legendados e manuais em áudio, para atender a diferentes perfis de usuário e ampliar o alcance da solução.

Bibliografia

- ACOSTA, L. H.; REINHARDT, D. A survey on privacy issues and solutions for voice-controlled digital assistants. *Pervasive and Mobile Computing*, Elsevier, v. 80, p. 101523, 2022.
- ALENCAR, A. J.; SCHMITZ, E. A.; CRUZ, L. T. *Assistentes Virtuais Inteligentes: Conceitos e estratégias*. Rio de Janeiro: Brasport, 2013.
- ALEPIS, E.; PATSAKIS, C. Monkey says, monkey does: Security and privacy on voice assistants. *IEEE Access*, v. 5, p. 17841–17851, 2017.
- APPALARAJU, V.; RAJESH, V.; SAIKUMAR, K.; SABITHA, P.; KIRAN, K. R. Design and development of intelligent voice personal assistant using python. In: IEEE. *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*. Greater Noida, India, 2021. p. 1650–1654.
- BERMUTH, D.; POEPEL, A.; REIF, W. Jaco: An offline running privacy-aware voice assistant. In: IEEE. *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. Sapporo, Japan, 2022. p. 618–622.
- BRUNIALTI, L.; PERES, S.; FREIRE, V.; LIMA, C. Aprendizado de máquina em sistemas de recomendação baseados em conteúdo textual: Uma revisão sistemática. In: *Anais do XI Simpósio Brasileiro de Sistemas de Informação*. Porto Alegre, RS, Brasil: SBC, 2015. p. 203–210. ISSN 0000-0000. Disponível em: <https://sol.sbc.org.br/index.php/sbsi/article/view/5818>.
- CAMBRIA, E.; WHITE, B. Jumping NLP curves: A review of natural language processing research [review article]. *IEEE Computational Intelligence Magazine*, v. 9, n. 2, p. 48–57, 2014.
- COUCKE, A.; SAADE, A.; BALL, A.; BLUCHE, T.; CAULIER, A.; LEROY, D.; DOUMOURO, C.; GISSELBRECHT, T.; CALTAGIRONE, F.; LAVRIL, T. et al. Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. *arXiv preprint arXiv:1805.10190*, 2018.
- DUA, M.; AKANKSHA; DUA, S. Noise robust automatic speech recognition: review and analysis. *International Journal of Speech Technology*, v. 26, p. 475–519, 2023.
- FERRARO, A.; GALLI, A.; GATTA, V. L.; POSTIGLIONE, M. Benchmarking open source and paid services for speech to text: an analysis of quality and input variety. *Frontiers in Big Data*, v. 6, p. 1210559, 2023. Disponível em: <https://doi.org/10.3389/fdata.2023.1210559>.
- GUZMAN, A. L. Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior*, 2019.
- HALAL, W. E. Artificial intelligence is almost here. *On the Horizon*, MCB UP Ltd, v. 11, n. 2, p. 37–38, 2003.

- HOY, M. B. Alexa, siri, cortana, and more: An introduction to voice assistants. *Medical Reference Services Quarterly*, Routledge, v. 37, n. 1, p. 81–88, 2018.
- KALHOR, B.; DAS, S. Evaluating the security and privacy risk postures of virtual assistants. *arXiv preprint arXiv:2312.14633*, 2023.
- LAZZARONI, L.; BELLOTTI, F.; BERTA, R. An embedded end-to-end voice assistant. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 136, p. 108998, 2024.
- MARTINS, D. L. Implementação de um assistente virtual para rede doméstica controlado por voz. 2020.
- MASNIKOSA, V. P. The fundamental problem of an artificial intelligence realization. *Kybernetes*, MCB UP Ltd, v. 27, n. 1, p. 71–80, 1998.
- METAXIOTIS, K.; ERGAZAKIS, K.; SAMOUILIDIS, E.; PSARRAS, J. Decision support through knowledge management: the role of the artificial intelligence. *Information Management & Computer Security*, MCB UP Ltd, v. 11, n. 5, p. 216–221, 2003.
- OKE, S. Artificial intelligence: A review of the literature. In: *Proceedings of the Fifth Asia Pacific industrial engineering and management systems conference*. [S.l.: s.n.], 2004.
- WEISS, K.; KHOSHGOFTAAR, T. M.; WANG, D. A survey of transfer learning. *Journal of Big Data*, v. 3, n. 9, 2016. Disponível em: <https://doi.org/10.1186/s40537-016-0043-6>.