

**FACULDADE DE TECNOLOGIA DE SÃO PAULO – FATEC SP**  
DEPARTAMENTO DE TECNOLOGIA DE INFORMAÇÃO

LUCAS VINICIUS MARTINS DO AMARAL

ENVIO E RECEBIMENTO DE DADOS NOS COMANDOS NA ASSISTENTE  
VIRTUAL ALEXA EM DISPOSITIVOS ECHO (ECHO DOT)

SÃO PAULO  
DEZEMBRO - 2021

LUCAS VINICIUS MARTINS DO AMARAL

ENVIO E RECEBIMENTO DE DADOS NOS COMANDOS NA ASSISTENTE  
VIRTUAL ALEXA EM DISPOSITIVOS ECHO (ECHO DOT)

Trabalho submetido como exigência parcial  
para a obtenção do Grau de Tecnólogo em  
Análise e Desenvolvimento de Sistemas  
Orientador: Prof(a). Me. Edméa Pujol Cantón

SÃO PAULO  
DEZEMBRO - 2021

FACULDADE DE TECNOLOGIA DE SÃO PAULO

LUCAS VINICIUS MARTINS DO AMARAL

ENVIO E RECEBIMENTO DE DADOS NOS COMANDOS NA ASSISTENTE  
VIRTUAL ALEXA EM DISPOSITIVOS ECHO (ECHO DOT)

Trabalho submetido como exigência parcial para a obtenção do Grau de  
Tecnólogo em Análise e Desenvolvimento de Sistemas.

Parecer do Professor Orientador

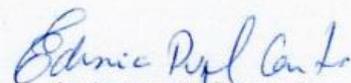
O TCC do aluno Lucas Vinicius Martins do Amaral  
atendeu a todas as exigências do DTI para Trabalhos  
de Conclusão de Curso.

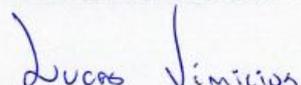
Conceito/Nota Final: 100 (Dez inteiros)

**Atesto o conteúdo contido na postagem do ambiente TEAMS pelo aluno e  
assinada por mim para avaliação do TCC.**

Orientador: Prof(a). Me. Edméa Pujol Cantón

SÃO PAULO, 02 de Dezembro de 2021.

  
Assinatura do Orientador

  
Assinatura do aluno

Dedico este trabalho à minha família por estar ao meu lado e me apoiar no começo desta jornada, aos meus colegas e amigos e aos professores que mais impactaram no meu caminho.

## **AGRADECIMENTOS**

Este trabalho não poderia ser terminado sem a ajuda de diversas pessoas e/ou instituições às quais presto minha homenagem. Certamente esses parágrafos não irão atender a todas as pessoas que fizeram parte dessa importante fase de minha vida. Portanto, desde já peço desculpas àquelas que não estão presentes entre estas palavras, mas elas podem estar certas de que fazem parte do meu pensamento e de minha gratidão.

À minha família, pelo carinho, incentivo e total apoio em todos os momentos da minha vida.

À minha orientadora, que me mostrou os caminhos a serem seguidos e pela confiança depositada.

A todos os professores e colegas do departamento, tanto da FATEC São Paulo, quanto da FATEC Jundiaí, que ajudaram de forma direta e indireta na conclusão deste trabalho.

Enfim, a todos os que, de alguma forma, contribuíram para a realização deste trabalho.

Primeira Lei: Um robô não pode ferir um ser humano ou, por omissão, permitir que um ser humano sofra algum mal.

Segunda Lei: Um robô deve obedecer às ordens que lhe sejam dadas por seres humanos, exceto nos casos em que tais ordens contrariem a Primeira Lei.

Terceira Lei: Um robô deve proteger sua própria existência desde que tal proteção não entre em conflito com a Primeira e Segunda Leis.

(ASIMOV, Isaac, 1950).

## RESUMO

[Alexa](#) é a Inteligência Artificial presente nos dispositivos Echo, usada como assistente virtual. O objetivo desse trabalho é analisar o tráfego dos dados dos comandos enviados pelo usuário. Coletamos todas as informações sobre o hardware e os serviços na Alexa Cloud através da pesquisa bibliográfica. O propósito deste trabalho de conclusão de curso é analisar o tráfego dos dados, a segurança do usuário, e tornar acessível para a língua portuguesa as informações sobre a [Alexa](#).

**Palavras-chave:** [Alexa](#). *Echo Dot*. *Alexa Cloud*.

## ABSTRACT

[Alexa](#) is the Artificial Intelligence present in Echo devices, used as a virtual assistant. The objective of this work is to analyze the data traffic of commands sent by the user. We collect all information about hardware and services on Alexa Cloud through literature search. The purpose of this course completion work is to analyze data traffic, user security, and make information about [Alexa](#) accessible to Portuguese.

**Keywords:** [Alexa](#). *Echo Dot. Alexa Cloud.*

## LISTA DE FIGURAS

FIGURA 1: RADIO REX DE 1920. PRIMEIRO DISPOSITIVO CAPAZ DO RECONHECIMENTO DE COMANDOS DE VOZ. ....	12
FIGURA 2: HISTÓRIA CURTA DA REVOLUÇÃO DA VOZ. ....	13
FIGURA 3: IMAGEM ILUSTRATIVA DE UM DISPOSITIVO <i>ECHO DOT</i> (3ª GERAÇÃO). ....	18
FIGURA 4: IMAGEM ILUSTRATIVA DE UM DISPOSITIVO <i>ECHO DOT</i> (3ª GERAÇÃO) COM MICROFONE DESATIVADO. ....	19
FIGURA 5: PLACA COM O PROCESSADOR DO <i>ECHO DOT</i> . ....	19
FIGURA 6: PLACA COM OS MICROFONES DO <i>ECHO DOT</i> . ....	20
FIGURA 7: DETECÇÃO DA PALAVRA "ALEXA" NOS DISPOSITIVOS <i>ECHO</i> . ....	23
FIGURA 8: DIAGRAMA DOS SERVIÇOS QUE PROCESSAM OS COMANDOS NA ALEXA. ....	24
FIGURA 9: EXEMPLO DE SSML. ....	26

## SUMÁRIO

<b>INTRODUÇÃO</b> .....	<b>12</b>
<b>1 DISPOSITIVOS ECHO DOT</b> .....	<b>17</b>
1.1 HARDWARE DO ECHO DOT (3ª GERAÇÃO) .....	17
1.2 FUNÇÕES.....	20
1.3 NUVEM .....	20
<b>2 DETECÇÃO DOS COMANDOS DE VOZ</b> .....	<b>22</b>
2.1 PROCESSO DE DETECÇÃO.....	22
<b>3 PROCESSAMENTO DOS COMANDOS</b> .....	<b>24</b>
3.1 O SISTEMA ALEXA.....	24
3.2 ALEXA CLOUD .....	24
3.3 ASR – AUTOMATIC SPEECH RECOGNITION.....	25
3.4 NLU – NATURAL LANGUAGE UNDERSTANDING .....	25
3.5 SKILL.....	26
3.6 SKILL DE TERCEIROS .....	26
3.7 TTS – Text to Speech.....	27
3.8 ARMAZENAMENTO DE DADOS DA ALEXA .....	27
<b>CONCLUSÃO</b> .....	<b>28</b>
<b>REFERÊNCIAS</b> .....	<b>29</b>
<b>REFERÊNCIAS COMPLEMENTARES</b> .....	<b>31</b>

## INTRODUÇÃO

### CONTEXTUALIZAÇÃO HISTÓRICA

Atualmente, estamos passando por grandes avanços tecnológicos, tornando, dessa maneira, mais acessível a posse de dispositivos capazes de reconhecer comandos por voz. As assistentes virtuais, ou assistentes por voz, podem parecer uma tecnologia muito atual, porém elas estão presentes desde 1922, com o brinquedo chamado *Radio Rex*, que pode ser visto na FIGURA 1.

FIGURA 1: Radio Rex de 1920. Primeiro dispositivo capaz do reconhecimento de comandos de voz.



Fonte: Emre Sevinc – Flickr – 2005.

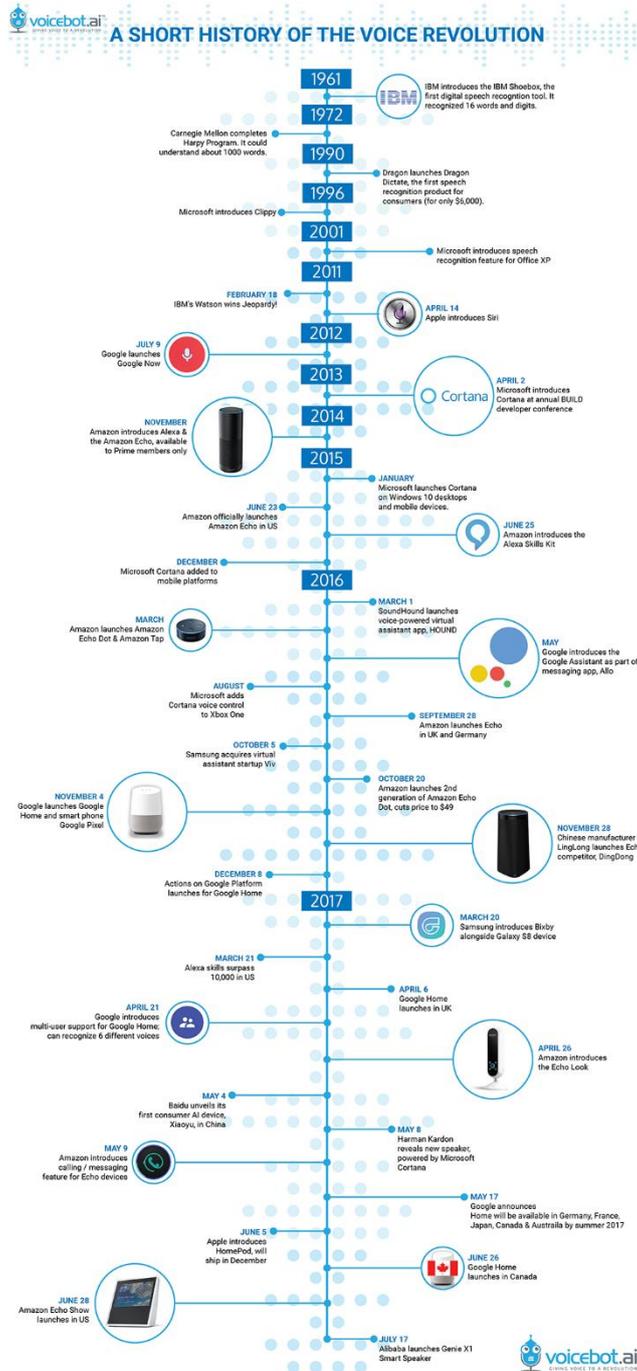
O brinquedo funcionava dessa maneira: o cachorro permanecia dentro da casa até ser chamado "Rex", então, ele saía da casa. O reconhecimento do comando pelo brinquedo é através da frequência da vogal ao falar o nome "Rex".

### PRIMEIROS ASSISTENTES POR RECONHECIMENTO DE VOZ

Como é possível visualizar na FIGURA 2, em 1962, a IBM anunciou a primeira assistente por reconhecimento de voz, chamada *IBM Shoebox*. O dispositivo era capaz de reconhecer números de zero a nove, e comandos matemáticos como soma, subtração, entre outras operações, semelhante à uma

calculadora. O reconhecimento das palavras era parecido com o mecanismo do brinquedo *Radio Rex*, no qual as palavras eram identificadas através da frequência do som.

FIGURA 2: História curta da revolução da voz.



Fonte: Voicebot.ai – 2017.

Por anos os dispositivos com reconhecimento de voz funcionaram da mesma maneira e com o mesmo propósito: auxiliar cálculos.

Avançando para o ano de 2008, temos o lançamento do primeiro dispositivo Android, a empresa *Google* começou a implementar pesquisas por voz, ocasionando, futuramente, na criação da *Google Voice Assistant*. Ainda, segundo a FIGURA 2, em 2010, temos o desenvolvimento da *Siri* pela *SRI International* como aplicativo para *iPhone*, com reconhecimento de voz, adquirida pela *Apple*, que fez a implementação nativamente do *software* em seus dispositivos, sendo a assistente padrão dos dispositivos da Apple a partir do iPhone 4s.

Em 2015, foi anunciado o lançamento da *Alexa*, assistente da *Amazon*. Seu nome é inspirado na Biblioteca de Alexandria, a consoante "x" no nome torna mais fácil o reconhecimento da voz quando é chamada, e, com o lançamento da *Alexa*, foi lançado os dispositivos Echo.

## **APRESENTAÇÃO DO PROBLEMA**

Este trabalho de conclusão de curso constitui-se em avaliar e analisar o tratamento dos dados enviados pela assistente *Alexa* presente nos dispositivos *Echo*, através dos comandos por voz, propondo responder a seguinte pergunta: qual a forma de envio e recebimento dos dados dos comandos?

## **OBJETIVO GERAL**

Este trabalho de conclusão de curso tem como objetivo geral o estudo do processo de recepção e envio de comandos por voz da *Alexa* nos dispositivos *Echo*.

## **OBJETIVOS ESPECÍFICOS**

Este trabalho de conclusão de curso tem como propósito estudar o tráfego dos comandos em dispositivos *Echo*, e, para atingir o objetivo geral, foram definidos três objetivos específicos:

1. Pesquisar quais são os serviços que processam os comandos até que chegue uma resposta ao usuário.
2. Verificar o tráfego dos dados dos comandos.
3. Analisar o processamento dos dados nos comandos enviados.

## JUSTIFICATIVA

A investigação proposta neste trabalho de conclusão de curso vem ao encontro da necessidade do estudo na segurança nessas novas tecnologias mais próximas ao usuário, onde podem existir dados sensíveis do usuário, como sua rotina, e, até mesmo, o que ele fala em momentos em que não estava enviando comandos.

Além disso, na literatura pesquisada, há poucos estudos sobre a assistente [Alexa](#), ainda em menor quantidade a disponibilidade em língua portuguesa, portanto este trabalho de conclusão de curso tem como propósito facilitar o acesso a informações envolvendo a [Alexa](#) em língua portuguesa.

## METODOLOGIA DE PESQUISA

A pesquisa científica apresenta várias modalidades, sendo uma delas a pesquisa bibliográfica que será abordada no presente Trabalho de Conclusão de Curso.

A pesquisa bibliográfica está inserida principalmente no meio acadêmico e tem a finalidade de aprimoramento e atualização do conhecimento, através de uma investigação científica de obras já publicadas.

Para Andrade (2010, p. 25): A pesquisa bibliográfica é habilidade fundamental nos cursos de graduação, uma vez que constitui o primeiro passo para todas as atividades acadêmicas. Uma pesquisa de laboratório ou de campo implica, necessariamente, a pesquisa bibliográfica preliminar. Seminários, painéis, debates, resumos críticos, monográficas não dispensam a pesquisa bibliográfica. Ela é obrigatória nas pesquisas exploratórias, na delimitação do tema de um trabalho ou pesquisa, no desenvolvimento do assunto, nas citações, na apresentação das conclusões. Portanto, se é verdade que nem todos os alunos realizarão pesquisas de laboratório ou de campo, não é menos verdadeiro que todos, sem exceção, para elaborar os diversos trabalhos solicitados, deverão empreender pesquisas bibliográficas (ANDRADE, 2010, p. 25).

## **ESTRUTURA DO TRABALHO**

O trabalho de conclusão de curso está estruturado da seguinte forma:

A introdução apresenta o problema, o objetivo geral, os objetivos específicos, a metodologia e a justificativa.

A revisão da literatura que representa o embasamento teórico-conceitual para o desenvolvimento da pesquisa, está dividida em três capítulos, além da conclusão:

Capítulo 1 – Dispositivos *Echo Dot*. Capítulo discorrendo sobre hardware, componentes e funções do dispositivo *Echo Dot* (3ª Geração).

Capítulo 2 – Detecção dos comandos de voz. Capítulo analisando o processo de detecção de voz até ativar a transmissão com a *Alexa Cloud*.

Capítulo 3 – Processamento dos comandos. Capítulo analisando os serviços em todas as etapas em que os dados dos comandos de voz passam, analisando a segurança dos dados e do usuário.

Conclusão.

## 1 DISPOSITIVOS ECHO DOT

Os dispositivos *Echo* são predominantes quando falamos em assistentes virtuais. De acordo com o blog de tecnologia estado-unidense *TheVerge*, a *Amazon* divulgou que em 2019 alcançaram a marca de 100 milhões de dispositivos vendidos, tornando-se algo comum na vida e na rotina de todos a cada dia.

O *Echo Dot* (3ª Geração), tem a seguinte descrição na página de venda na *Amazon* brasileira:

- O *Echo Dot* é o nosso *smart speaker* de maior sucesso. Controlado por voz com a *Alexa*, ele é perfeito para qualquer ambiente. Você pode pedir músicas, notícias, informações e muito mais. Além de ligar para amigos e familiares e controlar dispositivos compatíveis de casa inteligente com sua voz.
- Peça para a *Alexa* tocar músicas, responder perguntas, ler as notícias, checar a previsão do tempo, criar alarmes, controlar dispositivos de casa inteligente compatíveis e muito mais.
- Ouça músicas do *Amazon Music*, *Apple Music*, *Spotify*, *Deezer* e outros usando o alto-falante integrado com um som envolvente.
- Faça chamadas para o aplicativo *Alexa* ou outros dispositivos *Echo* compatíveis sem usar suas mãos. Envie avisos para outros dispositivos *Echo* em sua casa.
- A *Alexa* está sempre aprendendo e adicionando novas *skills*, como jogos, notícias e muito mais.
- Com 4 microfones de longo alcance, a *Alexa* ouve você do outro lado do cômodo. E com dispositivos *Echo* compatíveis em diferentes cômodos, você pode ouvir música em toda a sua casa.
- Também é possível conectar o *Echo* aos seus próprios alto-falantes por *Bluetooth* ou com um cabo de áudio de 3,5mm.
- Use sua voz para acender as luzes, trancar as portas e muito mais com dispositivos de casa inteligente compatíveis.

Por conta de tudo que o dispositivo tem a oferecer, não é à toa que vendeu tantas unidades. Um simples aparelho com 4,3 cm de altura x 9,9 cm de largura, não tem a capacidade de oferecer esses serviços e processar os dados diretamente nele. Por conta disso, é necessário que o aparelho sempre esteja alimentado com energia e conectado à uma rede *Wifi*, sendo assim, o dispositivo é responsável pelo recebimento dos comandos de voz, envio do comando para a *Alexa* (Inteligência Artificial da *Amazon*), recebimento da resposta e entrega dela para o usuário.

### 1.1 HARDWARE DO ECHO DOT (3ª GERAÇÃO)

Hardware é a parte física de um aparelho, como a carcaça, placa com os circuitos, processador, microfone e botões. O *Echo Dot* possui 4,3cm de altura e 9,9cm de largura; como pode ser visto na FIGURA 3, possui um led para sinalizar

quando for receber os comandos, e quando está enviando os dados para a [Alexa](#) na nuvem; 4 botões, sendo eles: ativar/desativar microfone, aumentar volume, diminuir volume e o botão de ação. O botão de controle do microfone ativa/desativa a funcionalidade do microfone e reconhecimento dos comandos por voz. O botão de ação serve para o caso de o usuário querer ativar sem chamar pela [Alexa](#) e ir direto para o comando.

FIGURA 3: Imagem ilustrativa de um dispositivo *Echo Dot* (3ª Geração).



Fonte: Página de venda da *Amazon* – 2021.

Os LEDs acendem em azul quando o dispositivo está ouvindo os comandos e enviando os dados para a [Alexa](#) processar, em vermelho quando está com o microfone desabilitado e em amarelo quando possui uma notificação, como exemplo ao pressionar o botão para desativar o microfone o led fica em vermelho assim como na FIGURA 4.

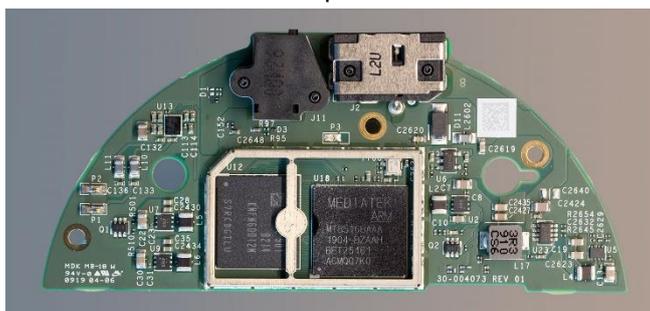
FIGURA 4: Imagem ilustrativa de um dispositivo Echo Dot (3ª Geração) com microfone desativado.



Fonte: Página de venda da Amazon – 2021.

Os principais componentes para o funcionamento dos comandos de voz são o microfone e o processador, eles são responsáveis pela detecção da palavra que ativa a [Alexa](#) para receber os comandos quando o usuário diz "Alexa", como pode ser visto, na FIGURA 5, a placa com o processador responsável pelo reconhecimento. Os processos da detecção de voz serão detalhados no Capítulo 2.

FIGURA 5: Placa com o processador do Echo Dot.

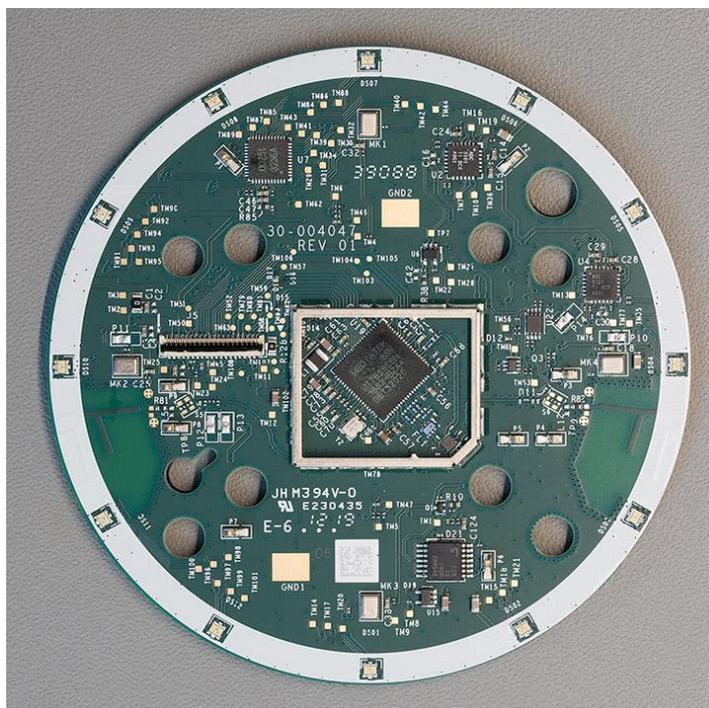


Fonte: briandorey.com – 2019.

O dispositivo tem 4 microfones no total, distribuídos pelos cantos da placa, a qual pode ser vista na FIGURA 6. O que facilita o processo de detecção, fazendo

com que a voz seja captada por todos os lados, o que resulta em não ser necessário falar próximo ao aparelho.

FIGURA 6: Placa com os microfones do Echo Dot



Fonte: briandorey.com – 2019.

## 1.2 FUNÇÕES

A construção pode parecer simples, porém é muito bem planejada e eficiente no papel que cumpre, já que a sua funcionalidade primária é receber os comandos e enviar à nuvem para que sejam processados pela [Alexa](#) e receber o retorno dos comandos enviados. O aparelho não grava o que é falado estando em espera até ser dita a palavra "Alexa", a partir do comando, como explicado anteriormente, o aparelho recebe os comandos e os envia para a nuvem. Por conta disso, deve-se manter o aparelho sempre alimentado com energia e conectado à uma rede *Wifi*.

## 1.3 NUVEM

O termo nuvem é muito comum atualmente para alguns casos envolvendo programação e tecnologia, as vezes ouve-se o termo *Cloud Computing* que é Computação na Nuvem, que é disponibilizar o que for necessário para o armazenamento, o processamento dos dados e a execução das tarefas através da internet, em servidores em várias regiões do mundo. Tornando possível aparelhos

como o *Echo Dot*, um aparelho com um hardware simples e que todo o processamento e execução das tarefas é na internet nos servidores da Amazon. Não sendo necessário um computador potente para executar os comandos de voz, e assim, enviando os comandos para a nuvem (servidores da Amazon) para processar.

## 2 DETECÇÃO DOS COMANDOS DE VOZ

Os dispositivos *Echo* foram projetados para serem entrada e saída da [Alexa](#), inteligência artificial da Amazon, para isso foi projetado com os microfones para a detecção dos comandos de voz dos usuários, assim como os outros assistentes virtuais, para ativar a [Alexa](#) é necessário dizer uma palavra específica, no caso da [Alexa](#) basta falar “Alexa” que já ativa a transmissão de áudio com a [Alexa](#) na nuvem e o envio dos comandos pelo usuário.

### 2.1 PROCESSO DE DETECÇÃO

Os dispositivos Echo enquanto não estão transmitindo o áudio para a [Alexa](#) na nuvem, detectam apenas a palavra que os ativa, o dispositivo possui uma memória pequena que aloca temporariamente o áudio ao seu redor e analisa através da frequência do som se a palavra “Alexa” foi dita. Está em um documento de privacidade da *Amazon* que não é gravado ou armazenado nenhum tipo de áudio fora as supostas palavras de ativação que ficam em uma memória temporária.

Enquanto não é detectado o processo fica em repetição até ser ativado e assim o Echo se prepara para receber o comando e enviar para a nuvem para a [Alexa](#) processar.

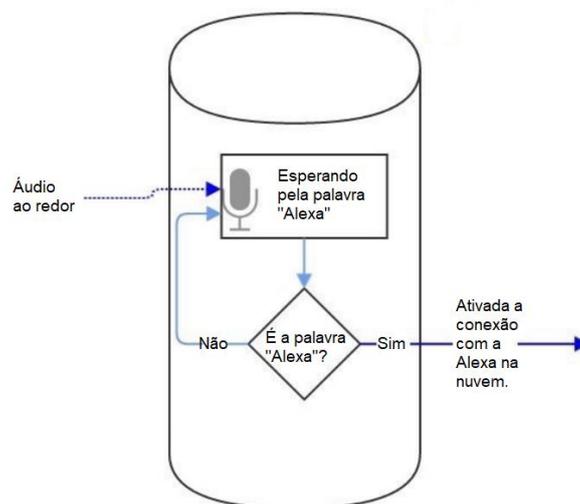
A palavra “Alexa” acaba facilitando a detecção por conta da consoante “x” que tem uma sonoridade muito específica, assim sendo mais eficiente na detecção pelo dispositivo.

De acordo com o documento de privacidade e manuseio de dados da [Alexa](#) disponibilizado pela *Amazon* o processo de detecção é dado em 6 etapas:

1. Ao menos que o microfone esteja desabilitado, o dispositivo Echo mantém gravando e analisando em espaços curtos na memória o áudio ao seu redor, até que seja dita a palavra que o ativa;
2. É detectada a palavra que o ativa;
3. É realizada uma transmissão de áudio com a [Alexa](#) na nuvem através da internet;
4. É acesa a luz azul no LED para demonstrar que está sendo feito a comunicação com a nuvem;
5. É analisado se o que está sendo falado é um comando para [Alexa](#), caso não seja é parado todo o processo e retorna para a etapa 1;
6. A transmissão do áudio é parada assim que a [Alexa](#) percebe que o usuário terminou de enviar o comando.

Os processos das etapas 1, 2 e 3 pode ser visto na FIGURA 7, onde o dispositivo *Echo* mantém verificando o áudio ao redor até ser detectada a palavra “Alexa” e assim iniciar uma transmissão de áudio com a [Alexa](#) na nuvem.

FIGURA 7: Detecção da palavra "Alexa" nos dispositivos Echo.



Fonte: Adaptado do diagrama em inglês no documento de privacidade da [Alexa](#) - 2021

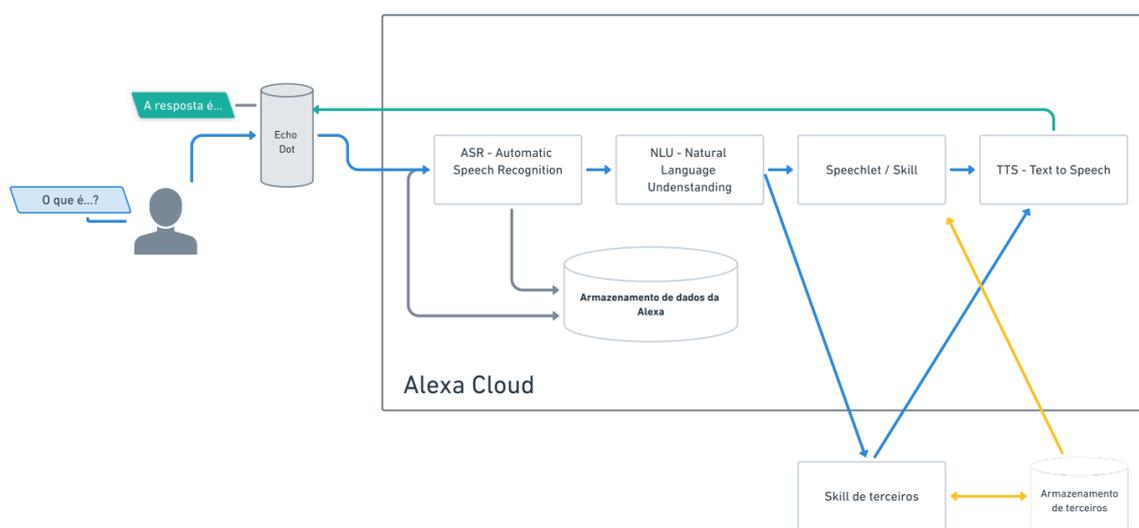
### 3 PROCESSAMENTO DOS COMANDOS

#### 3.1 O SISTEMA ALEXA

O sistema da [Alexa](#) é complexo e baseado na *Cloud Computing*, por conta disso é dividido em vários serviços e etapas desde a detecção até o retorno da resposta para o usuário, podendo comunicar com serviços externos ao da [Alexa](#) (serviços de terceiros), os quais também serão abordados.

Os serviços que processam o comando até chegar uma resposta final ao usuário pode ser acompanhada no diagrama contido na FIGURA 8.

FIGURA 8: Diagrama dos serviços que processam os comandos na [Alexa](#).



Fonte: Adaptado do diagrama em inglês no documento de privacidade da [Alexa](#) – 2021.

#### 3.2 ALEXA CLOUD

*Alexa Cloud* são os serviços da [Alexa](#) na nuvem, disponíveis através do *Cloud Computing* usando os servidores da *Amazon* para a execução das tarefas e armazenamento dos dados, a *Alexa Cloud* é composta por 4 serviços, **ASR – Automatic Speech Recognition** (Reconhecimento Automático de Fala), **NLU – Natural Language Understanding** (Compreensão de Linguagem Natural), **Skill** e **TTS – Text to Speech** (Texto para Fala), além disso ainda existe o **Armazenamento de dados da Alexa**.

A [Alexa](#) em algumas tarefas utiliza serviços externos, para isso existe no diagrama o **Skill de terceiros** e **Armazenamento de terceiros**.

### 3.3 ASR – AUTOMATIC SPEECH RECOGNITION

Reconhecimento automático da fala é uma tecnologia que converte palavras faladas em texto, é o primeiro tratamento que é feito no comando enviado pelo usuário.

A tecnologia **ASR** permite que computadores nos entendam através da fala, a transcrição da fala para texto ocorre quando enviamos um comando, um exemplo: o usuário perguntando “como está o clima em São Paulo?”, a inteligência artificial transcreve essa fala em texto de várias formas, e cada forma teria uma nota de confiança, como:

- “como está o clima em São Paulo”
- “como tá o clima em São Paulo”
- “como és tá o clima em São Paulo”
- “como está o clima em São Paolo”

É escolhido o texto que tiver a melhor nota, envia para o próximo serviço, o **NLU**, e caso o usuário esteja usando o aplicativo da **Alexa**, o texto aparecerá na tela.

É armazenada a detecção do que foi falado para o que foi transcrito no **Armazenamento de dados da Alexa** para *machine learning*, ou seja, para treinar e aprimorar a máquina para próximas transcrições.

### 3.4 NLU – NATURAL LANGUAGE UNDERSTANDING

O serviço **NLU** interpreta o texto enviado pelo **ASR**, a interpretação serve para descobrir o que o comando quer exatamente, e prepara para enviar para a **Skill**, nesse caso é enviado para a **Skill** ou **Skill de terceiros** que o usuário quer saber o **clima agora** em **São Paulo**.

Pode existir casos em que o usuário especifica o horário, supondo que enviou um comando seja “qual o clima em São Paulo às quatro da tarde”, nesse caso o que seria enviado para a **Skill** ou **Skill de terceiros** seria que o usuário quer saber o **clima às 16:00** em **São Paulo**.

Os dados da interpretação do texto são armazenados para *machine learning*, ou seja, para treinamento e aprimoramento da máquina para as próximas interpretações.

### 3.5 SKILL

Skill é onde é realizado o comando do usuário, no exemplo que abordamos, é pesquisado o **clima agora em São Paulo**, é retornado uma resposta para o serviço de Skill e então através dos dados recebidos pela pesquisa, o serviço de Skill formata uma resposta em SSML (*Simple Speech Markup Language*), uma linguagem de marcação simples para fala, que será enviada para o próximo serviço, **TTS**, preparando a forma da [Alexa](#) responder com voz, assim que é gerada a resposta é enviada para a próxima etapa.

#### 3.5.1 SSML

SSML é uma linguagem de marcação usada para casos onde queira que um texto seja lido por uma voz pelo computador, um exemplo de SSML no exemplo que estamos usando ficaria parecido com a FIGURA 9:

FIGURA 9: Exemplo de SSML.

```
< speak >  
  Em São Paulo, agora faz 30°C com previsão de chuva pela tarde  
< /speak >
```

Fonte: Exemplo desenvolvido por Lucas Amaral – 2021.

### 3.6 SKILL DE TERCEIROS

Em casos em que é solicitado para [Alexa](#) algo que não é da sua base de Skills e sim para algum serviço externo, como, anotar na agenda do *Google*, chamar um *Uber* e outros serviços externos. Nesse caso são acionadas as Skills de terceiros, mas para esses casos a [Alexa](#).

As skill tem como base a mesma função das Skills padrões da [Alexa](#), receber a entrada com o comando enviado pelo **NLU** e formatar a resposta para SSML para enviar para o **TTS**, porém no caso de Skills de terceiros a [Alexa](#) tem um tratamento diferente com os dados do usuário.

#### 3.6.1 MANUSEIO DOS DADOS

De acordo com o documento de privacidade e manuseio de dados da [Alexa](#) os dados pessoais do usuário não são enviados para os serviços de terceiros ao menos que o usuário permita através de uma solicitação que o serviço de terceiros faça.

### 3.6.2 **SEGURANÇA DO USUÁRIO**

Sempre que o usuário envia um comando para a [Alexa](#) e o comando faz o uso de uma Skill, é passado um token de usuário para a sua identificação dentro dessa Skill, porém o *token* de identificação é diferente para cada Skill, tornando segura a identificação do usuário.

### 3.7 **TTS – Text to Speech**

O serviço de TTS recebe o SSML gerado pela skill e cria um arquivo de áudio que é enviado para o dispositivo *Echo*, em alguns casos a interação termina nessa etapa, mas existem alguns comandos que são de pergunta e resposta, onde o usuário pode receber uma pergunta nessa etapa e responder e passar por todas as etapas novamente.

### 3.8 **ARMAZENAMENTO DE DADOS DA ALEXA**

Os comandos enviados são armazenados nos serviços de **ASR**, **NLU** e **TTS**; São armazenados para o treinamento da inteligência artificial através de *machine learning*, e para que os funcionários da *Amazon*, confirmem se está tudo correndo perfeitamente nos seus serviços.

Durante o serviço de **TTS** é armazenado o áudio gerado através do SSML, para a *Amazon* conferir se está gerando os áudios de fala apropriadamente.

## CONCLUSÃO

Este trabalho de conclusão de curso possibilitou entender o tráfego e processamento dos dados dos comandos enviados pelo usuário para a [Alexa](#) através dos dispositivos *Echo Dot*. Em termos de segurança os dispositivos *Echo* e a [Alexa](#) são seguros, com tokens de identificação do usuário no serviço de Skills, não gravar o que é falado no ambiente do dispositivo até o momento que a [Alexa](#) for ativada, e, quando ativada é feita uma transmissão direta com a *Alexa Cloud* via internet.

Para compreender os processos e o tráfego, definiu-se um objetivo. Pesquisar os serviços que processam os comandos enviados até que chegue uma resposta para o usuário, cujo objetivo demandou uma pesquisa bibliográfica sobre os documentos disponibilizados pela *Amazon* e outras referências.

Através dos documentos foi analisado o processo identificando possibilidades da falta de segurança do usuário, verificando se o áudio é gravado em algum momento quando o dispositivo *Echo* está ligado e se é armazenado algum dado desses comandos.

Como discorrido no capítulo do processamento dos comandos, sabe-se quais e quando são armazenados os dados pela *Amazon*, portanto em pesquisas futuras, pretendemos estudar a vulnerabilidade do dispositivo na rede local, antes mesmo que seja realizada uma transmissão com a *Alexa Cloud*, sendo um assunto mais técnico e exigindo mais do conhecimento sobre Segurança da Informação e Redes.

## REFERÊNCIAS BIBLIOGRÁFICAS

Alexa, Echo Devices, And Your Privacy. **Amazon**, 2021. Disponível em: <<https://www.amazon.com/gp/help/customer/display.html?nodeId=GVP69FUJ48X9DK8V>>. Acesso em: 25, novembro 2021.

BOHN, Dieter. Amazon Says 100 Million Alexa Devices Have Been Sold – What's Next?. **The Verge**, 2021. Disponível em: <<https://www.theverge.com/2019/1/4/18168565/amazon-alexa-devices-how-many-sold-number-100-million-dave-limp>>. Acesso em: 28, novembro 2021.

Echo Dot (3ª Geração): Smart Speaker com Alexa - Cor Preta. **Amazon**, 2021. Disponível em: <<https://www.amazon.com.br/Echo-Dot-3%C2%AA-Gera%C3%A7%C3%A3o-Cor-Preta/dp/B07PDHSJ1H/?th=1>>. Acesso em: 23, novembro 2021.

Echo Dot 3rd Gen Smart Speaker Teardown. **Briandorey**, 2021. Disponível em: <<https://www.briandorey.com/post/echo-dot-3rd-gen-smart-speaker-teardown>>. Acesso em: 19, novembro 2021.

Gonfalonieri, Alexandre. How Amazon Alexa works? Your guide to Natural Language Processing (AI). **towards data science**, 2021. Disponível em: <<https://towardsdatascience.com/how-amazon-alexa-works-your-guide-to-natural-language-processing-ai-7506004709d3>>. Acesso em: 23, novembro 2021.

MARR, Bernard. Machine Learning In Practice: How Does Amazon's Alexa Really Work?. **Forbes**, 2021. Disponível em <<https://www.forbes.com/sites/bernardmarr/2018/10/05/how-does-amazons-alexa-really-work/?sh=39c7d15b1937>>. Acesso em: 18, novembro 2021.

MUTCHLER, Ava. Voice Assistant Timeline: A Short History of the Voice Revolution. **voicebot.ai**, 2021. Disponível em:<<https://voicebot.ai/2017/07/14/timeline-voice-assistants-short-history-voice-revolution/>>. Acesso em: 23, novembro 2021.

RAMOS, Diana. Voice Assistants: How Artificial Intelligence Assistants Are Changing Our Lives Every Day. **Smartsheet**, 2021. Disponível em: <<https://www.smartsheet.com/voice-assistants-artificial-intelligence>>. Acesso em: 28, novembro 2021.

Speech Synthesis Markup Language (SSML) Reference. **Amazon Developer**, 2021. Disponível em: <<https://developer.amazon.com/en-US/docs/alexa/custom-skills/speech-synthesis-markup-language-ssml-reference.html>>. Acesso em: 28, novembro 2021.

What Do the Lights on Your Echo Device Mean?. **Amazon**, 2021. Disponível em: <<https://www.amazon.com/gp/help/customer/display.html?nodeId=GKLDRFT7FP4FZE56>>. Acesso em: 24, novembro 2021.

Whats Is Automatic Speech Recognition?. **Amazon Developer**, 2021. Disponível em: <<https://developer.amazon.com/en-US/alexa/alexa-skills-kit/asr>>. Acesso em: 24, novembro 2021.

White Paper – Alexa Privacy and Data Handling Overview. **D1 Awsstatic**, 2021. Disponível em: <<https://d1.awsstatic.com/product-marketing/A4B/White%20Paper%20-%20Alexa%20Privacy%20and%20Data%20Handling%20Overview.pdf>>. Acesso em: 24, novembro 2021.

## REFERÊNCIAS BIBLIOGRÁFICAS COMPLEMENTARES

GOGONI, Ronaldo. O que é a Alexa? [ou melhor, quem é], **tecnoblog**, 2019. Disponível em: <<https://tecnoblog.net/295738/o-que-e-a-alexa-ou-melhor-quem-e/>>. Acesso em: 29, novembro 2021.

KLEINA, Nilton. Quem é Alexa, a assistente virtual da Amazon. **Tecmundo**, 2021. Disponível em: <<https://www.tecmundo.com.br/produto/217404-alexa-assistente-virtual-amazon.htm>>. Acesso em: 26, novembro 2021.

What is a voice assistant?. **AlanBlog**. Disponível em: <<https://alan.app/blog/voiceassistant-2/>>. Acesso em: 27, novembro 2021.

When is Alexa Listening?. **Amazon**, 2021. Disponível em: <<https://www.amazon.com/is-alexa-always-listening/b?ie=UTF8&node=21137869011>>. Acesso em: 20, novembro 2021.