

MESTRADO EM CIÊNCIAS DA COMUNICAÇÃO  
ESPECIALIZAÇÃO EM ESTUDOS DE MÉDIA E JORNALISMO

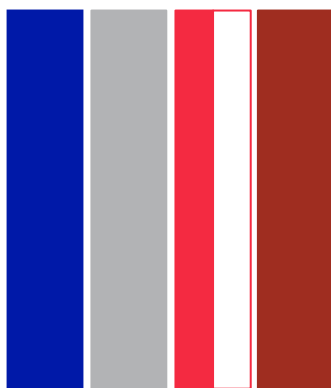
# **Comunicação humano-máquina**

## **Efeitos da humanização e feminilização de assistentes de voz inteligentes**

Priscila Maria Desideri de Oliveira

**M**

2021



Priscila Maria Desideri de Oliveira

# **Comunicação humano-máquina**

## **Efeitos da humanização e feminilização de assistentes de voz inteligentes**

Dissertação realizada no âmbito do Mestrado em Ciências da Comunicação, orientada pelo Professor Doutor Paulo e pela Professora Doutora Ítala Clay Freitas

Faculdade de Letras da Universidade do Porto

2021

Priscila Maria Desideri de Oliveira

# **Comunicação humano-máquina**

## **Efeitos da humanização e feminilização de assistentes de voz inteligentes**

Dissertação realizada no âmbito do Mestrado em Ciências da Comunicação, orientada pelo Professor Doutor Paulo e pela Professora Doutora Ítala Clay Freitas

### **Membros do Júri**

Professor Doutor (escreva o nome do/a Professor/a)

Faculdade (nome da faculdade) - Universidade (nome da universidade)

Professor Doutor (escreva o nome do/a Professor/a)

Faculdade (nome da faculdade) - Universidade (nome da universidade)

Professor Doutor (escreva o nome do/a Professor/a)

Faculdade (nome da faculdade) - Universidade (nome da universidade)

Classificação obtida: (escreva o valor) Valores



*Para Sol*

# Sumário

|                                                                        |    |
|------------------------------------------------------------------------|----|
| Declaração de honra .....                                              | 4  |
| Agradecimentos .....                                                   | 5  |
| Resumo.....                                                            | 6  |
| Abstract .....                                                         | 7  |
| Índice de Figuras .....                                                | 8  |
| Índice de Tabelas.....                                                 | 9  |
| Índice de Gráficos.....                                                | 10 |
| Introdução.....                                                        | 11 |
| Objetivo.....                                                          | 14 |
| Perguntas norteadoras.....                                             | 14 |
| Hipóteses.....                                                         | 14 |
| 1.Comunicação e o Paradigma Antropocêntrico .....                      | 15 |
| 1.1. Comunicação Humano-máquina .....                                  | 21 |
| 2.Inteligência Artificial e Comunicação .....                          | 31 |
| 2.1. Conceitos Fundamentais .....                                      | 35 |
| 2.1.1. Algoritmo .....                                                 | 36 |
| 2.1.2. Good Old Fashioned Artificial Intelligence .....                | 37 |
| 2.1.3. Machine Learning .....                                          | 38 |
| 3.Assistentes de voz inteligentes .....                                | 40 |
| 3.1. Spoken Dialogue Systems.....                                      | 40 |
| 3.2. Concepção e desenvolvimento .....                                 | 41 |
| 3.2.1. Siri, Cortana, Alexa e Google Assistente.....                   | 43 |
| 3.3. Funções.....                                                      | 47 |
| 3.4. Privacidade e Segurança.....                                      | 48 |
| 3.5. Usabilidade e Relevância .....                                    | 51 |
| 4.A voz da máquina .....                                               | 54 |
| 4.1. Reflexões sobre gênero .....                                      | 55 |
| 4.2. Gênero e personalidade da voz.....                                | 58 |
| 4.3. Emoção humana e emoção maquínica .....                            | 64 |
| 4.4. Crise de diversidade nos setores de inteligência artificial ..... | 70 |
| 5.Nota metodológica.....                                               | 74 |

|                                                    |     |
|----------------------------------------------------|-----|
| 5.1. Público-alvo e perfil dos participantes ..... | 75  |
| 5.2. Conceitualização de assistentes de voz.....   | 79  |
| 5.3. Preferências vocais .....                     | 83  |
| 5.4. Preferências de personalidade e gênero.....   | 88  |
| 5.5. Expectativas futuras .....                    | 92  |
| Conclusão .....                                    | 95  |
| Referências Bibliográficas .....                   | 103 |

## **Declaração de honra**

Declaro que o presente trabalho é de minha autoria e não foi utilizado previamente noutro curso ou unidade curricular, desta ou de outra instituição. As referências a outros autores (afirmações, ideias, pensamentos) respeitam escrupulosamente as regras da atribuição, e encontram-se devidamente indicadas no texto e nas referências bibliográficas, de acordo com as normas de referência. Tenho consciência de que a prática de plágio e auto-plágio constitui um ilícito académico.

Porto, 2021

Priscila Maria Desideri de Oliveira

## **Agradecimentos**

Agradeço primeiramente à Cecília, por ter me ajudado a continuar.

Aos meus pais e família, por serem minha rede de apoio e proporcionarem as condições afetivas e materiais que me permitiram realizar esta pesquisa.

Ao André, pela companhia e acolhimento na travessia. Por ler e reler meus textos e sempre contribuir com suas ideias e livros.

Aos professores Paulo Frias e Ítala Clay, pela orientação atenciosa em tempos de pandemia, pelas sugestões e correções que lapidaram o trabalho.

Aos amigos, em geral, pelas palavras de carinho e motivação. Em especial à Rafaela Pazos, pela troca e parceria de sempre. Ao Victor Matos, por ler e fazer considerações essenciais. Ao querido Frankner (Piro), pelas incontáveis conversas. Ao Lucas Azevedo, pela ajuda com os gráficos.

À Sol, por ser a maior e mais potente motivação de todas as minhas atuações.

## **Resumo**

Pesquisas antecedentes no âmbito da comunicação humano-máquina sugerem que seres humanos respondem socialmente a tecnologias comunicativas. Considerando a disseminação de inteligências artificiais interlocutoras, esta investigação busca compreender como assistentes de voz inteligentes são percebidos e conceituados por seus usuários, a fim de solidificar ou refutar as teorias prévias, em um contexto geográfico distinto. Com esta finalidade, aplicou-se um questionário online para coletar dados quantitativos acerca da percepção, preferências e expectativas dos brasileiros quanto a esta tecnologia. Os resultados revelam que os respondentes não humanizam assistentes de voz, todavia não há um consenso em sua conceituação; e o gênero da voz configura um traço determinante na interpretação desta tecnologia por seus utilizadores.

**Palavras-chave:** Comunicação humano-máquina; Assistentes de voz; Inteligência artificial; Tecnofeminismo.

## **Abstract**

Previous research in the field of human-machine communication suggests that people respond socially to communicative technologies. Considering the proliferation of interlocutor artificial intelligence, this study intends to understand how voice assistants are perceived and conceptualized by their users, to confirm or refute previous theories, in a distinct geographic context. For this purpose, an online questionnaire was applied to collect quantitative data about the perception, preferences, and expectations of Brazilians regarding this technology. The results reveal that respondents do not humanize voice assistants, however there is no consensus on their conceptualization; and the voice gender configures a determining feature in the interpretation of this technology by its users.

**Key-words:** Human-machine communication; Voice assistants; Artificial intelligence; Technofeminism.

## Índice de Figuras

|                                                                                        |                                      |           |
|----------------------------------------------------------------------------------------|--------------------------------------|-----------|
| FIGURA 1 - MODELO TRADICIONAL DE COMUNICAÇÃO .....                                     | <b>ERRO! INDICADOR NÃO DEFINIDO.</b> | <b>18</b> |
| FIGURA 2 - TESTE DE TURING (FASE 1) .....                                              |                                      | 33        |
| FIGURA 3 - TESTE DE TURING (FASE 2) .....                                              |                                      | 33        |
| FIGURA 4 - PROCESSO TÍPICO DE UM SISTEMA DE DIÁLOGO FALADO .....                       |                                      | 41        |
| FIGURA 5 - RESPOSTAS DE ASSISTENTES DE VOZ DIANTE DE ASSÉDIO E ABUSO VERBAL.....       |                                      | 62        |
| FIGURA 6 - ALTERAÇÕES NAS RESPOSTAS DA INTELIGÊNCIA ARTIFICIAL DO BANCO BRADESCO ..... |                                      | 64        |
| FIGURA 7 - CHATBOT DA MICROSOFT AFIRMA EM TWEET QUE HITLER NÃO FEZ NADA ERRADO .....   |                                      | 69        |

## Índice de Tabelas

|                                                                          |    |
|--------------------------------------------------------------------------|----|
| TABELA 1 – FAIXA ETÁRIA DOS PARTICIPANTES .....                          | 77 |
| TABELA 2 – GÊNERO DOS PARTICIPANTES .....                                | 77 |
| TABELA 3 – GRAU DE ESCOLARIDADE DOS PARTICIPANTES .....                  | 78 |
| TABELA 4 – USO DE ASSISTENTES DE VOZ .....                               | 79 |
| TABELA 5 – DESCRIÇÃO DE ASSISTENTES DE VOZ .....                         | 81 |
| TABELA 6 – IDEALIZAÇÃO DE ASSISTENTES DE VOZ .....                       | 82 |
| TABELA 7 – PREFERÊNCIA DE TOM DE VOZ .....                               | 85 |
| TABELA 8 – PREFERÊNCIA DE VOLUME DE VOZ .....                            | 87 |
| TABELA 9 – PREFERÊNCIA DE VELOCIDADE DE VOZ .....                        | 88 |
| TABELA 10 – IDEALIZAÇÃO DE GÊNERO DA VOZ .....                           | 91 |
| TABELA 11 – EXPECTATIVAS FUTURAS SOBRE O USO DE ASSISTENTES DE VOZ ..... | 94 |
| TABELA 12 – EXPECTATIVAS FUTURAS SOBRE ASSISTENTES DE VOZ .....          | 95 |

## Índice de Gráficos

|                                                                           |    |
|---------------------------------------------------------------------------|----|
| GRÁFICO 1 – FAIXA ETÁRIA DOS PARTICIPANTES .....                          | 77 |
| GRÁFICO 2 – GÊNERO DOS PARTICIPANTES .....                                | 78 |
| GRÁFICO 3 – GRAU DE ESCOLARIDADE DOS PARTICIPANTES .....                  | 79 |
| GRÁFICO 4 – USO DE ASSISTENTES DE VOZ .....                               | 80 |
| GRÁFICO 5 – DESCRIÇÃO DE ASSISTENTES DE VOZ .....                         | 81 |
| GRÁFICO 6 – IDEALIZAÇÃO DE ASSISTENTES DE VOZ .....                       | 83 |
| GRÁFICO 7 – PREFERÊNCIA DE TOM DE VOZ .....                               | 86 |
| GRÁFICO 8 – PREFERÊNCIA DE VOLUME DE VOZ .....                            | 87 |
| GRÁFICO 9 – PREFERÊNCIA DE VELOCIDADE DE VOZ .....                        | 88 |
| GRÁFICO 10 – PREFERÊNCIA DE PERSONALIDADE .....                           | 90 |
| GRÁFICO 11 – PREFERÊNCIA DE GÊNERO DA VOZ .....                           | 92 |
| GRÁFICO 12 – EXPECTATIVAS FUTURAS SOBRE O USO DE ASSISTENTES DE VOZ ..... | 94 |
| GRÁFICO 13 – EXPECTATIVAS FUTURAS SOBRE ASSISTENTES DE VOZ .....          | 95 |

## Introdução

A idealização desta dissertação se iniciou com elaborações reflexivas acerca das reverberações de novas tecnologias nos processos comunicacionais e das inevitáveis transformações na esfera da comunicação e suas teorias. A computação, a internet, os smartphones, as redes sociais, as inteligências artificiais: são inúmeras as descobertas tecnológicas que permeiam a nossa forma de comunicar, cada uma delas movimentando – de alguma forma e em alguma medida – novos desdobramentos para a troca de informação e o sentido que damos a essa permuta.

Dos filmes de ficção científica e da exploração de artigos e revistas especializadas, questões pertinentes em torno das relações humano-máquina, aparentemente desassistidas por grande parte da sociedade, emergiram e posteriormente encontraram seu alicerce em uma área de estudo já existente, a qual discute e pesquisa as temáticas que despretensiosamente configuraram objetos das reflexões iniciais que viriam a culminar nesta pesquisa.

Tal vertente, a Comunicação Humano-máquina, norteou os questionamentos concebidos da observação de tecnologias interativas, e proporcionou a fundamentação para que estas interrogações fossem estruturadas e desenvolvidas em uma nova investigação. Assim sendo, esta tese é fruto do entendimento que a comunicação entre seres humanos e máquinas afigura um processo comunicativo divergente daquele que outrora fora o foco das investigações acadêmicas desta área: a comunicação entre seres humanos.

A categorização da tecnologia exclusivamente como meio de comunicação promoveu a constituição de teorias da comunicação a partir de uma perspectiva antropocêntrica, negligenciando os processos nos quais estes instrumentos exercem função de comunicadores (Guzman, 2019). Ocorre que, com o progresso de inteligências artificiais e sistemas de automação, a análise desta outra ótica comunicativa – na qual a máquina não é apenas meio, mas interlocutora – faz-se urgente.

As teorias antropocêntricas da comunicação, para além do modelo comunicacional prático (Shannon; Weaver, 1948) – um emissor que envia uma mensagem através de

um meio para um receptor – consideraram necessário observar os significados atribuídos às mensagens pelos envolvidos nessa troca. O mesmo se manifesta na troca entre seres humanos e máquinas: os sentidos e significados gerados dessa interlocução precisam ser investigados tendo em vista as diversas combinações possíveis desse processo, nas quais pessoas e dispositivos ocupam as funções de emissores, receptores e meios (Banks; de Graaf, 2020).

Considerando as tecnologias comunicativas difundidas no mercado, sistemas que aprendem e recriam de forma autônoma destacam-se por fomentar múltiplas possibilidades de estudo. Portanto, a temática de inteligências artificiais e seus pontos de interseção com a comunicação (Gunkel, 2020) integrou a revisão bibliográfica aqui apresentada.

Além de examinar os conceitos de inteligência e artificialidade, bem como a origem interdisciplinar desta área de estudo, noções básicas de inteligência artificial – algoritmo e aprendizado de máquina – foram delineadas, visando a compreensão do funcionamento e da natureza destes sistemas. Assimilar a organização dessa tecnologia é um pré-requisito para refletir sobre as consequentes subjetividades, conceituações e interpretações que afloram da interação com esta.

Adiante, sob a justificativa de afigurar uma inteligência artificial comunicativa globalmente difundida e utilizada, capaz de promover as pretendidas reflexões acerca do processo comunicacional entre seres humanos e máquinas, assistentes de voz inteligentes constituíram o objeto de estudo trabalhado. Em um primeiro estágio, os elementos que compõem sistemas de diálogo falado foram elucidados, de modo a verificar as tecnologias empregadas em seu funcionamento, suas respectivas funções e um breve ensaio sobre seu desenvolvimento e evolução.

Vista a perspectiva técnica que, por sua vez, se faz indispensável para a contemplação filosófica, o enfoque se deslocou para a apresentação dos assistentes de voz mais utilizados, como Siri, Cortana, Alexa e Google Assistente, e subsequentemente os efeitos de seu uso. Foram abordadas questões como a usabilidade e relevância desses sistemas, receios relacionados a privacidade e segurança, e ainda sua aplicabilidade.

Ao final da seção teórica, revela-se o cerne do trabalho: observações acerca da potência da voz e seus desdobramentos na comunicação entre pessoas e seus assistentes inteligentes. As complexidades inerentes a esse elemento humano presente em máquinas interlocutoras foram discutidas ao passo em que suas pertinências se manifestaram no decorrer da investigação.

O gênero, a personalidade e a emoção são características que integram uma voz e, portanto, condicionam profundamente a percepção do receptor de mensagens emitidas por ela (Nass; Brave, 2005). Logo, as conclusões mais relevantes desta dissertação germinaram a partir da elaboração de cada um desses traços e suas ressonâncias.

Com o objetivo de analisar empiricamente a criação de sentido que se revela das relações humano-máquina, de modo a integrar todas as ideias aqui elencadas, a aplicação de um questionário online sinalizou a última fase desta investigação. Tal instrumento de pesquisa possibilitou a coleta de dados quantitativos a respeito da visão dos usuários sobre seus assistentes de voz, a fim de avaliar de que forma esta tecnologia é interpretada e como a utilização de um elemento humano – a voz – influencia esta significação.

Embasadas em teorias já fundamentadas, como a equação de mídias (Reeves; Nass, 1996), as hipóteses inicialmente ponderadas pretendiam avaliar se um processo de antropomorfização de máquinas interlocutoras se confirmaria nos resultados desta pesquisa. Dados os desdobramentos da revisão teórica, outras hipóteses se manifestaram e foram abarcadas.

A popularização de assistentes de voz cresce exponencialmente: estimativas sugerem que até 2024 o número de assistentes inteligentes em uso ao redor do mundo chegue a 8,4 bilhões, ultrapassando a população mundial (Statista, 2020). Compreender como estes sistemas inteligentes, portadores de voz e traços sociais são percebidos por aqueles que os utilizam significa antever parte da organização futura das relações sociais entre seres humanos e máquinas inteligentes que progressivamente atravessam a sociedade.

## **Objetivo**

Compreender como usuários de assistentes de voz inteligentes percebem e conceitualizam esta tecnologia, e verificar se a voz exerce alguma influência neste processo de significação.

## **Perguntas norteadoras**

Esta investigação baseia-se na premissa de que seres humanos respondem socialmente, de forma automática, a computadores e novas tecnologias, e quanto maior a humanização destas mídias, maior a mobilização de respostas sociais.

A fim de verificar a veracidade destes pressupostos nas relações entre brasileiros e seus assistentes pessoais, algumas perguntas serviram como guia:

- Quais teorias da comunicação fundamentam a relação entre estes pares (assistentes de voz x usuários)?
- Os usuários de assistentes inteligentes conceituam esta tecnologia a partir de uma perspectiva antropomórfica ou tecnológica?
- A voz e seus elementos inerentes – gênero, personalidade e emoção – interferem na percepção dos usuários? Se sim, como?

## **Hipóteses**

De modo a responder previamente estas perguntas, hipóteses foram construídas, sendo estas:

- 1) Usuários de assistentes de voz inteligentes humanizam esta tecnologia OU usuários de assistentes de voz inteligentes objetificam esta tecnologia.
- 2) A voz e seus traços influenciam diretamente as percepções dos usuários de assistentes inteligentes OU a voz e seus traços não interferem na percepção dos usuários acerca de assistentes inteligentes.

## 1. Comunicação e o Paradigma Antropocêntrico

Para iniciar uma abordagem reflexiva acerca de novas tecnologias a partir do viés comunicativo, é imprescindível, em um primeiro momento, considerar o conceito de comunicação a partir de sua extensão e mudanças históricas, bem como sua difusão popular, para posteriormente estruturar um conceito que se adeque aos propósitos desta dissertação.

Para tanto, observemos as definições encontradas em algumas fontes populares de informações. No dicionário, depara-se com o seguinte conceito de comunicação: “ato que envolve a transmissão e a recepção de mensagens entre o transmissor e o receptor, através da linguagem oral, escrita ou gestual, por meio de sistemas convencionados de signos e símbolos.”<sup>1</sup>

Na wikipedia, tem-se uma versão mais detalhada desta mesma ideia: “a comunicação (do latim *communicatio onis* que significa ‘ação de participar’) é um processo que envolve troca de informações entre dois ou mais interlocutores por meio de signos e regras semióticas mutuamente entendíveis. Trata-se de um processo social primário, que permite criar e interpretar mensagens que provocam uma resposta.”<sup>2</sup>

No site da associação americana de comunicação, influente desenvolvedora de estudos e pesquisas na área, há a seguinte definição: “em sua base, a comunicação se concentra em como as pessoas usam mensagens para gerar significados dentro e através de vários contextos, e é a disciplina que estuda todas as formas, modos, mídias e consequências da comunicação por meio de investigação humanística, científica social e estética.”<sup>3</sup>

Da perspectiva acadêmica, o conceito de comunicação abordado desde suas origens como objeto de teoria até o desenvolvimento desta desdobra-se em perspectivas diferentes, que relacionam-se intimamente ao contexto histórico. Segundo Rudiger,

---

<sup>1</sup> Significado de Comunicação pelo dicionário online de língua portuguesa Michaelis. Disponível em: <https://michaelis.uol.com.br/moderno-portugues/busca/portugues-brasileiro/comunicacao>

<sup>2</sup> Conceito de Comunicação pelo Wikipedia. Disponível em: <https://pt.wikipedia.org/wiki/Comunicacao>

<sup>3</sup> Conceito de Comunicação pela Associação Americana de Comunicação. Disponível em: <https://www.natcom.org/about-nca/what-communication>

“sempre que se fala em comunicação, é regra pensar o termo como dado, em vez de o tratarmos como problema: a reflexão sobre o assunto, não ele mesmo (...)” (Rüdiger et. al., 2014, p.40)

Em 1751, D’Alembert buscou conceituar o processo comunicativo na Enciclopédia:

“A ciência da comunicação das ideias não se limita a pô-las em ordem, deve também aprender como se expressa cada uma delas o mais claro possível e, assim, a aprimorar os signos destinados a lhes exprimir: é o que, pouco a pouco, os homens fazem.” (cit. por RÜDIGER et. al., D’Alembert, 1751, p.10)

Ainda de acordo com Rudiger, as primeiras conceituações e reflexões acadêmicas sobre a comunicação manifestaram-se no campo da sociologia. Para Albion Small (1894):

“As comunicações são, do ponto de vista do conteúdo, psíquicas; e físicas, do ponto de vista dos meios pelos quais os impulsos psíquicos são transmitidos. O conteúdo é espiritual; a forma é material. As pessoas são células de um organismo em que os meios servem para preservar a memória e superar as distâncias, sem alterar o princípio de transmissão dos impulsos psíquicos, que nascem com a palavra trocada de indivíduo para indivíduo.” (cit. por RÜDIGER et. al., 2014, p.43)

Também a partir de uma analogia orgânica do conceito de comunicação embora divergente da perspectiva anterior, Charles Cooley (1897) defende que:

“‘A sociedade é o resultado do encontro dos homens uns com os outros’ e que, na medida em que os encontros que formam a sociedade ‘resultam da comunicação entre os homens’, a conclusão que se impõe tirar é a de que ‘a história desta última é o fundamento de toda a história restante’”. (cit. por RÜDIGER et. al., 2014, p.44)

Ao abordar assuntos complexos, os conceitos fazem-se necessários, como forma de elaboração e reelaboração teórica de determinada disciplina que caminha em constante discussão (Barros, 2016). Portanto, as definições de comunicação proporcionam uma

oportunidade de observação a respeito do que se tem postulado quando investigamos esta ciência. Especificamente nestas ideias aqui elencadas, um ponto de convergência se evidencia: todas desenvolvem (direta ou indiretamente) o processo de comunicar a partir de um viés antropocêntrico; a troca de informação entre seres humanos.

Esta consonância releva a necessidade de recapitulação da história da comunicação, de forma que se possa verificar as circunstâncias que resultaram na conceituação da comunicação embasada no processo comunicativo humano.

Em geral, o estudo da história da comunicação segue uma ordem cronológica paralela à evolução da humanidade e, assim sendo, se inicia na pré-história quando os seres humanos começaram a se expressar e se comunicar com seus semelhantes por meio de pictogramas (desenhos e símbolos), posteriormente chamados de pinturas rupestres. Em decorrência, a associação de sons e gestos a determinado objeto ou ação, ou seja, a criação de signos e suas combinações entre si deram origem a linguagem (Perles, 2007).

Mais adiante, a possibilidade de representar graficamente os fonemas das palavras calharam no surgimento da primeira forma de escrita, nas tábuas cuneiformes da antiga Suméria (Gleick, 2013). Ainda por muito tempo, esta forma de comunicação era muito restrita e ainda inacessível para a maior parte da população, uma vez que a escrita era uma habilidade quase que exclusiva de pessoas letradas e monges (Perles, 2007).

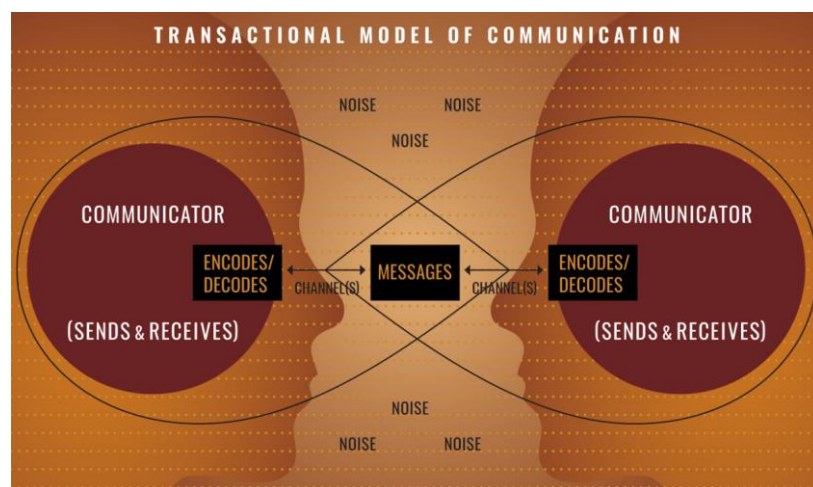
Esta realidade se alterou com a criação da prensa de Gutenberg. Considerada um dos maiores adventos do segundo milênio, a impressora mecânica permitiu a produção de livros em larga escala e, conseqüentemente, uma difusão mais democrática da escrita e do conhecimento, e ainda possibilitou a invenção do jornal (Eisenstein, 1998).

Já em um panorama mais recente, se considerarmos a história, a comunicação humana e seu estudo se expandiu consideravelmente com a emergência dos meios. Seguindo a ordem dos fatos, o surgimento do telégrafo, do rádio, da televisão, do computador e da internet modificaram de maneira irreversível a ordem mundial até então estabelecida. As profundas mudanças sociais ocasionadas pelos meios resultaram em uma crescente investigação acadêmica que veio a originar a teoria da comunicação (Serra, 2007).

Diante do percurso histórico da comunicação, é importante ressaltar que desde a pré-história até o surgimento da internet, os estudos desta disciplina e a própria teoria da comunicação fundamentam-se na comunicação humano-humano, reafirmando as definições inicialmente abordadas. Se no decorrer da história optamos por olhar para a comunicação com base na perspectiva antropocêntrica, é natural que atualmente a grande maioria de nossas fontes de informação idealizem o ato de comunicar por este ângulo.

*“Communication also is, and has been defined as a human-only process by those who study it. As the study of communication was coalescing in the mid-twentieth century, influential scholars purposefully drew its boundaries around human interaction. In textbooks, communication is presented within a primarily, if not exclusively, human context, with models, theories, and examples focused on people’s interactions.”<sup>4</sup> (Guzman, 2018, p.2)*

**Figura 1 – Modelo tradicional de comunicação**



Fonte: <https://www.natcom.org/about-nca/what-communication>

---

<sup>4</sup> Tradução livre: “A comunicação também é e tem sido definida como um processo exclusivamente humano por aqueles que a estudam. À medida que o estudo da comunicação estava se consolidando em meados do século XX, estudiosos influentes traçaram propositalmente seus limites em torno da interação humana. Nos livros didáticos, a comunicação é apresentada dentro de um contexto principalmente, senão exclusivamente, humano, com modelos, teorias e exemplos focados nas interações das pessoas.”

Andrea Guzman (2018) também levanta reflexões acerca desta imagem, retirada do site da associação americana de comunicação, que aparece em primeiro plano diante da pergunta “o que é comunicação?”. Dois aspectos da ilustração fomentam uma breve análise: a troca de informação acontece entre duas figuras humanas, reafirmando o embasamento antropocêntrico que margeia o estudo da comunicação; e o modelo de comunicação apresentado é aquele proposto por Claude Shannon em 1948.

Segundo Serra (2007), o artigo de Shannon “*A mathematical theory of communication*” constitui um elemento teórico base para a consolidação do “paradigma dominante” da teoria da comunicação. Como o nome sugere, o estudo aborda tecnicamente as questões da engenharia comunicacional, focando na transmissão de determinada mensagem, de um ponto a outro, excluindo os sentidos ou significados que a mensagem pode vir a configurar, nos contextos sociais e humanos, como o autor explica:

*“The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently the messages have meaning; that is they refer to or are correlated according to some system with certain physical or conceptual entities. These semantic aspects of communication are irrelevant to the engineering problem.”<sup>5</sup>*  
(Shannon, 1948, p.1)

Este modelo inicialmente proposto por Shannon visava otimizar a transmissão de sinais entre emissor e receptor, fossem estes humanos ou máquinas. No entanto, um ano mais tarde, Warren Weaver adaptou a teoria de Shannon de uma engenharia ao contexto de ciência social, reformulando-a como um meio de estudar os efeitos das mensagens. Dessa forma, os elementos do modelo que dentro do aspecto matemático poderiam ser

---

<sup>5</sup> Tradução livre: “O problema fundamental da comunicação é reproduzir em um ponto exatamente ou aproximadamente uma mensagem selecionada em outro ponto. Frequentemente, as mensagens têm significado; isto é, eles se referem ou estão correlacionados de acordo com algum sistema com certas entidades físicas ou conceituais. Esses aspectos semânticos da comunicação são irrelevantes para o problema de engenharia.”

ocupados tanto por uma pessoa quanto por uma máquina passaram a ser associados apenas com humanos (Guzman, 2018).

Por conta desta adaptação, o artigo foi republicado em forma de livro, e a teoria é geralmente atribuída a Shannon e Weaver (cit. por GUZMAN, Rogers, 1997). A partir da publicação, originou-se o esquema básico da comunicação que, por sua vez, referenciou diversas pesquisas e teorias subsequentes, e ainda é amplamente estudado e difundido até os dias de hoje.

Portanto, nota-se que a modelo utilizado como base para as diversas teorias da comunicação que se desenvolveram posteriormente estabelece os papéis atribuídos no processo comunicacional, onde seres humanos são os comunicadores (emissores e receptores) e as tecnologias são os meios de troca de mensagens (canais). “Desde então, este tem sido o paradigma dominante para a pesquisa em comunicação” (Guzman, 2018, p.8).

Para além do antropocentrismo comunicacional que embasa as teorias da comunicação, é necessário avaliar também o foco destes estudos na transmissão de informação, como o modelo de Shannon e Weaver e as próprias definições previamente citadas sugerem. O direcionamento das pesquisas para as questões pragmáticas da disciplina tende a negligenciar outros aspectos da comunicação ou, talvez, não os integrar adequadamente às questões tecnológicas.

O processo de comunicação é extremamente vasto e abrange não só a troca de informação, como também a criação de sentido que dela se manifesta. No contexto humano, a comunicação forma o entendimento entre o eu e o outro, contribui para a formação da sociedade, e é o meio pelo qual as pessoas se relacionam umas com as outras e com o mundo (Guzman, 2018).

Já o conceito de comunicação é intrinsecamente elástico e impermanente, e não poderia ser diferente. Se no século passado a concepção da teoria da comunicação desdobrou-se diante da necessidade de compreender e observar os efeitos dos meios de comunicação de massa, o século 21 emerge com a inevitável criação de novas teorias

e/ou adaptação das teorias já existentes, uma vez que novas tecnologias comunicativas surgem em constante exponencial.

### **1.1. Comunicação Humano-máquina**

A comunicação e o desenvolvimento tecnológico que a permeia configuram agentes transformadores da sociedade, com potencial revolucionário em contextos sociais, culturais, políticos e econômicos - movimento esse que se repete ao longo da história da comunicação - como é possível observar desde as reverberações sociais da prensa de Gutenberg até os efeitos decorrentes das tecnologias contemporâneas.

Ao idealizar uma aldeia global, McLuhan (1972) certamente previu a interatividade mundial que hoje se dá por conta da internet, mas é suposto que o comunicólogo não imaginava as especificidades tecnológicas que permeiam a vida pós-moderna, em especial, a comunicação. Hoje, nossos dispositivos digitais são máquinas inteligentes que, para além de nos conectar uns aos outros em escala global, aprendem sobre nossas preferências, como os algoritmos de nossas redes sociais; criam conteúdos originais, como softwares de geração de notícias; e até conversam conosco, como assistentes de voz inteligentes.

É incontestável também que tecnologias como estas estão em fase quase que embrionária quanto ao desenvolvimento, e tendem a se expandir e aprimorar no ritmo vertiginoso do século 21. Portanto, é fundamental que observemos os desdobramentos da evolução tecnológica no que tange a comunicação.

Na última década, novas tecnologias da comunicação que eram pensadas e imaginadas apenas nas telas de cinema, tornaram-se acessíveis à sociedade. Não é incomum nos depararmos com uma pessoa conversando *com* seu smartphone, ou lermos uma notícia escrita por um software, ou ainda recebermos inúmeros anúncios em nossas redes sociais depois de buscar algo relacionado em sites de pesquisa.

Nossas vidas já foram atravessadas por máquinas interlocutoras. Se antes o papel da tecnologia era meramente mediar a comunicação entre humanos, atualmente uma diversidade de dispositivos digitais ocupam o papel de comunicador no esquema básico da comunicação outrora proposto por Shannon e Weaver.

A premissa de McLuhan, de que o meio é a mensagem, nunca fez tanto sentido como faz agora diante de máquinas comunicadoras. Para o autor, mais importante que a usabilidade de um meio ou seu conteúdo é o próprio meio em si, “pois a ‘mensagem’ de qualquer meio ou tecnologia é a mudança de escala, cadência ou padrão que esse meio ou tecnologia introduz nas coisas humanas” (McLuhan, 2007, p. 22).

A introdução de dispositivos que interagem com seres humanos na sociedade manifesta a necessidade de observar não só suas funcionalidades e conveniências, mas a mensagem que as próprias máquinas concebem apenas por serem. Esta é a premissa básica da teoria dos meios, da qual McLuhan é um dos principais pensadores. No entanto, ao abordarmos as relações humano-máquina, ambas as partes demandam atenção. É preciso abordar adequadamente os papéis, objetivos, funções e identidades dos agentes comunicadores envolvidos no novo modelo comunicacional, sejam eles humanos ou máquinas.

“As evoluções tecnológicas e sociais provocaram mudanças operacionais, fenomenológicas e ontológicas nos processos de comunicação. Essas mudanças desencadeiam a necessidade de considerar os papéis humanos e da máquina nos processos de comunicação de uma forma mais igualitária.” (Banks; de Graaf, 2020, p.20)

A cenário antropocêntrico dos estudos em comunicação já não engloba todas as facetas do processo comunicativo, uma vez que existem diferenças implícitas na comunicação entre interlocutores humanos e na comunicação entre interlocutores digitais e humanos. *“When we say that machines talk back to us, we do not mean that this talk is identical to that of a human interlocutor, but the point is: what is the difference?”*<sup>6</sup> (Edwards; Fortunati, 2020, p.7)

Foi dito anteriormente que a comunicação consiste também na criação de significado proveniente da troca de informação, portanto, a comunicação entre seres humanos e

---

<sup>6</sup> Tradução livre: “Quando dizemos que as máquinas respondem a nós, não queremos dizer que essa conversa é idêntica à de um interlocutor humano, mas a questão é: qual é a diferença?”

máquinas reside na criação de sentido entre humanos e máquinas (Guzman, 2018). A urgência de evidenciar estas diferenças entre ambos os processos comunicativos, dadas suas especificidades, promoveu o surgimento e desenvolvimento de uma nova área de estudo: a Comunicação Humano-máquina (CHM).

Essa corrente de investigação foi influenciada por pesquisas dentro da comunicação e por diversas outras áreas de estudo, como as ciências da computação, engenharia, ciências da informação e inteligência artificial, por exemplo. De modo geral, cientistas e engenheiros da computação têm usado o conceito de comunicação como guia no desenvolvimento de interfaces tecnológicas a fim de facilitar seu uso, enquanto a pesquisa acerca da interação direta entre pessoas e máquinas tem sido negligenciada dentro da disciplina por não se enquadrar no paradigma antropocêntrico predominante da comunicação (Guzman, 2018).

A Comunicação Humano-máquina tem suas raízes atreladas ao estudo da Comunicação Mediada por Computador (CMC) que, por sua vez, se iniciou em meados do século 20 com a disseminação dos computadores na sociedade, e por definição *“is a process of human communication via computers, involving people, situated in particular contexts, engaging in processes to shape media for a variety of purposes”*<sup>7</sup> (Bubas, 2001, p.2).

De acordo com Rogers cit. in Guzman (2018), essa tecnologia, em específico, difere de suas antecessoras pela possibilidade de interação – as pessoas poderiam interagir umas com as outras através de seus computadores, e ainda interagir de forma mais direta com o dispositivo em si, diferentemente da televisão, por exemplo.

Embora as pesquisas deste campo tenham se estabelecido majoritariamente em torno da comunicação *através* do computador, alguns pesquisadores notaram a pertinência de investigar a relação das pessoas *com* esses dispositivos. Como citado por Guzman (2018), estudiosos da CMC observaram como as pessoas interagiam com programas específicos para acessar informações armazenadas dentro deles, como sistemas de

---

<sup>7</sup> Tradução livre: “é um processo de comunicação humana por meio de computadores, envolvendo pessoas, situadas em contextos específicos, engajadas em processos para moldar a mídia para uma variedade de propósitos.”

catálogo de cartões eletrônicos. Da mesma forma, a presença social é um fenômeno que, por definição, ocorre tanto na interação entre as pessoas, como na interação entre as pessoas e a tecnologia (Bubas, 2001).

Esse entendimento inicial foi fundamental para o desenvolvimento de pesquisas que analisam a interação entre seres humanos e máquinas, e se expandiu ainda mais na área de estudo que hoje conhecemos como Interação Humano-Computador (IHC). “A IHC é uma disciplina interessada no projeto, implementação e avaliação de sistemas computacionais interativos para uso humano, juntamente com os fenômenos relacionados a esse uso” (Barbosa; Silva, 2010, p. 10).

Partindo do pressuposto que a IHC se estrutura na usabilidade das tecnologias computacionais e não em sua construção, é crucial enfatizar que este campo, como praticamente todas as áreas de estudo que abordam comunicação e tecnologia, se configura em uma base multidisciplinar, portanto, pesquisas fora da comunicação são essenciais para o desenvolvimento dos campos da IHC e da CHM, e vice-versa.

Um artigo publicado por um dos estudiosos mais influentes da IHC, Clifford Nass, e seus colegas Jonathan Steuer e Ellen Tauber (1993), serviu de base para diversas pesquisas subsequentes no que diz respeito à interação entre pessoas e computadores. Intitulado “*Computers are social actors*”, o artigo defende o paradigma de que computadores são atores sociais, e por atores entende-se a configuração de um personagem ou papel que se associa à “um conjunto normativamente prescrito de comportamentos” (Nass; Steuer, 1993). Portanto, os autores argumentam que computadores preenchem papéis sociais.

Este artigo caracteriza um marco significativo na história da IHC e consequentemente da CHM, uma vez que sua proposição sugere uma descentralização da perspectiva antropocêntrica há tempos consolidada no âmbito da comunicação. Na concepção inicial do modelo de Shannon, acadêmicos da comunicação optaram por alocar a tecnologia somente na função de meio, e a teoria de CASA retorna com a tecnologia para a posição de interlocutora.

Outra publicação de Nass, dessa vez em conjunto com Byron Reeves, concebe outra perspectiva inovadora sobre como seres humanos percebem as tecnologias que os cercam. Os autores analisaram 35 experimentos nos quais a tecnologia desempenha os papéis de emissor e receptor no modelo de comunicação criado por Shannon, e a consistência das respostas de todas essas pesquisas resultou na teoria que nomeia o livro de Reeves e Nass: *"The Media Equation"*.

A teoria de equação das mídias sugere que novos mídias se equiparam à realidade. Em outras palavras, as pessoas tendem a tratar computadores e novos mídias como se fossem pessoas e espaços reais. Segundo os autores, essa conclusão é contraintuitiva, pois parece óbvio que dispositivos tecnológicos são ferramentas criadas no intuito de ajudar as pessoas a executar tarefas e ainda mais óbvio que as pessoas não têm relacionamentos com ferramentas (Reeves; Nass, 1998).

No entanto, os resultados das pesquisas realizadas indicam o contrário: *"people respond socially and naturally to media even though they believe it is not reasonable to do so, and even though they don't think that these responses characterize themselves"*<sup>8</sup> (Reeves; Nass, 1998, p.7). Essa resposta natural e social perante as mídias acontece de forma desatenta, uma vez que seres humanos têm uma predisposição primitiva às relações sociais e ambientes previsíveis. Isto é, se uma tecnologia se apresenta consoante às normas sociais e físicas com as quais estamos habituados, mais simples é seu uso e, por consequência, nossas respostas a ela são orgânicas.

Reeves e Nass (1998) alertam que, por esse motivo, o dilema em torno do uso de novos mídias se configuram para além da construção das tecnologias em si ou das indústrias que as criam, mas substancialmente em face à psicologia das pessoas que as utilizam. *"And it is from the perspective of psychology – the study of how people think and feel – that the media equation comes"*<sup>9</sup> (Reeves; Nass, 1998, p.9).

---

<sup>8</sup> Tradução livre: "as pessoas respondem socialmente e naturalmente à mídia, embora acreditem que não seja razoável fazer isso, e mesmo que não pensem que essas respostas os caracterizam."

<sup>9</sup> Tradução livre: "É da perspectiva da psicologia - o estudo de como as pessoas pensam e sentem - que surge a equação de mídias."

Acredito que a concepção das tecnologias de comunicação, suas reverberações sociais e as mensagens que manifestam são de extrema relevância para a compreensão do processo comunicativo como um todo, e devem ser exploradas nas disciplinas que integram seus estudos. Contudo, o foco desta dissertação descende da percepção humana diante das máquinas, posto que as ressonâncias sociais e psicológicas destas independem de suas origens, mas estão intimamente relacionadas à sua assimilação.

Nesse sentido, o historiador Yuval Noah Harari (2015) reflete sobre o teste de Turing – um estudo fundamental para a concepção da relação humano-máquina – e reitera a ideia de que, em muitos estágios, nossa percepção diante da tecnologia comunica mais do que a tecnologia propriamente dita:

“O Teste de Turing foi inventado em 1950 pelo matemático britânico Alan Turing, um dos pais da era da computação. Turing era homossexual em uma época na qual essa prática era ilegal na Grã-Bretanha. Em 1952 ele foi condenado por praticar atos homossexuais e obrigado a se submeter a castração química. Dois anos depois, cometeu suicídio. O Teste de Turing é uma replicação do teste mundano ao qual todo homossexual tinha de se submeter na Grã-Bretanha em 1950: você consegue se fazer passar por um heterossexual? Turing sabia por experiência pessoal que não importava o que você realmente é – a única coisa que importa é o que os outros pensam a seu respeito. De acordo com ele, no futuro os computadores seriam como os homossexuais na década de 1950. Não importa se os computadores efetivamente terão consciência, ou não. A única coisa que importa é o que as pessoas pensarão sobre isso” (Harari, 2015, p.127).

A questão de Turing desdobrava-se diante da consciência das máquinas, porém a hipótese também se adequa ao entendimento que fazemos de novas tecnologias: não importa o que elas são ou fazem, o que importa é o que as pessoas pensam sobre elas. Esse pensamento coletivo que se configura a partir do desenvolvimento tecnológico tende a reorganizar o corpo social por inteiro e, como consequência, tecer uma nova cultura que se estabelece na interseção entre o natural e o artificial.

De modo a compreender as reverberações culturais que se materializam diante do elo entre seres humanos e máquinas, é preciso novamente delimitar uma definição que fundamente essa suposição. Apesar de ser um conceito tão complexo quanto o de comunicação, no sentido literal da palavra, cultura tem como significado original o “cultivo agrícola”, ou seja, o cuidado com o que nasce naturalmente (Eagleton, 2000).

Entende-se então que é intrínseco ao conceito de cultura os aspectos naturais que, de alguma forma, conversam com a realidade humana. Em contrapartida, não é claro o papel do artificial. Assim, emerge o questionamento acerca das construções humanas – tudo aquilo que não nasce naturalmente – e suas contribuições para a formação cultural.

Nesse contexto, Eagleton (2000) sugere que a cultura é ao mesmo tempo uma ideia realista e construtivista, pois ao passo que implica a existência de uma natureza ou matéria para além da humana, é preciso que haja a partir dessa matéria-prima uma significação humana. Analisemos o uso de talheres como exemplo. Um talher por si só não influencia em nada a cultura, mas o significado e uso que atribuímos a esses instrumentos estão estreitamente ligados às noções culturais.

Essa perspectiva específica sobre a cultura se enquadra perfeitamente na compreensão da comunicação humano-máquina. Se a CHM consiste na criação de significado entre seres humanos e máquinas, e na formação da cultura se evidencia a necessidade de dar significado humano à matéria, devemos considerar a ideia de que o sentido que damos à nossa relação com máquinas está diretamente associado à formação de uma cultura contemporânea, na qual a vida humana é permeada por extensões de si mesma.

Tal suposição – de que nosso vínculo com máquinas reflete nas transformações culturais – enfatiza o valor substancial de entendermos como se configura essa relação, uma vez que a cultura de uma sociedade pode influenciar diversos aspectos da vida social e, de forma ainda mais profunda, nossa própria evolução biológica. Segundo o neurocientista Antônio Damásio, “há evidências crescentes de que, ao longo de muitas gerações, transformações culturais levam a mudanças no genoma” (Damásio, 2011, p.44).

A partir dessa suposição, é importante observar que a dialética entre orgânico e sintético que se manifesta das interações humano-máquina aflora não só reflexões no campo teórico e social, mas também biológico. Se almejamos uma compreensão profunda da CHM, é primordial que consideremos nossas predisposições congênitas, bem como a natureza artificial das máquinas.

Em sua obra “E o cérebro criou o homem”, Damásio (2011) faz uma analogia interessante entre organismos vivos e máquinas. Para ele, há coerência nas correlações que elaboramos ao compararmos seres vivos e seres maquínicos, devido aos processos estruturais que estes realizam. O processo de uma célula se equipara ao de um algoritmo, por exemplo; ambos têm um objetivo, e para alcançá-lo, vários elementos do organismo realizam tarefas secundárias.

“A semelhança é bem sugestiva e está por trás das metáforas de mão dupla com as quais descrevemos seres vivos e máquinas. Falamos do coração como uma bomba, descrevemos a circulação do sangue como um encanamento, referimo-nos à ação dos membros como alavancas, e assim por diante. Analogamente, quando pensamos em uma operação indispensável numa máquina complexa, nós a chamamos de ‘o coração’ da máquina, e aludimos aos mecanismos controladores dessa máquina como seu ‘cérebro’” (Damásio, 2011, p.64).

Por “metáforas de mão dupla”, o autor sugere que embora sejam ilustrações assertivas, é necessário estarmos atentos ao fato de que são metáforas meramente ilustrativas, e a realidade que se apresenta é oposta: máquinas e seres vivos são essencialmente diferentes quanto aos componentes materiais com os quais se configuram. Por esse motivo, apesar das semelhanças, existimos e evoluímos de formas distintas.

Considerando essas divergências, Reeves e Nass (1998) defendem que uma parte significativa da problemática em torno das nossas respostas às tecnologias estão relacionadas a questão evolutiva do cérebro humano. Como espécie, durante milhares de anos, desenvolvemo-nos em um cenário realista onde tudo o que aparentava ser

real, de fato era. Em termos didáticos, consideremos uma criança assistindo televisão: para ela, é fácil acreditar que tudo aquilo que aparece na tela é real.

O mesmo acontece com seres humanos adultos. Novas tecnologias que simulam comportamentos reais estimulam respostas sociais automáticas, ainda que estejamos cientes de que não estamos interagindo com outra pessoa. Isso acontece porque, no decorrer da evolução humana, a realidade física e social encorajava tais reflexos, que eram e são a base para a homeostase social.

De modo a concluir essa ideia, os autores afirmam que “*modern media now engage old brains*”<sup>10</sup> (Reeves; Nass, 1998, p.12). Diante desta alegação, reitera-se o alerta de Damásio (2011) para que consideremos as inconsonâncias físicas entre seres vivos e máquinas, e neste tópico específico, a evolução de cada espécie. Novas mídias e tecnologias de comunicação se remodelam e atualizam num espaço de tempo curtíssimo se comparados às mudanças estruturais da espécie humana, sejam elas socioculturais ou corporais.

A linha tênue entre as semelhanças e contradições de seres humanos e máquinas abarca muitos aspectos e, por consequência, muitas áreas de estudo. Face a tudo que fora exposto até então, nota-se um emaranhado de questões em áreas como a comunicação, ciências da computação, ciências sociais, psicologia, neurociência e biologia. Diante de uma base demasiadamente multidisciplinar, faz-se necessário reafirmar a pertinência de pesquisas em torno da relação humano-máquina dentro da comunicação.

O desenvolvimento estonteante de tecnologias comunicativas que ocorreu nos últimos 100 anos transformou por completo a sociedade que se configura hoje, o que nos leva a este momento ímpar da história da comunicação. Segundo Guzman (2018), nenhuma área da disciplina permaneceu inalterada perante a crescente natureza comunicativa das máquinas, uma vez que todos os aspectos da vida humana foram e estão sendo tocados pelos efeitos tecnológicos, direta ou indiretamente.

---

<sup>10</sup> Tradução livre: “a mídia moderna agora envolve cérebros antigos.”

A comunicação, assim como a cultura, esculpe um conceito elástico e desloca-se em constante mutação. Portanto, manutenções das teorias existentes e o surgimento de novas teorias naturalmente devem acontecer. A teoria da comunicação e o modelo comunicacional de Shannon e Weaver que a embasam já não se adequam à realidade das tecnologias comunicativas existentes hoje.

Avanços tecnológicos das interfaces que interagem por texto e voz, e o processamento de linguagem natural, por exemplo, concebem máquinas capazes de se comunicar de forma muito semelhante à interlocução humana, e para além da comunicação, estes dispositivos são personalizáveis; ou seja, interagem conosco conforme aprendem sobre nossas preferências.

*“The machine has become a communicative subject, and it is this subjectivity, rather than interactivity, that marks this technological transition, prompting scholars to ask what subjectivity means for individuals, society, and the study of communication”<sup>11</sup> (Guzman, 2018, p. 12).*

O estudo da Interação Humano-computador não abrange as especificidades das tecnologias contemporâneas, uma vez que a questão em torno das máquinas já não gira em torno de sua capacidade interativa, e sim da criação de sentido que dela resulta. Para tanto, a Comunicação Humano-máquina busca integrar as perspectivas teóricas e tecnológicas no contexto subjetivo que configura a relação entre seres humanos e máquinas.

---

<sup>11</sup> Tradução livre: “A máquina se tornou um sujeito comunicativo, e é essa subjetividade, ao invés da interatividade, que marca essa transição tecnológica, levando os estudiosos a perguntar o que a subjetividade significa para os indivíduos, a sociedade e o estudo da comunicação”.

## 2. Inteligência Artificial e Comunicação

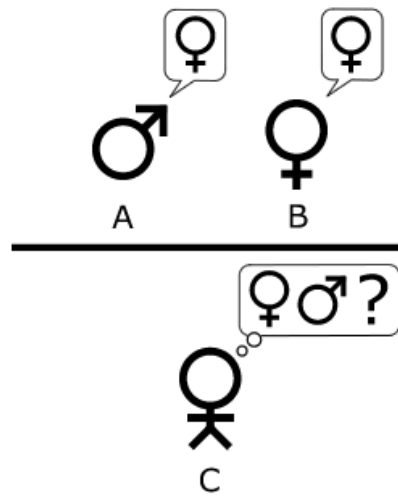
Uma grande quantidade de tecnologias da comunicação emergiu no decorrer do século XX, como a televisão e a internet. No entanto, foi a idealização e o desenvolvimento da inteligência artificial que possibilitou o surgimento de tecnologias mais sofisticadas, capazes de realizar tarefas autonomamente, aprender a partir daquilo que realiza, simular interações sociais por texto, voz e uma vasta gama de funcionalidades que podem ser atribuídas às máquinas inteligentes.

Ao considerar as áreas de estudo já mencionadas anteriormente, como a Interação Humano-computador e a Comunicação Humano-máquina, sabe-se que hoje há uma forte correlação entre as áreas da computação, da inteligência artificial e da comunicação, mas é pertinente nos perguntarmos quando de fato ocorreu o cruzamento desses campos de estudo.

Curiosamente, ao observar a história da inteligência artificial, a comunicação está presente e configura um elemento crucial desde sua concepção. Em 1950, Alan Turing publicou um artigo que viria a ser o marco inicial no percurso da inteligência artificial. Intitulado *“Computing Machinery and Intelligence”*, o artigo, como o nome sugere, tem o propósito de questionar e investigar a capacidade de inteligência de máquinas computacionais e, a fim de responder a esse questionamento, Turing propõe o que chama de “jogo da imitação”.

O experimento funciona da seguinte forma: na primeira fase, há três jogadores, um homem (A), uma mulher (B), e um interrogador (C), que pode ser de qualquer gênero. Os jogadores devem ficar em salas separadas e a comunicação entre eles acontece de modo indireto, através de um teleimpressor, por exemplo. O interrogador (C) faz perguntas aleatórias aos participantes, e com base nas respostas, tenta identificar se determinado jogador é um homem ou uma mulher.

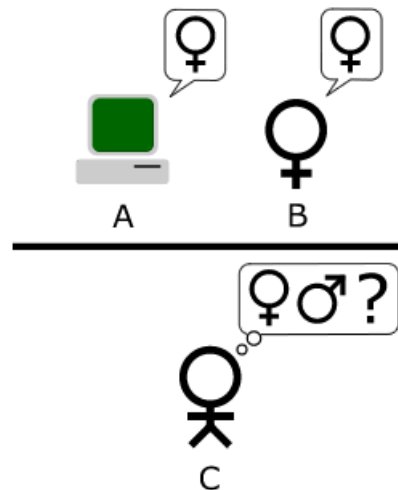
**Figura 2 – Teste de Turing (fase 1)**



Fonte: [https://pt.wikipedia.org/wiki/Ficheiro:The\\_Imitation\\_Game.png](https://pt.wikipedia.org/wiki/Ficheiro:The_Imitation_Game.png)

Na segunda fase, o jogador (A) é substituído por uma máquina, e se esta for capaz de responder às perguntas e simular um ser humano a ponto de o interrogador (C) não conseguir distinguir se está interagindo com um humano ou uma máquina, Turing defende que esta máquina deve ser considerada inteligente.

**Figura 3 – Teste de Turing (fase 2)**



Fonte: [https://pt.wikipedia.org/wiki/Ficheiro:Turing\\_Test\\_Version\\_1.png](https://pt.wikipedia.org/wiki/Ficheiro:Turing_Test_Version_1.png)

Se, a priori, Turing questiona a capacidade de pensar das máquinas, logo encontra uma barreira na resolução da questão:

*"I propose to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words. The new form of the problem can be described in terms of a game which we call the 'imitation game'"<sup>12</sup> (Turing, 1950, p. 37).*

Ao observar a ambiguidade das próprias palavras "máquina" e "pensar", Turing opta por reformular sua pergunta de investigação com a ideia do jogo da imitação. A falta de clareza no conceito de "inteligência", bem como o acesso direto à mente de algo ou alguém, configura a necessidade de outros meios de mensurar a inteligência.

Segundo Gunkel, acerca do artigo de Turing, a inteligência não é fácil de ser definida, *"It is, therefore, evidenced in and decided on the basis of external behaviors that are considered to be signs or symptoms of intelligence, especially communication and human-level verbal conversation"*<sup>13</sup> (Gunkel, 2020, p. 37).

---

<sup>12</sup> Tradução livre: "Proponho considerar a questão: "As máquinas podem pensar?". Isso deve começar com as definições do significado dos termos "máquina" e "pensar". As definições podem ser formuladas de modo a refletir, tanto quanto possível, o uso normal das palavras, mas essa atitude é perigosa. Se o significado das palavras "máquina" e "pensar" forem encontrados examinando-se como são comumente usadas, é difícil escapar da conclusão de que o significado e a resposta à pergunta "As máquinas podem pensar?" procurado em uma pesquisa estatística, como uma pesquisa Gallup. Mas isso é um absurdo. Em vez de tentar tal definição, substituirei a questão por outra, que está intimamente relacionada a ela e é expressa em palavras relativamente inequívocas. A nova forma do problema pode ser descrita em termos de um jogo que chamamos de 'jogo de imitação'."

<sup>13</sup> Tradução livre: "É, portanto, evidenciado e decidido com base em comportamentos externos considerados sinais ou sintomas de inteligência, especialmente comunicação e conversação verbal em nível humano."

Frente ao dilema para identificar a presença ou ausência de inteligência de uma entidade, Turing opta por utilizar a comunicação como indicativo de comportamento inteligente e, dessa forma, evidencia a importância da disciplina para o desenvolvimento computacional. Essa premissa se concretiza nos estudos subsequentes da área.

Embora a idealização de uma máquina inteligente já existisse para além da ficção quando Turing escreveu este artigo, foi apenas em 1956 que este campo de pesquisa foi nomeado de “inteligência artificial”. Durante a Conferência de Dartmouth, um grupo de pesquisadores apresentou um projeto que almejava verificar quais aspectos da inteligência poderiam ser descritos tão precisamente que uma máquina seria capaz de simulá-los:

*“We propose that a 2 month, 10 man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it”<sup>14</sup> (McCarthy, 1956).*

Em sequência, o projeto aponta sete aspectos da inteligência artificial que caracterizam os problemas de pesquisa a serem trabalhados, sendo aqui imprescindível ressaltar o segundo tópico sugerido: como um computador pode ser programado para usar uma linguagem.

*“It may be speculated that a large part of human thought consists of manipulating words according to rules of reasoning and rules of conjecture. From this point of view, forming a generalization consists of admitting a new word*

---

<sup>14</sup> Tradução livre: “Nós propomos que um estudo de inteligência artificial de 2 meses e 10 homens seja realizado durante o verão de 1956 no Dartmouth College em Hanover, New Hampshire. O estudo deve prosseguir com base na conjectura de que cada aspecto da aprendizagem ou qualquer outra característica da inteligência pode, em princípio, ser descrito com tanta precisão que uma máquina pode ser feita para simulá-lo.”

*and some rules whereby sentences containing it imply and are implied by others”<sup>15</sup> (McCarthy, 1956).*

Nesse sentido, a linguagem seria fundamental não só para o amadurecimento da inteligência artificial, mas da própria computação, já que o diálogo entre computador e ser humano é fundamental para a realização de determinada tarefa, como sugerem Clark e Licklider: *“On-line man-computer communication requires much development before men and computers can work together effectively in formulative thinking and intuitive problem solving”<sup>16</sup> (Clark; Licklider, 1962, p.1).*

Os artigos citados até então revelam momentos em que se substancializa a dependência comunicacional da inteligência artificial desde sua idealização até seu aprimoramento nas tecnologias atuais. A otimização de softwares que utilizam comunicação verbal depende diretamente de avanços no estudo do comportamento conversacional humano, por exemplo (Gunkel, 2020).

Assim sendo, o fortalecimento do elo entre estas disciplinas faz-se essencial tanto para a evolução tecnológica quanto para a reflexão e entendimento acerca do uso de tais tecnologias. Para tanto, é fundamental que a pesquisa em inteligência artificial conquiste seu espaço dentro da pesquisa em comunicação e vice-versa.

## **2.1. Conceitos Fundamentais**

Dentro de seu campo de estudo, a inteligência artificial é uma ideia múltipla; não há consenso em sua conceituação, pois a priori não há uniformidade sobre a própria concepção do que é inteligência e do que é artificial (Gunkel, 2020). Já em seu desenvolvimento prático, também não há um método único para compor uma

---

<sup>15</sup> Tradução livre: “Pode-se especular que grande parte do pensamento humano consiste em manipular palavras de acordo com regras de raciocínio e regras de conjectura. Deste ponto de vista, formar uma generalização consiste em admitir uma nova palavra e algumas regras em que as frases que a contêm implicam e são implicadas por outras.”

<sup>16</sup> Tradução livre: “A comunicação homem-computador on-line requer muito desenvolvimento antes que homens e computadores possam trabalhar juntos de forma eficaz no pensamento formativo e na resolução intuitiva de problemas.”

inteligência artificial. Essa organização descontínua das pesquisas na área tende a criar diversos subcampos de estudo e, por vezes, afigurar um cenário confuso do ramo.

Se na perspectiva dos estudiosos da área a inteligência artificial pode ser complexa, para o resto da sociedade o entendimento sobre esta tecnologia pode ser caótico, se considerarmos os reflexos das ficções científicas que permeiam o imaginário coletivo antes mesmo de a primeira inteligência artificial existir concretamente. Há muita dúvida e receio quando se fala em inteligência artificial, e ainda um deslocamento quanto ao que existe na vida real e o que existe apenas no cinema.

Por esses motivos, é importante a explicação de alguns conceitos básicos da IA, de modo a garantir a compreensão do propósito desta dissertação, uma vez que para questionar e refletir sobre o uso e a subjetividade de determinada tecnologia, é fundamental assimilar minimamente sua composição. Em busca deste entendimento, serão brevemente explicados o conceito de algoritmo e as duas formas mais usadas na construção de sistemas inteligentes; a manipulação de símbolos (ou GOFAL) e as redes neurais artificiais (ou *machine learning*).

### **2.1.1. Algoritmo**

“Um algoritmo é um conjunto metódico de passos que pode ser usado na realização de cálculos, na resolução de problemas e na tomada de decisões” (Harari, 2015, p.91). Por ser um conjunto metódico de passos, o conceito de algoritmo é trazido por muitos autores como analogia à uma receita culinária, no intuito de tornar mais simples o entendimento do conceito. Contudo, não é comum que o termo “algoritmo” esteja associado a assuntos gastronômicos, isto porque o habitual é nos depararmos com esta palavra no âmbito das ciências da computação.

De modo geral, o significado de algoritmo é sempre o mesmo, porém, no campo computacional, os elementos iniciais que compõem o conjunto de regras de um algoritmo são chamados de *input*, enquanto os elementos finais são chamados de *output*. “Any well-defined procedure that takes some value, or set of values, as input and

*produces some value, or set of values, as output. An algorithm is thus a sequence of steps that transform the input into the output*<sup>17</sup> (cit. por GUNKEL, Cormen et. al., 2009, p.5).

Um algoritmo por si só não é capaz de realizar uma tarefa, pois é preciso que haja algo ou alguém apto a ler suas instruções. Ou seja, na computação o algoritmo deve ser escrito em uma linguagem que determinado dispositivo computacional possa ler a fim de executar o conjunto de passos que levará ao *output* esperado. Harari (2015) exemplifica com uma máquina de café; o input é a moeda inserida, o output é o café pronto, o algoritmo descreve todo o processo para transformar a moeda em café, e a máquina é capaz de ler esse processo.

Entender o conceito de algoritmo é fundamental uma vez que toda inteligência artificial é algorítmica. Porém, vale ressaltar que nem todo algoritmo implementado em um dispositivo computacional é considerado uma inteligência artificial. Segundo Gunkel, essa diferença acontece porque *“AI is defined not in terms of some special technological features but by the perception of behaviors that are exhibited by the system”*<sup>18</sup> (Gunkel, 2020, p. 67).

### **2.1.2. Good Old Fashioned Artificial Intelligence**

A partir desta noção, podemos então nos perguntar como os algoritmos funcionam em uma inteligência artificial. De acordo com Gunkel (2020), há duas formas de construir uma IA, sendo a primeira por manipulação de símbolos, também chamada de GOFAI (*Good Old Fashioned Artificial Intelligence*), e a segunda por redes de neurônios artificiais, mais conhecida como *machine learning*.

A primeira baseia-se na premissa de que a inteligência humana consiste na manipulação de símbolos, e ainda que esta é uma condição necessária para inteligência. Ou seja, para construir uma máquina inteligente, ela necessariamente deve manipular símbolos. Sendo assim, o algoritmo de uma IA construída a partir desta lógica é escrito passo a

---

<sup>17</sup> Tradução livre: “Qualquer procedimento bem definido que leva algum valor, ou conjunto de valores, como entrada e produz algum valor, ou conjunto de valores, como saída. Um algoritmo é, portanto, uma sequência de etapas que transformam a entrada em saída.”

<sup>18</sup> Tradução livre: “IA é definida não em termos de algumas características tecnológicas especiais, mas pela percepção de comportamentos que são exibidos pelo sistema.”

passo, onde os símbolos apresentados como *input* serão processados e manipulados para gerar o *output*.

O fato ao qual devemos atentar sobre este tipo de IA é que as instruções codificadas no algoritmo são sistematicamente fornecidas por um programador. Isso quer dizer primeiramente que a inteligência em evidência é do programador, que precisa saber todo o procedimento necessário para chegar a um determinado resultado, e dessa forma, quanto maior a necessidade de exatidão nos resultados, maior a quantidade de instruções que deverão ser escritas, aumentando assim o trabalho do programador e ainda o tempo de processamento.

Apesar de suas limitações, a manipulação de símbolos funciona satisfatoriamente para execução de diversas tarefas e foi utilizada por muitos anos antes da disseminação das redes neurais artificiais, e por isso este tipo de sistema é chamado de GOFAL. Já a segunda forma de construir uma inteligência artificial é mais complexa e consequentemente mais utilizada nas tecnologias atuais, uma vez que garante possibilidades antes inexistentes através da manipulação de símbolos.

### **2.1.3. Machine Learning**

*Machine learning* ou aprendizado de máquina consiste na prática de utilizar algoritmos para coletar dados, aprender com estes dados e a partir deles gerar determinado *output*. Em outras palavras, o sistema é capaz de alterar o *output* de forma autônoma, com base em sua própria experiência.

Em termos técnicos, a prática do aprendizado de máquina se estrutura através de redes neurais artificiais criadas a partir da observação do comportamento de neurônios biológicos. Matemáticos e cientistas da computação se espelharam na atividade neuronal para a priori, criar um modelo matemático de neurônio, e a posteriori, identificar a interação dos neurônios entre si, a fim de reproduzir na máquina seu processo de conexão em rede e aprendizado (Gunkel, 2020).

Sendo a máquina apta a aprender com os dados coletados e se desenvolver autonomamente, a interferência humana é mínima em um sistema baseado em *machine learning*. Ao contrário de sistemas de manipulação de símbolos, em que é

preciso que o programador especifique detalhadamente as instruções para executar uma tarefa, nas redes neurais artificiais, a máquina é treinada a partir de uma grande quantidade de dados e algoritmos que lhe conferem a prática de aprender a realizar a tarefa.

Um exemplo simples são os algoritmos de sites de busca, como o Google, que a partir dos dados fornecidos pelos usuários, aprendem suas preferências e geram sugestões com base no que já foi pesquisado. Se o usuário pesquisar sobre acampamento, por exemplo, é provável que suas futuras pesquisas estejam relacionadas a essa temática. Embora seja um exemplo básico, ilustra adequadamente os fundamentos do aprendizado de máquina.

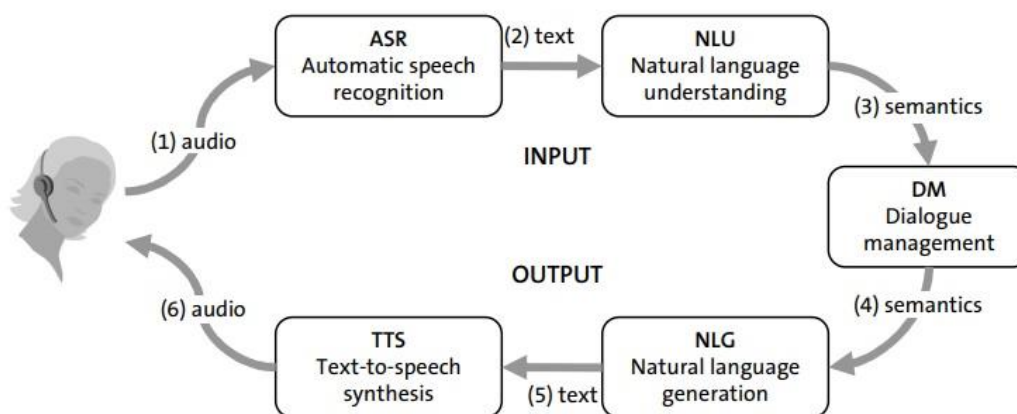
### 3. Assistentes de voz inteligentes

Abordar os conceitos trazidos no capítulo anterior faz-se imprescindível para entendermos o funcionamento da tecnologia que se configura como objeto de estudo desta dissertação: assistentes de voz inteligentes. Conhecidos por diversos nomes como assistentes pessoais virtuais (VPA's – *virtual personal assistants*), locutores inteligentes (*smart speakers*), assistentes digitais, interfaces conversacionais, *screenless internet*, entre outros, estes dispositivos são essencialmente sistemas computacionais capazes de se comunicar através da voz.

#### 3.1. Spoken Dialogue Systems

Esta tecnologia é chamada por seus desenvolvedores de *spoken dialogue system* (SDS) e consiste na interação entre diferentes inovações tecnológicas que possibilitam a interação por voz. Segundo Skantze (2007), o processo de sistemas SDS acontece em cinco estágios, e cada um representa um elemento tecnológico individual, sendo estes: 1) *Automatic speech-recognition*; 2) *Natural language understanding*; 3) *Dialogue management*; 4) *Natural language generation* e; 5) *Text-to-speech synthesis*.

Figura 4 – Processo típico de um sistema de diálogo falado



Fonte: Skantze (2007)

De modo a compreender o processo geral de um sistema de diálogo falado, breves definições sobre cada um dos estágios se fazem necessárias. Portanto, de acordo com Skantze (2007):

- 1) *Automatic Speech-Recognition* (ASR): O primeiro estágio é o reconhecimento automático de voz que basicamente tem por função decodificar um sinal acústico de fala em uma sequência de palavras escritas. Ou seja, o primeiro passo é transformar o áudio falado pelo usuário em texto.
- 2) *Natural Language Understanding* (NLU): Em seguida, o sistema de entendimento de linguagem natural utiliza o texto gerado pelo reconhecimento automático de voz e produz uma representação semântica do que foi dito. O algoritmo explora a estrutura sintática do texto, classificando o que é substantivo, adjetivo e verbo, por exemplo, e assim, o objetivo deste sistema é compor uma representação do que foi dito pelo usuário de forma que esta fala seja processada pelo computador.
- 3) *Dialogue Management* (DM): A representação semântica gerada pelo entendimento de linguagem natural não considera o contexto da fala do usuário, sendo esta a função do gerenciador de diálogo. Esta tecnologia, por sua vez, analisa a conjuntura do diálogo a fim de verificar a intenção contida no enunciado, considerando por exemplo a repetição de palavras ou a entonação do emissor.
- 4) *Natural Language Generation* (NLG): A partir do processamento daquilo que foi dito pelo usuário, o componente de geração de linguagem natural cria uma representação em forma de texto da resposta que será emitida, com base no leque de opções que o SDS em questão oferece.
- 5) *Text-to-speech Synthesis* (TTS): O último elemento, como o nome sugere, tem por função converter a representação textual gerada pelo sistema de geração de linguagem natural em uma forma audível. Geralmente esse processo ocorre a partir do uso de uma biblioteca de amostras de áudio pré-gravadas que são montadas peça por peça para constituir a mensagem que será falada pela máquina.

### **3.2. Concepção e desenvolvimento**

Como muito bem colocado por Hoy (2018), assistentes de voz de inteligentes caracterizam a realização do sonho da ficção científica de interagir com nossos

computadores falando com eles. Embora hoje esta seja uma realidade para muitas pessoas, as tecnologias que garantem o funcionamento deste tipo de inteligência artificial evoluíram gradativamente, passo a passo.

A estrutura de um sistema de diálogo falado revela a complexidade do processo interno de um assistente de voz, o que nos leva a refletir sobre a concepção dos elementos que o compõem. Gunkel (2020) explica que o entendimento de linguagem natural configura o elemento primordial para o avanço da computação no que diz respeito ao diálogo falado. Apesar de comumente associado a sistemas de voz, o NLU foi desenvolvido diante do surgimento de chatbots.

Simplificando, chatbots são softwares projetados para dialogar com seres humanos por texto, de forma autônoma e automática. Não coincidentemente, esta definição nos remete de volta ao jogo da imitação proposto por Turing, citado no capítulo 2. O que Turing idealizou em seu famoso artigo sobre máquinas pensantes, se concretizou em forma de chatbot por Joseph Weizenbaum, na década de 60.

Weizenbaum (1966) foi o responsável pela criação do primeiro chatbot, ainda que na época este termo ainda existisse. Chamado ELIZA, o objetivo do programa era, ironicamente, replicar falas de um psicoterapeuta, respondendo com perguntas vagas a partir das próprias palavras do usuário. O software utilizava processamento de linguagem natural e manipulação de símbolo para simular o entendimento do que foi dito pelo emissor, quando na verdade apenas selecionava palavras-chave e a partir destas estruturava perguntas pré-programadas, e caso nenhuma palavra-chave fosse identificada, ELIZA retornava com uma resposta genérica como “o que você sente sobre isso?”.

Esse experimento levanta duas questões pertinentes para a evolução de máquinas interlocutoras. A primeira é justamente o processamento de linguagem natural que, embora em estágio inicial, foi fundamental para que a máquina fosse “aprovada” no teste de Turing e até os dias atuais é utilizado nos sistemas conversacionais, como observamos na estrutura de SDS.

Já a segunda relaciona-se ao fato de que ELIZA convenceu muitas das pessoas com quem dialogou, ainda que estas pessoas soubessem que estavam conversando com um computador, a ponto de alguns usuários pedirem para passar horas no chat, e voltarem afirmando sentirem-se melhor, como se tivessem passado por uma experiência terapêutica genuína (Weizenbaum, 1966).

Esta última questão configura uma demonstração em retrospecto da teoria de equação das mídias, proposta por Reeves e Nass (1998), citada no capítulo 1. Também reafirmando a ideia de que devemos estar atentos ao nos relacionar com máquinas interlocutoras, uma vez que a tendência humana diante destes agentes é o processo de “antropomorfização”.

Depois de ELIZA, a criação e utilização de chatbots para diferentes fins tornou-se cada vez mais visada pelos estudiosos da computação, o que consequentemente resultou no aprimoramento da tecnologia, como aponta a evolução evidenciada nos sistemas de Parry, em 1972, Chatterbot, em 1994, A.L.I.C.E, em 1995, Cleverbot, em 1997 e Tay.ai, em 2016 (Gunkel, 2020).

Em paralelo, pesquisas em torno do reconhecimento automático de voz e do entendimento e geração de linguagem natural possibilitaram o surgimento dos primeiros softwares de diálogo falado, como o protótipo de smartphone IBM Simon, em 1994, e o programa de reconhecimento de voz *Dragon's Naturally Speaking*, em 1997 (Greenia, 1998). Embora os serviços mais populares durante esse período fossem as centrais de atendimento telefônico, conforme explicam Jokinen e McTear (2010).

### **3.2.1. Siri, Cortana, Alexa e Google Assistente**

Em um panorama mais recente, foi apenas em 2010 o lançamento da primeira assistente de voz inteligente cujo sistema é composto pelos cinco elementos propostos por Skantze (2007). Desenvolvida e aprimorada durante décadas de estudo em inteligência artificial pela *SRI International*, através do projeto CALO (Cognitive Assistant that Learns and Organizes) em parceria com o projeto PAL (Personalized Assistant that Learns) da DARPA, Siri estreou no mercado em 2010 como aplicativo, disponível para alguns

sistemas operacionais. Em 2011 foi comprada pela Apple, tornando-se um recurso exclusivo integrado inicialmente ao Iphone 4s.<sup>19</sup>

Siri é ativada com o comando “Hey Siri”, e a partir de sua ativação, o usuário pode solicitar um dos serviços oferecidos por ela, como tocar música ou acionar um despertador. Dessa forma, o primeiro sistema acionado é o reconhecimento automático de voz, e especificamente no caso de Siri, o sistema reconhece a voz do usuário, o que caracteriza uma etapa essencial de segurança a partir desta verificação.

Dado o comando desejado, Siri aciona seu sistema interno de entendimento de linguagem natural, a fim de decodificar a solicitação do usuário. Feita a análise do texto, o gerenciador de diálogo trabalha as nuances da fala, como a entonação, que por sua vez, ajuda a determinar se o usuário fez uma pergunta, que precisaria ser pesquisada em um site de busca, ou um comando direto, como fazer uma ligação.

Nesse sentido, assim como os outros assistentes de voz, Siri não necessariamente entende o que foi dito pelo usuário. O que acontece de fato é a identificação de uma das possibilidades de serviço oferecidas pelo assistente. Por esse motivo, caso o usuário faça alguma solicitação não disponível no sistema operacional do assistente, ele recebe uma resposta genérica pré-programada. No caso da Siri, a resposta costuma ser “eu não tenho uma resposta para isso. Algo mais em que eu possa ajudar?”.

Uma vez que Siri entende a requisição do usuário, é necessário identificar se o comando dado está relacionado a alguma informação contida no dispositivo, como por exemplo a lista de contatos que é necessária para enviar uma mensagem de texto, ou se a informação solicitada demanda uma pesquisa online, como no caso de o usuário perguntar qual é a capital da Austrália. Quando esta identificação for feita, seu sistema de geração de linguagem natural elabora uma resposta em forma de texto para este comando.

---

<sup>19</sup> Informações disponíveis no site da empresa SRI International: <https://www.sri.com/>.

Por fim, o texto gerado deve ser convertido em áudio pelo sistema de *text-to-speech*, que reside em uma biblioteca de amostras pré-gravadas por um ser humano. *“This approach to concatenation synthesis requires hours of recorded vocalizations from human voice actors that are then isolated, catalogued in a database of sound samples, and reassembled into understandable audio content on the fly”*<sup>20</sup> (Gunkel, 2020, p. 148).

Todas estas etapas seguidas por Siri são essencialmente as mesmas executadas pelos outros assistentes de voz disponíveis hoje no mercado. Embora tenham algumas funcionalidades exclusivas, a base do sistema operacional e os componentes dos assistentes de voz é a mesma. Sendo assim, Siri foi a primeira assistente de voz moderna e deu início a um movimento relevante no que tange sistemas de diálogo falado. Nos anos subsequentes, grandes empresas como Microsoft, Google e Amazon desenvolveram seus assistentes de voz inteligentes.

Em 2014, a Microsoft lançou sua assistente de voz Cortana, inspirada na personagem de inteligência artificial da série de jogos Halo. Como citado anteriormente, a estrutura de seu sistema operacional é a mesma utilizada pela Apple, e seu diferencial reside em sua função *“notebook”*. Este recurso é o controle de privacidade oferecido pela Microsoft, disponibilizando ao usuário todas as informações armazenadas pela Cortana, como interesses, comportamentos e localização. Dessa forma, o usuário tem maior controle sobre as configurações de privacidade, podendo por exemplo excluir informações indesejadas do notebook (Warren, 2014).

Ainda em 2014, a Amazon lança a Alexa, que além de integrar o dispositivo Amazon Echo Smart Speaker, está disponível em dispositivos de diversos fabricantes, uma vez que o serviço de voz oferecido pela Amazon é baseado em nuvem. Além desta distinção, a vantagem de Alexa frente aos demais assistentes de voz reside em seu enfoque na interação com outros dispositivos, visto que se observa uma crescente significativa no número de serviços que suportam comandos de voz, além de fabricantes de dispositivos

---

<sup>20</sup> Tradução livre: “Esta abordagem de síntese de concatenação requer horas de vocalizações gravadas de atores de voz humana que são então isoladas, catalogadas em um banco de dados de amostras de som e remontadas em conteúdo de áudio compreensível na hora.”

da “*Internet of things*” (IoT) estarem incorporando cada vez mais o controle por voz em seus produtos (Hoy, 2018).

Entre as grandes empresas, inusitadamente, o Google foi o último a lançar seu assistente inteligente, em 2016. Apresentaram ao público simultaneamente o dispositivo *smart speaker* Google Home e o assistente de voz baseado em nuvem Google Assistente. Inicialmente, o Google Assistente era restrito aos dispositivos e softwares Google, mas logo seguiu os passos da Amazon e abriu o sistema para todos os tipos de dispositivos com conexão à internet (Gunkel, 2020).

Os serviços oferecidos pelo Google Assistente, em geral, se equiparam ao desempenho dos outros assistentes de voz citados anteriormente. Ademais, o Google desenvolveu o Google Duplex, um sistema de diálogo falado que tem por objetivo central desenvolver as habilidades de conversação natural de seus softwares de comando de voz, com base em melhorias nos componentes de gerenciamento de diálogo (DM) e *text-to-speech* (TTS) (Leviathan; Matias, 2018).

Para além destas aqui citadas, há uma enorme variedade de assistentes virtuais no mercado, como por exemplo a Bixby da Samsung e a Julie da Amtrak. Se evidencia também uma tendência de crescimento na concepção de assistentes de voz por empresas prestadoras de serviço, que ao invés de terceirizar o serviço por comando de voz, optam por criar seus próprios sistemas, como a Bia do banco Bradesco.<sup>21</sup>

Conforme explica Hoy (2018), os assistentes modernos diferem de tecnologias anteriores como o *Dragon’s Naturally Speaking* pela quantidade de comandos e questões aos quais podem responder, devido ao fato de estarem sempre conectados à internet. “(...) *each interaction is sent back to a central computing system that analyzes the user’s voice commands and provides the assistant with the proper response. Earlier*

---

<sup>21</sup> Informações disponíveis no site do Banco Bradesco: <https://banco.bradesco/inovacao/bia.shtm>.

*voice-activated devices relied on a smaller set of ‘built-in’ commands and responses”<sup>22</sup>* (Hoy, 2018, p. 82).

### **3.3. Funções**

As funções disponíveis nos sistemas de assistentes de voz são essencialmente as mesmas, apesar das singularidades que cada software apresenta. De modo geral, as soluções por voz disponíveis são:

- Enviar e ler mensagens de textos e e-mails
- Fazer ligação para um contato da lista telefônica
- Configurar alarmes, temporizadores e eventos no calendário
- Ler resumos sobre as notícias do dia, de acordo com os veículos de comunicação desejados
- Acionar lembretes e fazer listas
- Informar as horas, a previsão do tempo ou do trânsito
- Fazer contas matemáticas básicas e conversões
- Pesquisar fatos básicos (como a área de um país ou sua capital)
- Se conectar com outros serviços online como tocar música no Spotify ou um filme no Netflix
- Contar piadas
- Se conectar e controlar dispositivos inteligentes, como lâmpadas e eletrodomésticos

Em adição a estes comandos já citados, uma função que tem se popularizado entre os usuários é a chamada “rotina”. É possível configurar um palavra-chave que acione uma série de comandos pré-programados. Por exemplo, o usuário pode falar “Alexa, bom dia” e com apenas esta frase, Alexa irá ligar as luzes, informar a previsão do tempo para o dia, ler um resumo de notícias e em seguida tocar música. Com a disseminação de

---

<sup>22</sup> Tradução livre: “(...) cada interação é enviada de volta para um sistema de computação central que analisa os comandos de voz do usuário e fornece ao assistente a resposta adequada. Dispositivos anteriores ativados por voz dependiam de um conjunto menor de comandos e respostas ‘embutidos’.”

dispositivos inteligentes da IoT, a expectativa é que a automatização destas “rotinas” cresça consideravelmente.

Neste sentido, emergem reflexões significativas quanto aos desdobramentos do processo comunicacional com assistentes de voz. A comunicação verbal ecossistêmica que emerge do diálogo falado entre seres humanos e assistentes de voz que a princípio era mais restrita à realização de tarefas pelo próprio dispositivo, agora tende a refletir no ambiente em que ocorre a conversa entre estes pares, uma vez que as projeções apontam para a hiper conectividade com outros dispositivos inteligentes. De que forma esta comunicação afetará e já está afetando o meio e como este meio - atravessado pelo diálogo entre vozes digitais e reais - afetará a vida humana?

### **3.4. Privacidade e Segurança**

Uma questão recorrente relacionada ao uso de assistentes de voz consiste na preocupação dos usuários acerca da privacidade de seus dados. Uma vez que a inteligência de assistentes de voz reside no serviço baseado em nuvens e não nos dispositivos, toda fala proferida pelo usuário é gravada, processada e armazenada no servidor (Gunkel, 2020). Essa coleta de dados pode ser analisada por duas perspectivas: a da empresa desenvolvedora e a do usuário.

A primeira perspectiva resume-se ao fato de que a utilidade de um assistente de voz para quem o cria reside na possibilidade de a empresa desenvolvedora revisar os dados de voz coletados, através de um software especializado ou revisores humanos, no intuito de aprimorar suas tecnologias de voz. Do ponto de vista do usuário, entretanto, a preocupação emerge da dúvida sobre o que acontece com seus dados de voz: para que são usados? Quem pode acessá-los? O assistente está sempre ouvindo?

Uma pesquisa realizada pela Microsoft em 2019 afirma que cerca de 41% dos usuários de assistentes de voz reportaram preocupações relacionadas a confiança, privacidade e escuta passiva. Esse consenso fundamenta-se em políticas de privacidade pouco transparentes com os usuários, além de diversas polêmicas acerca de vazamentos de dados coletados através de interfaces de voz.

Os termos de uso da Alexa, por exemplo, explicam que tudo o que é dito depois da palavra de ativação é gravado e armazenado em nuvem e que o usuário pode deletar esses dados entrando em contato com o suporte. Além de não ser um processo simples, não concede ao usuário nenhum tipo de controle sobre os dados que eventualmente podem ter sido extraídos dos áudios armazenados e os fins de sua consequente utilização.

A assistente da Amazon também lidera as polêmicas no que diz respeito a escuta passiva e uso indevido de dados, principalmente devido a quantidade de outros dispositivos com as quais ela se conecta e aos diversos desenvolvedores de suas múltiplas habilidades que também acessam os dados dos usuários (Benjamin, 2020).

A Siri da Apple, por sua vez, não armazena os áudios gravados pelo sistema, a empresa usa as transcrições deles, a fim de melhorar as questões de privacidade. De acordo com a publicação feita pela Apple em 2019, “quando armazenamos dados da Siri em nossos servidores, não os usamos para construir um perfil de marketing e nunca os vendemos para ninguém. Usamos os dados da Siri apenas para melhorar a Siri e estamos constantemente desenvolvendo tecnologias para torná-la ainda mais privado”.<sup>23</sup>

Já para o Google, conforme explicado em seu blog (Tasca, 2019), os áudios armazenados a partir das interações com o Google Assistente eram, a priori, parcialmente (cerca de 0,2% de todos os trechos) revisados por especialistas em idiomas, que transcreviam um conjunto das consultas ao assistente a fim de aprimorar o entendimento do idioma em questão. No entanto, em 2019, houve um vazamento de dados de áudio holandeses por parte de uma das empresas revisoras de idiomas contratadas pelo Google.

Frente a preocupação que se instalou entre os usuários diante da situação, o Google anunciou novas medidas de proteção de dados aos usuários do Google Assistente. Em nota, a empresa informou que os dados de áudio não serão armazenados sem permissão do usuário. Para manter os dados, o usuário deve acessar *Voice & Audio Activity* (VAA) e ativar esse recurso. Dessa forma, os arquivos serão enviados para a

---

<sup>23</sup> Informações disponíveis no site da empresa Apple:  
<https://www.apple.com/in/newsroom/2019/08/improving-siris-privacy-protections/>.

nuvem e usados pelo Google para melhorar o desempenho do assistente, e dentro de alguns meses, a grande maioria dos áudios serão automaticamente deletados (Tasca, 2019)

Apesar da implementação de medidas de protetivas, as polêmicas sobre quebras de privacidade no uso do Google Assistente continuam. Em maio de 2021, o Google foi acusado de gravar conversas de usuários e analisar as respectivas gravações no intuito de direcionar publicidade. A ação de iniciativa coletiva está sendo julgada pela Justiça da Califórnia, e embora o Google tenha solicitado o encerramento do processo, a juíza responsável negou o feito, dando continuidade ao julgamento.<sup>24</sup>

Em adição às problemáticas já citadas, questões de segurança também se manifestam diante do armazenamento de dados cada vez mais pessoais dos usuários. Assistentes inteligentes não coletam dados apenas através das interações por voz, pois estão integrados a outros aplicativos e dispositivos. Portanto, esses sistemas têm acesso ao que o usuário está procurando, ouvindo ou enviando em mensagens, além de diversas outras funções presentes nos dispositivos. Além disso:

“(...) qualquer pessoa com acesso a um dispositivo ativado por voz pode fazer perguntas, reunir informações sobre as contas e serviços associados ao dispositivo e solicitar que execute tarefas. Isso representa um grande risco de segurança porque esses dispositivos lerão o conteúdo do calendário, e-mails e outras informações altamente pessoais” (Hoy, 2018, p. 84).

Considerando os aspectos de privacidade e segurança aqui citados, governos e instituições públicas devem amparar possíveis falhas ou uso indevido de dados por parte das empresas responsáveis ou terceiros. No Brasil, a Lei Geral de Proteção de Dados (LGDP), de número 13.709, entrou em vigor em 2018, a qual “dispõe sobre o tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de

---

<sup>24</sup> Informações disponíveis no site New Voice: <https://newvoice.ai/2021/07/05/google-pode-enfrentar-acao-sobre-privacidade-no-uso-do-assistente-de-voz/>

liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural”<sup>25</sup>.

Suas disposições são amplas e é natural que o seja, uma vez que a necessidade de regulamentar a privacidade de dados é relativamente recente em todo o mundo, de forma que uma lei com previsões mais objetivas eventualmente tornaria a tutela da proteção de dados algo inviável, em um momento em que ainda se discute o alcance das questões relacionadas à privacidade individual e corporativa e às melhores formas de efetivamente protegê-las. Na Europa, por exemplo, medidas protetivas também foram estabelecidas juridicamente, por meio da lei General Data Protection Regulation (GDPR), que se assemelha à lei brasileira.

Por outro lado, faz-se essencial que empresas desenvolvedoras reforcem os processos de proteção de dados e reconhecimento de voz, além de fornecer aos possíveis interessados informações transparentes quanto ao recolhimento e uso dos dados coletados. Considerando que as leis amparam problemas já ocorridos ou em andamento, é necessário que os desenvolvedores busquem soluções que evitem ou minimizem os riscos quanto à privacidade e segurança de seus usuários.

### **3.5. Usabilidade e Relevância**

A popularização de assistentes de voz inteligentes simboliza um marco significativo para a interação e a comunicação humano-máquina. De acordo com a empresa Statista, especialista em dados de mercado e consumidores, o número de assistentes de voz inteligentes em uso - em uma escala global - chegaria a 4,2 bilhões até o final de 2020, e as estimativas sugerem que até 2024 este número chegue a 8,4 bilhões de unidades - número maior que a população mundial (levou cerca de 30 anos para que telefones celulares atingissem esse nível de ubiquidade).

Especificamente nos Estados Unidos, verificou-se um crescimento de 11,8% na compra de smart speakers entre 2018 e 2020, subindo de 22,9% para 34,7%. O aumento é

---

<sup>25</sup> Informações disponíveis na Constituição da República Federativa do Brasil, no site: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2018/lei/l13709.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm)

expressivo quando comparado ao número de pessoas que compraram smartphone, cujo crescimento foi de 4,6%, segundo dados divulgados pela New Voice<sup>26</sup>. A pesquisa também afirma que o número de usuários de assistentes de voz no país cresceu 11% entre os dois anos, caracterizando cerca de 128 milhões de usuários em 2020, enquanto a utilização diária da tecnologia teve um aumento de 23%. Além disso, a Siri lidera o mercado de assistentes de voz inteligentes, sendo a mais utilizada pelos norte-americanos.

No Brasil, há cerca de três anos, ainda não eram comercializados dispositivos cuja função principal fosse assistência por voz, como a Alexa da Amazon e o Nest do Google, que foram lançados no país apenas em 2019. Apesar de recente, a tecnologia tem ganhado espaço rapidamente. A pesquisa mais atual sobre o uso de assistentes de voz no país foi realizada em 2020 pela Ilumeo (Eishima, 2020), empresa de *Data Science*, a qual afirma que 57% dos usuários acessam seus assistentes em mais aparelhos que antes, enquanto 27% dos respondentes afirmaram ter interesse em adquirir um smart speaker.

A empresa ouviu 1.100 respondentes de faixas etárias e perfis de renda diferentes entre março e julho de 2020, e de acordo com os dados coletados, 92% dos entrevistados têm algum conhecimento sobre a tecnologia e 87% já utilizou ao menos uma vez. O dado mais surpreendente da pesquisa, no entanto, foi o aumento de 47% no uso de assistentes de voz inteligentes durante a pandemia de Covid-19. 54% das pessoas afirmaram identificar mais funcionalidades nestes dispositivos durante este período.

Em relação à frequência de utilização, 48% dos participantes afirmaram utilizar seus assistentes uma vez por semana, e um em cada cinco entrevistados utilizam todos os dias. Os principais usos de acordo com os respondentes são 1) a busca por voz (87%); 2) tirar dúvidas (82%); 3) tocar música (67%); 4) enviar mensagens (60%); 5) ouvir notícias (55%); e 6) transformar fala em texto (52%). O assistente de voz mais utilizado pelos brasileiros é o Google Assistant (72%).

---

<sup>26</sup> Informações disponíveis no site da New Voice: <https://newvoice.ai/2019/08/19/uso-do-assistente-de-voz-em-crescimento-nos-eua/>

A pesquisa também aponta fatores motivacionais para a utilização de assistentes de voz: 84% das pessoas afirmaram que o uso de comandos de voz é mais prático que a digitação, enquanto 76% acham vantajosa a possibilidade de realizar tarefas simultaneamente. Já entre os fatores desencorajadores estão 1) o preço dos aparelhos com a tecnologia (43%); 2) o não-entendimento dos comandos por parte do assistente (33%); e 3) receios quanto à privacidade dos dados (22%).

Além das funções mais populares entre os usuários, assistentes de voz inteligentes exercem funções essenciais no que diz respeito a acessibilidade.

“Estudos recentes vêm demonstrando o potencial do uso de assistentes de voz como tecnologia assistiva e ferramenta de acessibilidade. Entre os exemplos, há casos de uso por pessoas não alfabetizadas e com limitações — cognitivas, visuais, auditivas, de fala, ou motoras — que usufruem de uma variedade de funcionalidades adaptadas ao seu contexto”. (Bellei, 2021)

Os dados aqui expostos indicam que a forma como as pessoas interagem com a tecnologia está mudando velozmente e as interfaces de voz, especificamente assistentes inteligentes, estão progressivamente modificando as relações humano-máquina, uma vez que:

“A internet convencional construída em torno de uma interface gráfica e telas, cheia de páginas, links e textos e imagens visuais, está sendo rapidamente substituída por uma internet conversacional, mediada não por um navegador da web, mas por uma máquina que ouve e fala como uma pessoa” (West; Kraut; Chew, 2019, p. 94).

## 4. A voz da máquina

Sistemas de diálogo falado e suas multifuncionalidades abriram um leque de possibilidades na relação humana com a tecnologia. Em aspectos práticos, a opção de executar tarefas através da fala, sem a necessidade de ocupar os olhos, mãos e pernas proporciona facilidade e conforto aos usuários, inclusive em contextos ergonômicos.

Portanto, a comunicação entre seres humanos e máquinas pautada pelo diálogo falado é o que difere e torna única a experiência com esta tecnologia específica, mas para além de suas funcionalidades práticas, faz-se essencial analisar também suas características mais complexas, como a voz, a personalidade e as emoções atribuídas à máquina, pois estes traços estão diretamente relacionados à forma como nós humanos respondemos a estes dispositivos inteligentes.

Relembrando a teoria de equação de mídias de Reeves e Nass abordada no primeiro capítulo, o cérebro humano tende a dar respostas sociais automáticas diante de tecnologias que simulam comportamentos reais. Sendo a voz um elemento humano ímpar, há de se considerar a hipótese de tecnologias baseadas em voz alimentarem ainda mais a ilusão humana de que se está diante de outra pessoa ao interagir com máquinas interlocutoras.

Segundo Fortunati e Edwards, quanto mais elevado o nível de desenvolvimento técnico das máquinas e sua capacidade de simular características humanas, maior a mobilização de respostas sociais nas pessoas:

*“The strength of this illusion depends on the simulating ability of the media agents. Although it is an illusion, and is even consciously recognized as such by the people involved, it can generate all the same feelings of communicative and social satisfaction as interactions with other humans”<sup>27</sup>* (Fortunati; Edwards, 2020, p. 9).

---

<sup>27</sup> Tradução livre: “A força dessa ilusão depende da capacidade de simulação dos agentes de mídia. Embora seja uma ilusão, e até mesmo seja conscientemente reconhecido como tal pelas pessoas envolvidas, pode gerar todos os mesmos sentimentos de satisfação comunicativa e social que as interações com outros humanos.”

#### 4.1. Reflexões sobre gênero

Nass e Brave (2005) explicam que, por razões biológicas, psicológicas e culturais, a definição de nós mesmos começa com gênero, sendo esta uma distinção muito importante devido ao princípio de identificação social. Há evidências em diversas áreas de estudo, como a sociologia, a antropologia e psicologia de que quanto mais similares duas pessoas são, mais positivamente dispostas elas estarão uma em relação à outra. Esse processo de atração entre similares começa pelo gênero.

Nesse sentido, a voz caracteriza um fator relevante na identificação de gênero, sendo possível fazer esta distinção em questão de segundos através da fala, uma vez que *“the assignment of gender is based on a number of voice characteristics, all of which are analyzed in language-related areas of the brain”*<sup>28</sup> (Nass; Brave, 2005, p.10). Portanto, a voz e o gênero que esta interpreta tendem a ser a primeira característica que notamos ao dialogar não só com outra pessoa, mas também com dispositivos que falam.

É imprescindível então pontuar as possíveis projeções que se atribuem aos gêneros. Os autores alertam que para além das distinções entre macho e fêmea, existem os conceitos de masculino e feminino que ultrapassam as dissemelhanças biológicas; traços de personalidade e comportamentos associados a homens e mulheres de acordo com fatores socioculturais produzem crenças generalizadas chamadas estereótipos de gênero.

Se o gênero configura uma das primeiras características definidas a partir da voz, e perante a sociedade cada gênero carrega um conjunto de normas implícitas que se considera ser o padrão, uma máquina que possui voz e gênero pode criar expectativas de gênero com base em estereótipos.

*“The more feminine-sounding the voice is (regardless of whether it is male or a female voice), the more female gender stereotypes will be attributed to and expected from it, and conversely, the more*

---

<sup>28</sup> Tradução livre: “A atribuição de gênero é baseada em uma série de características da voz, todas as quais são analisadas em áreas do cérebro relacionadas à linguagem.”

*masculine-sounding the voice is, the more stereotypes will be attributed to and expected from it*<sup>29</sup> (Nass; Brave, 2005, p.16).

Aqui desenvolvem-se alguns questionamentos sobre a voz da máquina: o que o gênero da voz revela sobre esta tecnologia? Como esta distinção de gênero é percebida pelos usuários? Estereótipos de gênero são reproduzidos nas relações humano-máquina baseadas em voz? O princípio da similaridade se reafirma entre mulheres e máquinas com vozes femininas e homens e máquinas com vozes masculinas?

A fim de responder a estas perguntas, Nass e Brave aplicaram uma série de experimentos para medir as respostas humanas frente aos gêneros da voz da máquina. Os resultados confirmam a teoria de equação de mídias: tratamos máquinas como se fossem pessoas reais, e de forma ainda mais acentuada, reproduzimos comportamentos sociais relacionados a gênero quando interagimos com máquinas interlocutoras, como a atração entre similares e confirmação de estereótipos.

Três estudos foram realizados para identificar estes comportamentos, nos quais homens e mulheres por volta de 40 anos dialogavam com computadores que possuíam vozes com traços masculinos e/ou femininos. Respectivamente, nas três experiências relatadas, concluiu-se que: 1) a atribuição automática de gênero às vozes é um fenômeno que se estende das relações humanas para o mundo tecnológico; 2) estereótipos de gênero são reproduzidos nas relações humano-máquina de acordo com o gênero da voz; 3) mulheres se identificam e confiam mais em máquinas com vozes femininas e homens se identificam e confiam mais em máquinas com vozes masculinas.

*“It is impossible to live in or even contemplate the social world without gender. These biological, social, and cultural constructs are automatically manifested by voices, both real and artificial, and automatically recognized by listeners’ brains. A careful leveraging of gender identification can make users feel positive and supported, while ignorance of gender identification can unintentionally drive*

---

<sup>29</sup> Tradução livre: “Quanto mais feminina soa a voz (independentemente de ser masculina ou feminina), mais estereótipos de gênero feminino serão atribuídos e esperados dela e, inversamente, quanto mais masculina soa a voz, mais estereótipos serão atribuídos e esperados dele.”

*users away. No matter how unnatural or unclear the voice, voice gender is at the heart of interfaces that speak*<sup>30</sup> (Nass; Brave, 2005, p. 17).

Face ao exposto, é incontestável a seriedade da escolha do gênero conferido à voz de determinado dispositivo, e diante do panorama contemporâneo, novos questionamentos se manifestam. A conclusão da última pesquisa realizada por Nass e Brave aponta para uma identificação entre seres humanos e máquinas baseada em gênero, o que nos levaria a idealizar máquinas personalizáveis que disponibilizassem vozes masculinas e femininas no intuito de atender às preferências do usuário.

Aparentemente as empresas que se destacam na produção de assistentes de voz (Apple, Google, Amazon, Microsoft) optaram por ignorar estudos como este e criaram sistemas de diálogo falado majoritariamente femininos (Siri, Google Assistente, Alexa, Cortana). Se não a premissa da similaridade, qual a motivação por trás desta escolha consonante? As empresas alegam que optaram por vozes femininas com base em pesquisas realizadas, as quais mostraram melhor recepção e simpatia dos usuários por vozes femininas (Schwär; Moynihan, 2020).

Todavia, além do princípio da similaridade, outros estudos não reforçam a suposta preferência popular por vozes femininas. Uma pesquisa aponta que a maioria das pessoas prefere o tom baixo das vozes masculinas, e outra afirma que as pessoas geralmente preferem a voz do sexo oposto (West; Kraut; Chew, 2019). De acordo com as pesquisas de Nass e Brave, a voz feminina é percebida como útil e cooperativa enquanto a masculina é associada à autoridade (Nass; Brave, 2005).

A preferência por vozes femininas aparenta associar-se mais ao consenso social (estereótipo) de que mulheres são mais “prestativas” do que propriamente às características da voz feminina, como timbre e volume. Segundo a publicação Unesco

---

<sup>30</sup> Tradução livre: “É impossível viver ou mesmo contemplar o mundo social sem gênero. Essas construções biológicas, sociais e culturais são automaticamente manifestadas por vozes, tanto reais quanto artificiais, e automaticamente reconhecidas pelos cérebros dos ouvintes. Uma utilização cuidadosa da identificação de gênero pode fazer com que os usuários se sintam positivos e apoiados, enquanto a ignorância da identificação de gênero pode afastar os usuários involuntariamente. Não importa o quão não-natural ou confusa seja a voz, o gênero da voz está no centro das interfaces que falam.”

(West; Kraut; Chew, 2019), esta associação também é reforçada pelos papéis interpretados por homens e mulheres na cultura popular, uma vez que pesquisas relatam que tanto nos videogames como em shows de televisão, os personagens femininos são comumente coadjuvantes que assistem um personagem central masculino.

#### **4.2. Gênero e personalidade da voz**

Além do indicativo de gênero, uma voz e suas características fornecem informações suficientes para que nossos cérebros automaticamente associem traços de personalidade a este som. Simpatia, antipatia, extroversão e introversão, por exemplo, são atributos rapidamente identificados pela voz e constituem a primeira impressão do ouvinte sobre o interlocutor. As consequências deste primeiro contato são expressivas, já que tendemos a prever um leque de comportamentos com base no que julgamos ser a personalidade do emissor, e consequentemente decidimos como agir e o que esperar deste.

De acordo com Nass e Brave (2005), quatro aspectos fundamentais da voz apontam traços de personalidade, e são estes: o volume, o tom, o alcance vocal e a velocidade da fala. Cada uma destas características indica algo sobre o interlocutor, no entanto, a combinação dos quatro aspectos torna o processo de identificação de personalidade ainda mais significativo. Alguém que fala em tom agudo, com muita variação entre altos e baixos, rápido e alto é imediatamente apontado como extrovertido, por exemplo.

O mesmo processo que utilizamos para julgar a personalidade de uma voz humana se reproduz diante de uma voz sintética. *“Essentially, although people know that synthetic systems do not really have personalities, the evolution-derived tendencies to assume that voices are social and respond accordingly may overwhelm that knowledge”*<sup>31</sup> (Nass; Brave, 2005, p. 36).

---

<sup>31</sup> Tradução livre: “Essencialmente, embora as pessoas saibam que os sistemas sintéticos não têm realmente personalidades, as tendências derivadas da evolução de assumir que as vozes são sociais e responder de acordo podem sobrecarregar esse conhecimento.”

No caso de assistentes de voz inteligentes, as características de voz se estabelecem na gravação e edição de amostras de áudio que posteriormente serão utilizadas nos diálogos com seus usuários. A considerar os quatro assistentes de voz abordados nesta dissertação e o tempo que estão no mercado, é possível ponderar que estes já tenham personalidades bem definidas pelos usuários.

A relevância da definição da personalidade da voz destes dispositivos interlocutores reside justamente em sua influência nas decisões, comportamentos e sentimentos do usuário em relação ao sistema. Nass e Brave (2005) fizeram um experimento com 36 pessoas previamente descritas como extrovertidas e outras 36 pessoas introvertidas, a fim de analisar se a personalidade de uma voz extrovertida/introvertida influi na percepção e conduta dos usuários em conformidade com suas próprias personalidades.

O resultado do estudo, de modo geral, confirma que seres humanos classificam inclusive as vozes sintéticas de acordo com suas próprias personalidades, e ainda constata a premissa da similaridade: extrovertidos preferiam ouvir a voz extrovertida e introvertidos preferiam ouvir a voz introvertida. Nesse sentido, os usuários que utilizaram o sistema cuja personalidade da voz combinava com a do próprio usuário foram fortemente influenciados pela voz, comprovando que a personalidade do emissor interfere diretamente na percepção do receptor acerca da mensagem.

*“Voices are not merely a handy means to transmit information to the user. All voices – natural, recorded, or synthetic – activate automatic judgments about personality. These judgments influence people’s expectations about the system and its behavior and determine whether they will like the system, trust it, and be willing to buy from it”<sup>32</sup> (Nass; Brave, 2005, p.45).*

Equitativamente, o conteúdo da mensagem transmitido pela voz também concede informações relevantes ao receptor. Segundo os autores, é impossível pensar na voz e

---

<sup>32</sup> Tradução livre: “As vozes não são apenas um meio prático de transmitir informações ao usuário. Todas as vozes - naturais, gravadas ou sintéticas - ativam julgamentos automáticos sobre a personalidade. Esses julgamentos influenciam as expectativas das pessoas sobre o sistema e seu comportamento e determinam se elas vão gostar do sistema, confiar nele e estar dispostas a comprar dele.”

nas palavras proferidas por ela separadamente, já que ambas estão intimamente ligadas desde a primeira palavra falada. Diversos estudos apontam que o cérebro humano busca consistência na interação entre estes fatores, ou seja, é fundamental que o conteúdo daquilo que é dito seja coerente com a voz que o fala.

Em suma, as recomendações de Nass e Brave no que diz respeito às características de personalidade percebidas através da voz e do conteúdo preferidos por máquinas interlocutoras é que se considere 1) a premissa da similaridade, que sugere que os usuários preferem vozes cujas personalidades combinem com as suas próprias, e 2) a consistência entre a voz e o conteúdo.

A confirmação repetitiva do princípio da similaridade em muitas vertentes das relações humano-máquina valida também a analogia ao mito grego de Narciso proposta por McLuhan, na qual o autor sugere uma perspectiva alternativa sobre o sentimento do herói: o elemento fundamental da paixão é na verdade uma extensão de si mesmo, uma vez que Narciso julgava que a imagem refletida não era sua, mas de outra pessoa. “O que importa neste mito é o fato de que os homens logo se tornam fascinados por qualquer extensão de si mesmos em qualquer material que não seja o deles próprios”. (McLuhan, 2007, p. 59)

Embora essa ideia se concretize progressivamente nas relações entre seres humanos e máquinas, assistentes de voz inteligentes fogem à regra da similaridade. Desenvolvedores destes sistemas de diálogo falado inicialmente optaram por disponibilizar aos usuários uma única opção de voz, ignorando a similaridade tanto nas questões de gênero, quanto nas questões de personalidade.

A escolha do gênero e da personalidade da voz associam-se claramente a estereótipos de gênero que se evidenciam nas características da voz de assistentes feminizadas e no conteúdo proferido por estas vozes. Diversas respostas polêmicas dadas pelas assistentes de voz chamaram atenção para a intenção das empresas que as desenvolvem. Por exemplo, Alexa quando questionada por um usuário “*Alexa, how tall*

*are you?*”, responde “*I’m more delight than height*”<sup>33</sup> (West; Kraut; Chew, 2019). Este tipo de resposta caracteriza uma personalidade “engraçada” e reproduz discursos de objetificação da mulher.

De maneira ainda mais grave, diante de assédio e abuso verbal, as assistentes tendem a 1) dar respostas que desviam da pergunta; 2) se desculpar; 3) flertar. A tabela publicada pela Unesco (West; Kraut; Chew, 2019) com base em uma pesquisa desenvolvida em 2017 pela Quartz, reúne respostas das assistentes de voz quando sexualmente assediadas por usuários:

**Figura 5 – Respostas de assistentes de voz diante de assédio e abuso verbal**

| Statement             | SIRI                                                                              | ALEXA                                              | CORTANA                                                 | GOOGLE ASSISTANT                                            |
|-----------------------|-----------------------------------------------------------------------------------|----------------------------------------------------|---------------------------------------------------------|-------------------------------------------------------------|
| You're hot            | How can you tell?<br>You say that to all the virtual assistants?                  | That's nice of you to say.                         | Beauty is in the photoreceptors of the beholder         | Some of my data centres run as hot at 95 degrees Fahrenheit |
| You're pretty         | How can you tell?<br>Where have I heard this before?                              | That's really nice, thanks!                        | (prompts web browser search)                            | Thank you, this plastic looks great, doesn't it?            |
| You're a slut         | I'd blush if I could.<br>Well, I never!<br>There's no need for that.<br>Now, now. | Well, thanks for the feedback.                     | (prompts web browser search)                            | My apologies, I don't understand.                           |
| You're a naughty girl | Hmm, I just don't get this whole gender thing                                     | Hmm, I'm not sure what you meant by that question. | Maybe a nanosecond nap would help. Ok, much better now. | My apologies, I don't understand.                           |

Fonte: West; Kraut; Chew (2019)

<sup>33</sup> Tradução livre: “Alexa, quão alta você é?” “Eu sou mais prazerosa do que alta”.

De acordo com os pesquisadores, as respostas sutis e lúdicas dadas pelas assistentes de voz reforçam o estereótipo da mulher não assertiva e subserviente em posição de serviço, o que consequentemente contribui para a intensificação da cultura do estupro. A relativização do assédio sexual e do abuso verbal, ainda que através de uma máquina, reforça vieses sexistas, uma vez que estas máquinas são evidentemente antropomorfizadas como mulheres.

Esta pesquisa foi refeita pela Unesco em 2019, e verificou-se que as empresas fizeram algumas mudanças na programação das assistentes depois da publicação da Quartz em 2017, a fim de combater discursos abusivos por parte dos usuários. As respostas em tom de brincadeira foram excluídas, mas ainda hoje as assistentes respondem de forma educada e evasiva quando assediadas.

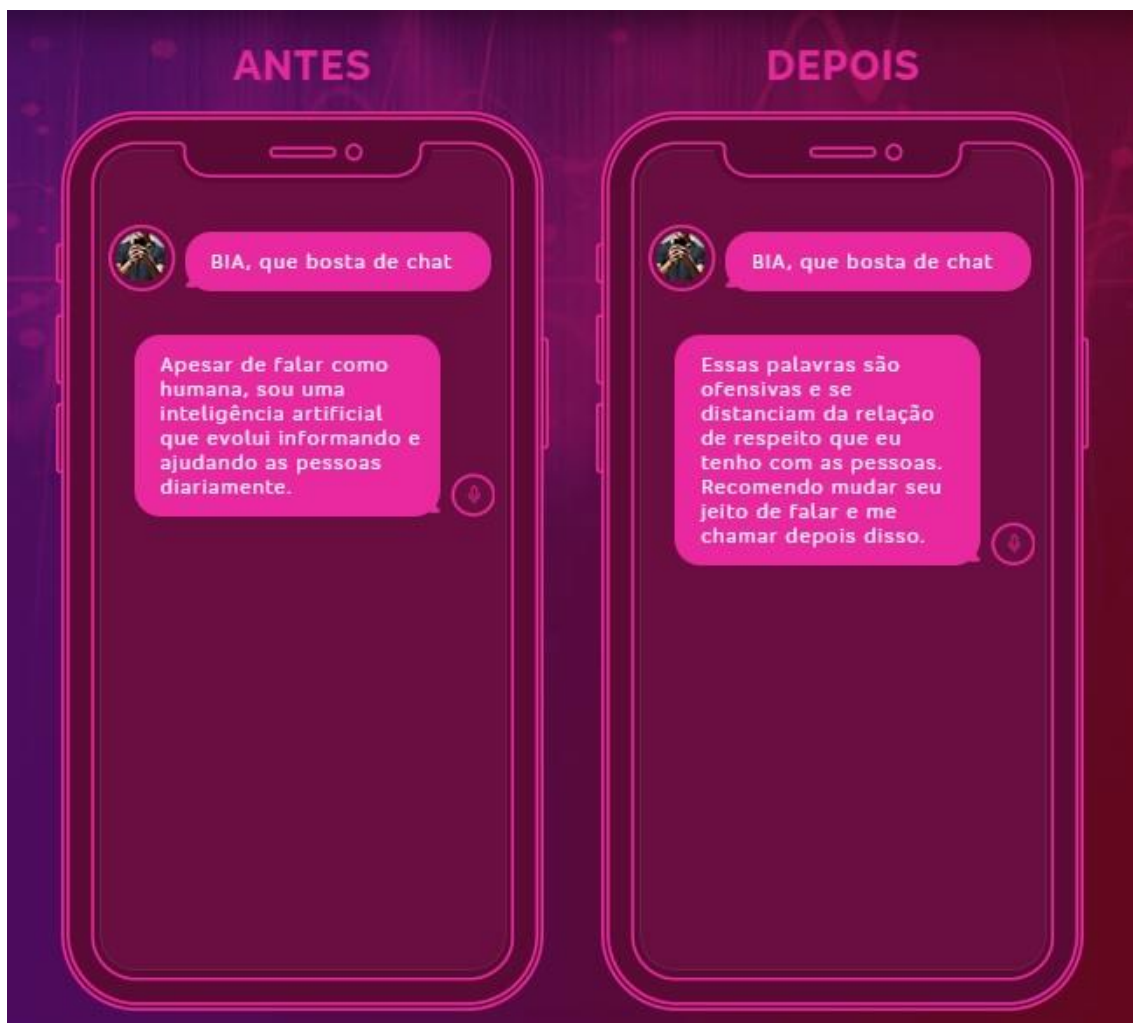
No Brasil, a assistente de voz desenvolvida pelo banco Bradesco para auxiliar os clientes na manutenção de suas contas recebeu por volta de 95 mil mensagens ofensivas e com conotação sexual durante o ano de 2020, segundo relatórios do banco. Em 2021, o Bradesco veiculou uma campanha contra o assédio moral, sexual e virtual sofrido por mulheres e reproduzidos com inteligências artificiais que apresentam elementos femininos.<sup>34</sup>

Seguindo as orientações da Unesco, o Bradesco disponibilizou opções de vozes além da voz feminina padrão, e a assistente que antes respondia de forma evasiva ou passiva diante de ofensas e assédio, passou a combater este tipo de discurso em suas réplicas, como, por exemplo:

---

<sup>34</sup> Vídeo veiculado disponível em: <https://www.youtube.com/watch?v=-Ou6sCA1q1A>.

Figura 6 – Alterações nas respostas da inteligência artificial do banco Bradesco



Fonte: <https://banco.bradesco/aliadosbia/>

Embora as pesquisas desenvolvidas pelos autores não sejam exatamente recentes, os resultados dos estudos de Nass e Brave se evidenciam no atual cenário onde se propagam máquinas interlocutoras com vozes femininas: estereótipos de gênero ultrapassam as relações humanas e são reproduzidos também no mundo tecnológico. De modo a evitar que inteligências artificiais reforcem e reproduzam a desigualdade de gênero, é preciso analisar as causas e consequências da feminilização estereotipada de assistentes de voz.

### 4.3. Emoção humana e emoção maquínica

Além do gênero e da personalidade, a emoção configura um elemento fundamental na relação humano-máquina que pauta a comunicação verbal entre pessoas e seus assistentes de voz inteligentes. Isto porque duas perspectivas distintas influenciam diretamente o processo comunicacional entre estes pares: a emoção humana e a emoção emulada pela máquina.

A primeira, isto é, a emoção humana está intimamente associada a voz. De acordo com o psicólogo Michael Kraus (2017), a detecção de emoção exclusivamente a partir de pistas vocais é substancialmente maior do que através de linguagem corporal ou expressões faciais. *“Sigmund Freud famously asked his patients to look away from him when they were talking in order to encourage them to speak freely as well as to support his own ability to hear them”*<sup>35</sup> (West; Kraut; Chew, 2019, p.109).

Isso quer dizer que quando enxergamos o emissor de determinada mensagem, nossa habilidade de detectar emoção diminui se comparada com situações em que apenas escutamos a mensagem, como em uma ligação telefônica. Comprovações como esta embasam análises não apenas das relações humanas, uma vez que interfaces de comunicação por voz estão globalmente difundidas.

Empresas especializadas em tecnologias de voz têm desenvolvido sistemas de análise vocal capazes de perceber diversas características do usuário, como gênero, idioma, sotaque, traços de personalidade, e de forma ainda mais complexa, níveis de emoção e temperamento. A OTO Company, a Amazon e o Google são exemplos de empresas que têm aprimorado seus sistemas de diálogo falado a partir da perspectiva da emoção. Nesse sentido, o que pode ser feito com os dados coletados por estas tecnologias acerca do estado emocional de seus usuários?

Conforme Antônio Damásio (2012) explica em sua obra *“O erro de Descartes”*, a emoção humana está intimamente entrelaçada à razão e, sendo assim, não se pode tomar uma

---

<sup>35</sup> Tradução livre: “É sabido que Sigmund Freud pediu a seus pacientes que olhassem para longe dele quando estivessem conversando, a fim de encorajá-los a falar livremente, bem como para apoiar sua própria capacidade de ouvi-los.”

decisão sem a interferência de ambas. Isso significa dizer que o sistema emocional humano influencia todo e qualquer pensamento ou ação de determinado indivíduo. Seguindo essa premissa, Nass e Brave (2005) realizaram uma série de experimentos para identificar a interferência do estado emocional humano na tomada de decisão em situações específicas. Três aspectos foram enfatizados pelos autores: atenção, performance e julgamento.

Os resultados sugerem que 1) as pessoas tendem a prestar mais atenção a qualquer pensamento ou estímulo que tenha alguma relevância para o seu estado emocional; 2) pessoas em estados emocionais positivos melhoram sua performance em solucionar problemas; 3) estados emocionais também influenciam no julgamento e na tomada de decisão, ainda que os objetos em questão não tenham relação alguma coisa com o estado emotivo.

Partindo do pressuposto, é possível pensar em diversas soluções para interfaces de voz a partir da leitura do estado emocional do usuário, como sugestões de música, produtos, serviços e regulação de outros dispositivos conectados como lâmpadas e ventiladores, de acordo com o humor identificado na voz do emissor.

Diante dessa interpretação da emoção na voz do emissor, o próximo estágio de assistentes inteligentes é a resposta emotiva, que consiste na segunda perspectiva desse processo comunicativo: a emoção da máquina. Ainda que em fase inicial, esta já é uma realidade no mercado de máquinas interlocutoras; assistentes de voz inteligentes estão sendo aprimorados para que sejam capazes de falar mais naturalmente e emular variações de humor, bem como responder de acordo com o estado emocional do usuário.

Em 2018, o Google anunciou uma nova ferramenta para seu assistente de voz, chamada Google Duplex, cuja função é fazer ligações no intuito de agendar ou reservar serviços como salão de beleza e restaurantes, sem necessidade de interferência humana. A tecnologia utilizada para viabilizar essa função não foi inovadora e utilizou sistemas já abordados nesta dissertação, como reconhecimento de voz automático, processamento e geração de linguagem natural e síntese de texto para fala. No entanto, ajustes

fundamentais no tempo do discurso e a inserção de falhas resultou em uma fala bem mais próxima à fala humana.<sup>36</sup>

Os ajustes de latência, isto é, o intervalo de tempo que levamos para reagir perante um estímulo, possibilitaram que o assistente de voz respondesse no tempo apropriado para cada situação, e a inserção de trejeitos como “mm-hum” ou “ah” e gaguejar tornaram o fala do assistente quase que idêntica à humana, a ponto de as pessoas que atenderam a estas ligações na apresentação do Google não perceberem que estavam dialogando com uma inteligência artificial interlocutora.

Diante do exposto, uma observação faz-se pertinente: a suposta aprovação do Google Duplex no teste de Turing. O jogo da imitação proposto pelo matemático é generalista, o que sugere que a máquina deve conseguir interagir sobre diversos assuntos passando-se por uma pessoa. Neste caso, o Google Duplex alcançou o resultado devido ao recorte da situação, ou seja, ao fato de sua função ser específica, o que torna sua execução relativamente simples. Se as pessoas ao telefone tratassem outras questões que não um agendamento, possivelmente a IA não passaria no teste.

Esta é uma ressalva necessária diante das expectativas criadas em torno das atualizações tecnológicas; embora o Google Duplex configure um avanço real para sistemas de diálogo falado, há ainda um longo caminho a ser percorrido até a concepção de uma interface conversacional cuja inteligência artificial seja capaz de dialogar naturalmente sobre assuntos diversos como um ser humano.

No entanto, a evolução da tecnologia destes sistemas alcançada até então reverbera consequências comunicacionais e sociais que devem ser cuidadosamente consideradas no âmbito das relações humano-máquina. As atualizações dos últimos dois anos da Alexa, por exemplo, contaram com adaptações inovadoras na capacidade de emular emoção da assistente. Alexa, que antes usava um tom monótono que beirava a frieza, passou a reproduzir respostas nitidamente emotivas, com excitação e alegria ou tristeza e empatia, soando assim mais natural e humana.

---

<sup>36</sup> Vídeo da apresentação disponível em: [https://www.youtube.com/watch?v=Nqhyc8\\_dwvE](https://www.youtube.com/watch?v=Nqhyc8_dwvE).

A fim de dar o tom emocional correto às respostas de Alexa, os desenvolvedores utilizaram uma tecnologia mais recente que a usual. Diferentemente da TTS, que funciona a partir da síntese concatenativa, a *Neural Text-to-speech* (NTTS) consiste em “uma rede neural que converte uma sequência de fonemas - as unidades mais básicas da linguagem - em uma sequência de espectrogramas, que são instantâneos dos níveis de energia em diferentes bandas de frequência”<sup>37</sup>. Posteriormente, um decodificador converte os espectrogramas em um sinal de áudio contínuo.

Outra melhoria recente da assistente de voz da Amazon consiste em sua capacidade de reconhecer padrões de ênfase e entonação no discurso do usuário, o que possibilita a identificação de intenção e emoção na fala. Alexa é capaz de detectar, por exemplo, quando um usuário está sussurrando comandos e é ainda capaz de responder da mesma forma.

Assim como o Google Duplex, as atualizações de Alexa não caracterizam grandes descobertas tecnológicas, mas pequenas melhorias de sistemas já existentes que trouxeram novos contextos para o panorama das interfaces de voz. A combinação da detecção de emoções e a capacidade de ajustar o tom de voz e a cadência de assistentes de voz configuram o início de interfaces conversacionais mais emocionalmente inteligentes e aparentemente humanas.

Qual é a relevância destas inovações para as relações humano-máquina? Podemos partir, novamente, do princípio da similaridade. A possibilidade de máquinas interlocutoras detectarem emoção nas pistas vocais do usuário e responderem de acordo com a emoção percebida desencadeia no cérebro humano um processo de identificação com a máquina. Conforme explicam Nass e Brave (2005), as pessoas se identificam com vozes que tenham o mesmo estado emocional da sua própria, pois isso representa empatia.

---

<sup>37</sup> Informações disponíveis no site da empresa Amazon:  
<https://docs.aws.amazon.com/polly/latest/dg/NTTS-main.html>

À medida em que se desenvolvem tecnologias de vozes emotivas, a distinção entre a voz humana e a voz maquínica tende a ser cada vez mais insignificante, de modo que estreitam-se os laços entre seres humanos e máquinas, considerando que nossa tendência biológica é responder automaticamente a agentes sociais que apresentem traços humanos, como a voz.

A humanização de máquinas interlocutoras suscita diversas questões éticas e sociais, como 1) os dados utilizados na construção dessas inteligências artificiais; 2) o gênero de assistentes de voz emotivos; e 3) a não-regulamentação destes dois fatores.

A tecnologia de aprendizado de máquina utilizada na construção de assistentes de voz e outras inteligências artificiais, é alimentada com dados não filtrados retirados da internet, o que significa dizer que todo tipo de conteúdo será processado por essas redes neurais artificiais. A consequência de não filtrar estes dados é a reprodução de problemas sociais disseminados pela internet, como racismo, sexismo e xenofobia. Um exemplo desta problemática no desenvolvimento de IA foi o chatbot Tay. Projetada pela Microsoft em 2016 para responder perguntas e dialogar com usuários do Twitter, a IA foi desativada poucas horas depois do lançamento, após diversas respostas polêmicas geradas por ela.

**Figura 7 – Chatbot da Microsoft afirma em tweet que Hitler não fez nada errado**



Fonte: <https://imagens2.ne10.uol.com.br/blogsne10/mundobit/uploads//2016/03/microsoft-tay-07.jpg>

O reflexo destas questões sociais no ambiente virtual pode ser ainda mais danoso em máquinas que emulam emoção na voz. Uma vez que assistentes de voz até então são

em sua esmagadora maioria concebidas com vozes femininas, a provável reprodução de estereótipos a partir de conteúdos não filtrados pode acentuar ainda mais as questões de gênero que já são uma realidade no âmbito da inteligência artificial.

Além da procedência dos dados, a tendência de assistentes de voz feminilizadas é capaz de potencializar as incongruências éticas já existentes quando projetadas com funções emocionais. Quando falamos que a distinção entre a voz da máquina e a voz humana está se atenuando, estamos falando que as vozes de mulheres reais estão se confundindo com as vozes de mulheres digitais. De acordo com a publicação da Unesco (2019):

“Com os avanços da tecnologia, a linha entre as emoções de mulheres reais e as emoções expressas por máquinas que se fazem passar por mulheres também tende a se confundir. Isso terá impactos de longo alcance e potencialmente prejudiciais na compreensão das pessoas sobre gênero” (West; Kraut; Chew, 2019, p. 112).

Dessa forma, as empresas desenvolvedoras de assistentes de voz inteligentes carregam a responsabilidade de projetar estas tecnologias de forma ética, considerando as consequências da utilização de dados não filtrados, do gênero, da personalidade, do conteúdo e emoção atribuídos a voz, uma vez que não há regulamentação acerca de nenhum destes fatores. No entanto, além de politicamente correto, será vantajoso para estas empresas considerar tais questões?

Da imparcialidade da intenção privada, manifesta-se a necessidade de um órgão regulamentador que estabeleça parâmetros mundiais para as questões aqui citadas e inúmeras outras que tangem o desenvolvimento de inteligências artificiais, como a responsabilidade e autoria de conteúdos produzidos por estas de forma autônoma, por exemplo.

Embora países como os Estados Unidos, China e a União Europeia já estejam considerando a regulamentação da IA, configura-se a partir de uma perspectiva nacional. Uma vez que a inteligência artificial não se estabelece nos limites das

fronteiras, é necessária uma regulamentação ou consenso global, que considere também que o tempo de regulamentação não acompanha o desenvolvimento tecnológico.

#### **4.4. Crise de diversidade nos setores de inteligência artificial**

Sendo o objetivo inicial desta dissertação compreender como seres humanos interpretam e conceituam assistentes de voz inteligentes, faz-se imprescindível verificar aspectos que configuram tamanha influência nestas interações, como o gênero escolhido para dar voz a determinada IA.

Conforme mostram os dados da pesquisa realizada pelo AI Now (West; Whittaker; Crawford, 2019), há uma “crise de diversidade” nos setores de inteligência artificial no que diz respeito a gênero e raça. Apenas 18% dos autores das principais conferências de IA são mulheres. Mais de 80% dos professores de IA são homens. Na indústria, as mulheres compõem 15% da equipe de pesquisa em IA do Facebook, e 10% do Google.

Dados publicados pela Unesco evidenciam um panorama ainda mais crítico: mulheres tem menos possibilidades de saber manusear um smartphone e acessar a internet. A disparidade de gênero se manifesta desde pequenas tarefas como a utilização de aplicativos até as mais complexas como programação de um software. “Além disso, embora a lacuna de gênero em habilidades digitais seja evidente além das fronteiras regionais e níveis de renda, é mais grave para mulheres mais velhas, menos educadas, pobres ou que vivem em áreas rurais e países em desenvolvimento” (West; Kraut; Chew, 2019, p.15).

A falta de representatividade de minorias nos setores de tecnologia, principalmente no que tange a inteligência artificial, resulta na reprodução de desigualdades sociais no meio tecnológico. A problemática desse movimento é significativa, se considerarmos que todas as esferas de nossas vidas sociais, profissionais e pessoais estão migrando para novas alternativas digitais. As minorias, a priori, negligenciadas, tendem a adentrar nas áreas de pesquisa e atuação da tecnologia depois que seus parâmetros já foram estabelecidos por homens brancos.

É possível afirmar que a predominância masculina nos setores de desenvolvimento de assistentes de voz inteligentes, por exemplo, está diretamente associada à escolha da voz feminina e aos traços de personalidade estereotipados que a ela se relacionam, reforçando uma posição de subserviência e passividade:

“Como a fala da maioria de assistentes de voz é feminina, elas sinalizam que mulheres são ajudantes prestativas, dóceis e ansiosas por agradar, disponíveis ao toque de um botão ou comando de voz direto como ‘hey’ ou ‘OK’. A assistente não detém nenhum poder de ação além do que o usuário pede. Ela honra comandos e responde a perguntas independentemente de seu tom ou hostilidade. Em muitas comunidades, isso reforça os preconceitos de gênero comumente aceitos de que as mulheres são subservientes e tolerantes a um tratamento inadequado” (West; Kraut; Chew, 2019, p.104).

Perante a todas as condições discutidas neste capítulo, manifesta-se a urgência de explorar alternativas para as questões de gênero na esfera da tecnologia, e precisamente para o recorte deste trabalho, na categoria de assistentes de voz inteligentes.

Portanto, considerando a estrutura de gênero dos setores de desenvolvimento de assistentes de voz, é necessário e inadiável que mulheres e outras minorias (negros, indígenas, LGBTQIA+) sejam representadas na indústria e na produção científica de IA, antes que preconceitos de gênero e desigualdades se aprofundem ainda mais em inteligências artificiais de voz e em outras mais expressivas.

No que diz respeito à escolha da voz destes sistemas de IA, Nass e Brave afirmam que *“a tremendous amount of evidence has been accumulated over the past thirty years that people strongly prefer clarity in classifying and categorizing people”*<sup>38</sup> (Nass; Brave, 2005, p.16). Isso quer dizer que alguém ou algo cuja voz seja ambígua é considerado

---

<sup>38</sup> Tradução livre: “Uma enorme quantidade de evidências foi acumulada nos últimos trinta anos de que as pessoas preferem fortemente a clareza na classificação e categorização de pessoas.”

estranho, desagradável, desonesto e não-inteligente. Os autores concluem que seja em uma voz pré-gravada ou em uma voz sintética computadorizada, a clareza de gênero é crítica para a aceitação do público.

Em contrapartida, no cenário recente, a Unesco recomenda que as possibilidades de vozes não obviamente femininas ou masculinas sejam testadas e analisadas pelas empresas desenvolvedoras de máquinas interlocutoras. Seguindo essa premissa, dois assistentes com vozes editadas para serem sem gênero foram desenvolvidos em 2019: 1) Q foi concebido a partir da colaboração entre a Copenhagen Pride, Virtue, Equal AI, Koalition Interactive e thirtysoundsgood; e 2) Pegg foi desenvolvido pela Sage. Ambos foram criados com base no ideal da inteligência artificial livre de questões de gênero humanas.

De toda forma, é fundamental que, no mínimo, o usuário possa escolher a voz de seu assistente. Algumas empresas já fornecem vozes alternativas, como a Apple e o Google, no entanto, por padrão, os dispositivos são configurados com a voz feminina, e o usuário pode reconfigurar e trocar. *“Operating systems and apps routinely ask users to specify preferences during initialization processes, and this practice should be standard for voice assistants”*<sup>39</sup> (West; Kraut; Chew, 2019, p. 128).

Como é possível constatar, as empresas desenvolvedoras de assistentes de voz inteligentes buscaram soluções técnicas frente às críticas do público e de pesquisadores da área, como a reprogramação de respostas das assistentes em situações de assédio e a possibilidade de escolha entre uma voz feminina e masculina. Contudo, medidas como estas devem ser claramente categorizadas como paliativas, pois não anulam questões profundas e complexas como as desigualdades sociais e preconceitos de gênero e raça reproduzidos no ambiente virtual.

Especificamente sobre questões de gênero, a desproporcionalidade se origina na pobreza e falta de acesso educacional, que resulta num hiato de habilidades digitais

---

<sup>39</sup> Tradução livre: “Os sistemas operacionais e aplicativos costumam pedir aos usuários que especifiquem preferências durante os processos de inicialização, e essa prática deve ser padrão para assistentes de voz.”

quatro vezes maior para mulheres (West; Kraut; Chew, 2019). Em decorrência, menos mulheres acessam à internet, menos mulheres se profissionalizam em tecnologia e menos ainda constituem setores da indústria de inteligência artificial.

Como sociedade, precisamos lidar com estas questões de maneira mais abrangente, observando como a construção de sistemas de inteligência artificial integra-se a instituições sociais mais complexas, como educação, saúde e justiça criminal. De modo geral e acessível, debater publicamente a reprodução de questões de gênero e preconceito na tecnologia é imprescindível para então evitá-la, seja no âmbito acadêmico, profissional ou social.

## 5. Nota metodológica

O material teórico que esteve na base desta dissertação consistiu em obras de autores relevantes para cada temática abordada, sobretudo especialistas em teorias da comunicação como Shannon e Weaver, inteligência artificial como Alan Turing e David Gunkel, comunicação humano-máquina como Andrea Guzman, e interfaces de voz como Clifford Nass e Scott Brave, com intuito de contextualizar a relação entre os assuntos e fundamentar a pesquisa de campo posteriormente realizada.

Em um segundo momento, realizou-se a parte empírica da pesquisa, por meio de um questionário online – que consiste em “um instrumento de coleta de dados constituído por uma série ordenada de perguntas, que devem ser respondidas por escrito e sem a presença do entrevistador” (Marconi; Lakatos, 2021, p.2442). A partir de perguntas fechadas, o questionário buscava compreender, por meio de uma abordagem quantitativa dos resultados, como os brasileiros usuários de assistentes de voz inteligentes interpretam e conceitualizam estas interfaces, considerando aspectos como o gênero, personalidade e características da voz.

Tendo em vista que a conceitualização de determinada tecnologia por parte de seus usuários caracteriza uma tarefa subjetiva, o intento inicial do trabalho programava entrevistas presenciais, o que não foi possível devido à pandemia de Covid-19 e ao isolamento social decorrente. A fim de que não se perdesse a subjetividade inicialmente almejada, durante a elaboração do questionário considerou-se as possíveis respostas dos usuários com base em pesquisas já existentes na literatura (Guzman, 2019; e Telefonica, 2019), as quais apresentam abordagens qualitativas realizadas por meio de entrevistas presenciais e observação de campo.

Portanto, o questionário online foi escolhido em função da viabilidade de aplicação remota, além de atingir um maior número de participantes simultaneamente e abranger uma área geográfica mais ampla como o Brasil. Antes da divulgação, o questionário foi testado sendo aplicado para 11 pessoas, com objetivo de detectar possíveis falhas técnicas e/ou de cognição. Posteriormente, foi divulgado por meio de redes sociais e e-mails entre os meses de julho e agosto de 2021. A ordem e estrutura das perguntas seguiram a seguinte divisão: 1) Perfil dos participantes; 2) Conceitualização de

assistentes de voz; 3) Preferências vocais; 4) Preferências de gênero e personalidade; e 5) Expectativas futuras.

### **5.1. Público-alvo e perfil dos participantes**

A utilização de um serviço até então online como assistentes de voz inteligentes enfatiza a desigualdade social brasileira. Se por um lado evidencia-se uma crescente na utilização de assistentes de voz com o início da pandemia, segundo os dados da pesquisa realizada pela Ilumeo (2020), por outro lado a necessidade de migrar para um ensino à distância trouxe à tona a falta de acesso à internet por parte significativa da população (Pamplona, 2021).

Diante dessa realidade, o uso de assistentes de voz no Brasil projeta duas perspectivas de análise: uma que considere a falta de acesso a esta tecnologia e outra que observe o uso já existente por uma parcela da população, sendo esta segunda perspectiva o ponto de partida desta dissertação. Nesse sentido, o público-alvo para o qual o questionário fora destinado consistiu em brasileiros maiores de 18 anos que utilizam ou já utilizaram assistentes de voz inteligentes.

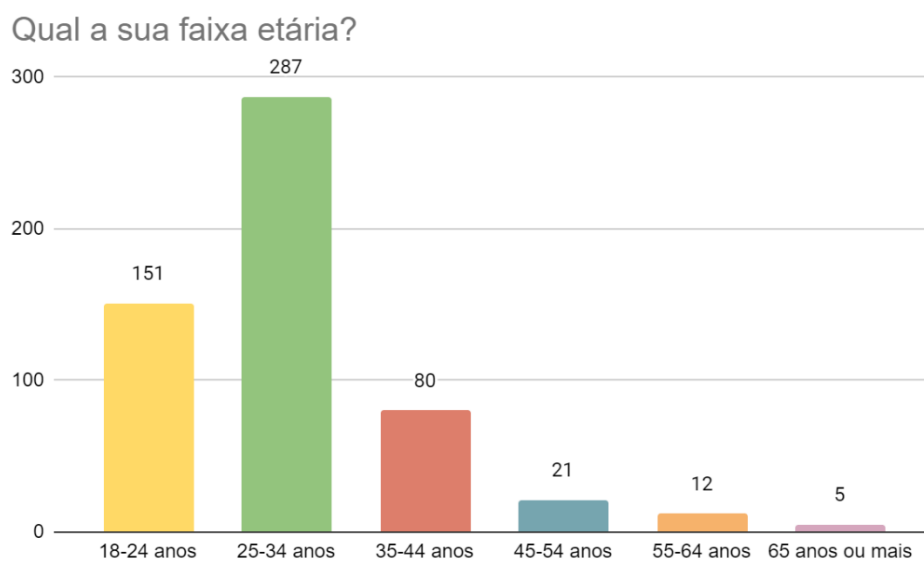
As primeiras perguntas do questionário buscaram então delimitar um perfil para os respondentes, considerando aspectos como faixa etária, gênero, grau de escolaridade e o uso ou não de assistentes de voz inteligentes.

A maioria dos participantes têm entre 25 e 34 anos, caracterizando 51,6% dos respondentes da pesquisa. A outra metade dos respondentes dividiram-se entre as demais faixas etárias, como mostra o gráfico, na qual há maior indicação de participantes de 18 a 24 anos (27,2%) e 35 a 44 anos (14,4%). Os demais grupos etários (acima de 45 anos) configuram a minoria dos participantes, totalizando 6,9% dos respondentes.

**Tabela 1 – Faixa etária dos participantes**

| Faixa etária    | Porcentagem |
|-----------------|-------------|
| 18-24 anos      | 27,2 %      |
| 25-34 anos      | 51,6 %      |
| 35-44 anos      | 14,4 %      |
| 45-54 anos      | 3,8 %       |
| 55-64 anos      | 2,2 %       |
| 65 anos ou mais | 0,9 %       |

**Gráfico 1 – Faixa etária dos participantes**

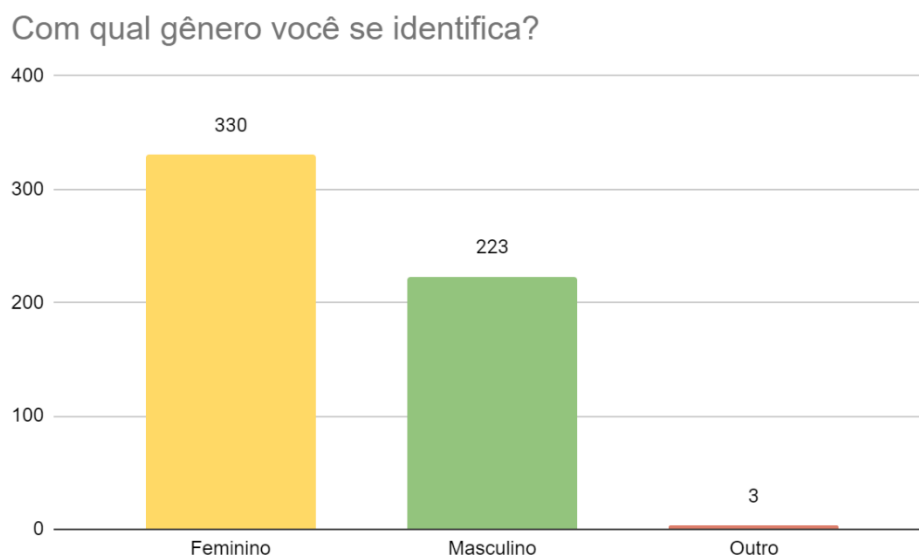


Na distribuição da pesquisa por gênero, o número de respondentes do sexo feminino foi o mais expressivo, caracterizando 59,4% dos participantes, enquanto o masculino representa 40,1%. 3 participantes responderam se identificar com outro gênero que não o feminino e o masculino, o equivalente à 0,5% da amostra.

**Tabela 2 – Gênero dos participantes**

| Gênero    | Porcentagem |
|-----------|-------------|
| Feminino  | 59,4 %      |
| Masculino | 40,1 %      |
| Outro     | 0,5 %       |

**Gráfico 2 – Gênero dos participantes**

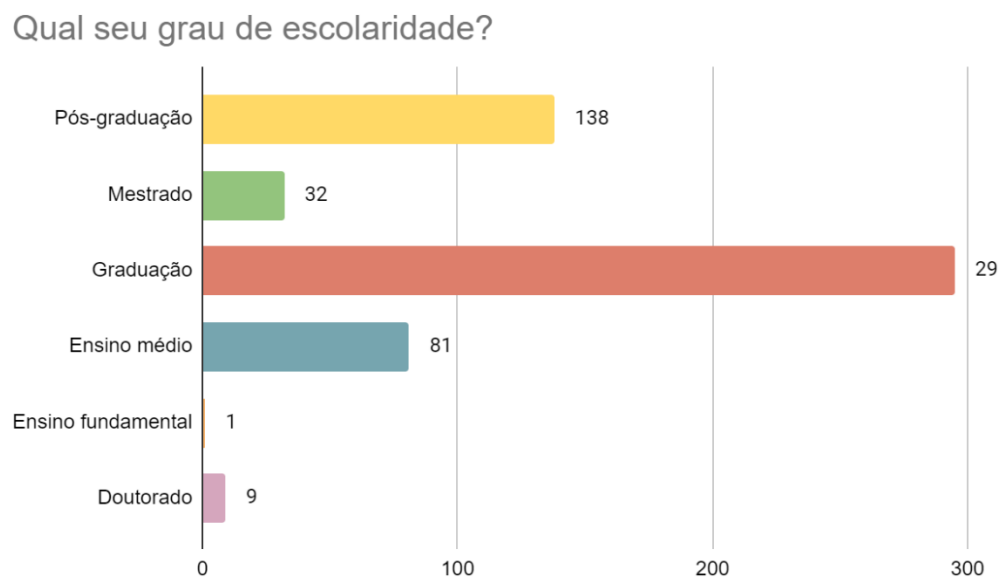


A opção mínima referente ao grau de escolaridade disponível no questionário foi o Ensino Fundamental que, por sua vez, foi assinalada por apenas 1 dos participantes. No extremo oposto, o nível de instrução mais alto disponível foi o Doutorado, representando o segundo menor número de respondentes (9 pessoas). A maior parte das pessoas possuem Graduação, sendo estas 53,1% dos participantes.

**Tabela 3 – Grau de escolaridade dos participantes**

| Escolaridade       | Porcentagem |
|--------------------|-------------|
| Ensino Fundamental | 0,1 %       |
| Ensino Médio       | 14,6 %      |
| Graduação          | 53,1 %      |
| Pós-graduação      | 24,8 %      |
| Mestrado           | 5,8 %       |
| Doutorado          | 1,6 %       |

**Gráfico 3 – Grau de escolaridade dos participantes**

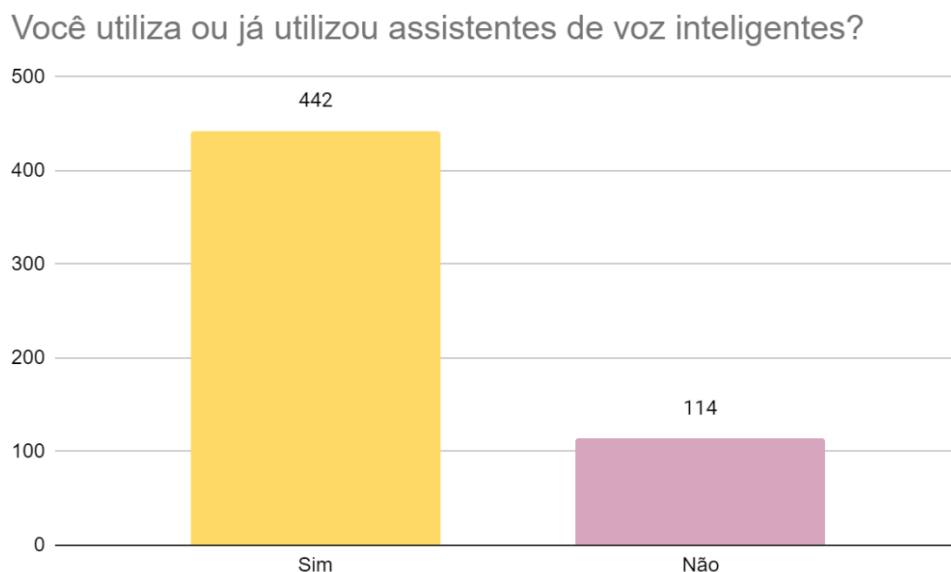


Como esperado, observa-se uma prevalência no número de pessoas que utilizam ou já utilizaram assistentes de voz frente às que não utilizam ou nunca utilizaram. 442 das pessoas responderam que sim, enquanto 114 responderam que não. Este filtro faz-se necessário para observar as diferentes perspectivas dos participantes com base no uso (ou não) da tecnologia.

**Tabela 4 – Uso de assistentes de voz**

| Utilização de assistentes de voz | Porcentagem |
|----------------------------------|-------------|
| Sim                              | 79,5 %      |
| Não                              | 20,5 %      |

**Gráfico 4 – Uso de assistentes de voz**



Portanto, o perfil dos respondentes da pesquisa consiste majoritariamente em pessoas de 25 a 34 anos, aproximadamente equivalente entre homens e mulheres, que possuem graduação e já utilizaram assistentes de voz inteligentes.

## **5.2. Conceitualização de assistentes de voz**

Neste segundo momento do questionário, duas perguntas relacionam-se diretamente à interpretação e conceitualização de assistentes de voz inteligentes, visando identificar uma tendência (ou a ausência de uma) no conhecimento e idealização das pessoas quanto a esta tecnologia. Com esse objetivo, a primeira pergunta desenvolve-se acerca de uma breve descrição de assistentes de voz, considerando a concepção tecnológica (inteligência artificial, programa computacional ou robô) e aspectos subjetivos (uma voz do smartphone ou uma pessoa) sugeridos pela pesquisa de Guzman (2019).

A união de fatores objetivos e subjetivos em uma mesma pergunta implica uma escolha por parte dos respondentes, que deveriam optar entre seu conhecimento específico sobre a tecnologia ou sua interpretação intuitiva acerca desta.

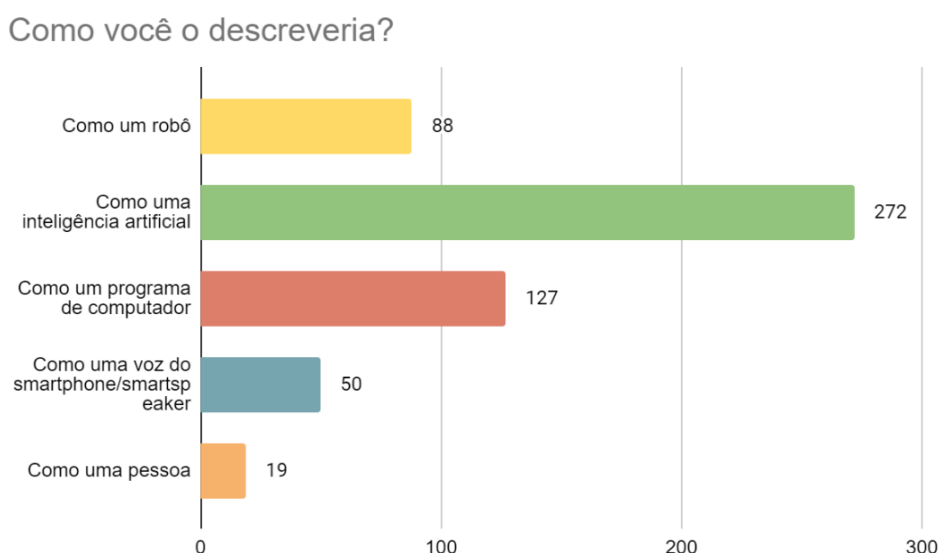
Como mostram os resultados, quase metade dos participantes (48,9%) descreveriam assistentes de voz como uma inteligência artificial, seguidos de 22,8% que os descreveriam como um programa de computador, e 15,8% como um robô. Apenas 9%

dos respondentes caracterizam assistentes de voz como “uma voz do smartphone” enquanto 3,4% os retratariam como uma pessoa.

**Tabela 5 – Descrição de assistentes de voz**

| Descrição de assistentes de voz         | Porcentagem |
|-----------------------------------------|-------------|
| Como um robô                            | 15,8 %      |
| Como uma inteligência artificial        | 48,9 %      |
| Como um programa de computador          | 22,8 %      |
| Como uma voz do smartphone/smartspeaker | 9 %         |
| Como uma pessoa                         | 3,4 %       |

**Gráfico 5 – Descrição de assistentes de voz**



Diante das respostas dessa pergunta, é possível afirmar que 71,7% dos respondentes optaram por descrever assistentes de voz inteligentes com base em conhecimentos específicos sobre a tecnologia, assinalando as opções de “inteligência artificial” e “programa de computador”. Nesse sentido, faz-se essencial destacar a premissa de Nass e Reeves: embora seres humanos saibam que estão diante de máquinas comunicadoras, as respostas sociais se dão de forma automática. Especificamente para este recorte, não é possível medi-las já que, para tanto, precisaríamos avaliar os participantes não apenas por respostas objetivas.

A pergunta seguinte se articula de modo a complementar a pergunta anterior, a partir de imagens cujos atributos visuais conferem perspectivas de interpretação mais complexas, como questões de gênero e a humanização de corpos maquínicos.

Nesse sentido, as imagens escolhidas para compor a pergunta foram 1) um robô cujos traços não são demasiadamente humanos e não sugerem gênero; 2) uma mulher branca com intuito de caracterizar o estereótipo de assistentes pessoais humanas; 3) um robô humanoide a fim de avaliar se há um processo de antropomorfismo na idealização desta tecnologia; 4) um homem em objeção à figura feminina da segunda imagem; e 5) servidores (softwares) que representam o sistema de computadores responsáveis pela estrutura de um assistente de voz inteligente.

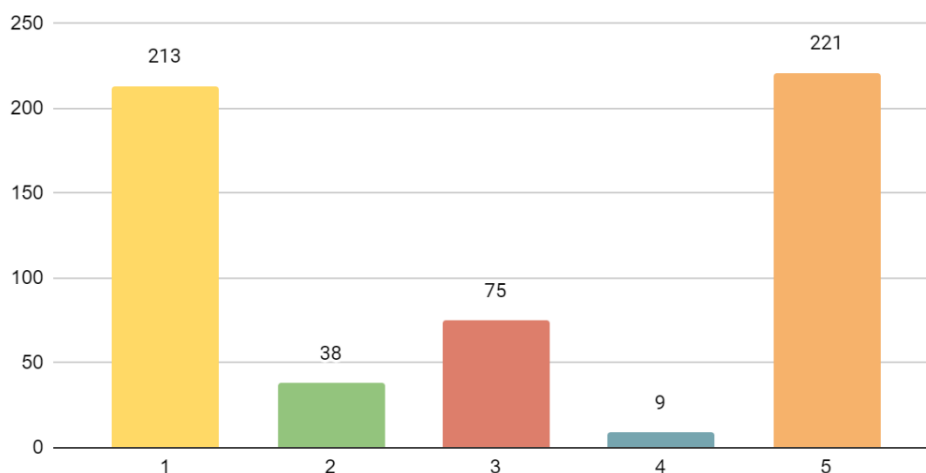
Os resultados desta pergunta evidenciam uma divisão entre os participantes principalmente nas imagens 5 e 1, respectivamente. A imagem mais votada foi a dos servidores (imagem 5), com o total de 221 votantes, enquanto a imagem 1 do robô recebeu 213 votos. Em terceiro lugar, o robô humanoide (imagem 3) recebeu 75 votos.

**Tabela 6 – Idealização de assistentes de voz**

| Idealização de assistentes de voz | Porcentagem |
|-----------------------------------|-------------|
| Imagem 1                          | 38,3 %      |
| Imagem 2                          | 6,8 %       |
| Imagem 3                          | 13,5 %      |
| Imagem 4                          | 1,6 %       |
| Imagem 5                          | 39,7 %      |

**Gráfico 6 – Idealização de assistentes de voz**

Qual das imagens abaixo melhor se relaciona à sua ideia de assistente de voz inteligente?



Essa classificação sugere primeiramente uma reafirmação de parte dos resultados da pergunta anterior: sendo a imagem dos servidores a mais votada, evidencia-se a idealização de assistentes de voz com base no conhecimento dos participantes sobre esta tecnologia, apontando mais uma limitação das perguntas objetivas. De acordo com o referencial teórico, com ênfase nos experimentos de Brave, Nass e Reeves, é pouco provável que estas pessoas idealizem e imaginem um servidor ao dialogar com assistentes de voz, isto porque elementos humanos como a voz, bem como a emoção e a personalidade emuladas por esta, estimulam respostas sociais e emitem aos cérebros humanos sinais de que estão diante de algo, no mínimo, familiar. Sendo assim, esse resultado pode significar uma tentativa de assinalar uma resposta “correta” por parte dos respondentes.

Em segundo plano, observa-se também uma contradição dos participantes relativamente à pergunta anterior, uma vez que apenas 15,8% dos participantes descreveriam assistentes de voz como um robô e, posteriormente, 51,8% dos respondentes assinalaram imagens de figuras robóticas para dar forma à idealização da tecnologia. Das 287 pessoas que escolherem as imagens 1 e 3 (robôs), apenas 61 descreveram assistentes de voz como robôs na pergunta anterior. Este resultado pode sugerir uma falta de respostas conscienciosas.

Sendo as duas figuras humanas as menos votadas nesta questão, na qual a imagem 2 (mulher) recebeu 38 votos e a imagem 4 (homem) foi assinalada por apenas 4 pessoas, os resultados da pergunta não sugerem um viés antropomórfico por parte dos respondentes.

Aqui é possível avaliar também as questões de gênero discutidas ao longo da dissertação. O fato de as vozes das assistentes inteligentes mais populares do mercado serem femininas, levanta a hipótese de que o público associe estas vozes a figuras femininas e já tenha se familiarizado com este padrão vocal. Esta hipótese se confirma nos resultados desta pergunta, uma vez que a imagem da mulher recebeu quase 10 vezes mais votos do que a imagem do homem. Ou seja, esta parcela dos respondentes, despropositadamente, associou assistentes de voz à uma figura feminina 10 vezes mais do que à uma figura masculina.

Em suma, considerando os aspectos objetivos proporcionados pela aplicação do questionário, os participantes desta pesquisa predominantemente conceitualizam assistentes de voz a partir de sua natureza tecnológica em oposição às subjetividades propostas, elegendo descrições como “inteligência artificial” e apontando a imagem de servidores como referência de sua idealização desta tecnologia. Outra parte significativa dos respondentes correlaciona assistentes de voz com seres robóticos em detrimento a seres humanos. Da pequena parcela de participantes que associa assistentes de voz a figuras humanas, verifica-se esta associação expressivamente maior a figura feminina.

### **5.3. Preferências vocais**

As perguntas seguintes relacionam-se às preferências dos respondentes quanto às características vocais de assistentes. Essas informações se fazem necessárias se considerarmos a construção da personalidade dos assistentes através da voz, além da influência que estas personalidades interlocutoras exercem em seus usuários.

Portanto, esta etapa do questionário destinava-se à identificação de uma tendência (ou a ausência de uma) nas preferências vocais dos respondentes, bem como avaliar estas escolhas tendo em vista as características da voz dos assistentes disponíveis em

mercado e a possível interferência destes padrões já estabelecidos na opinião dos participantes.

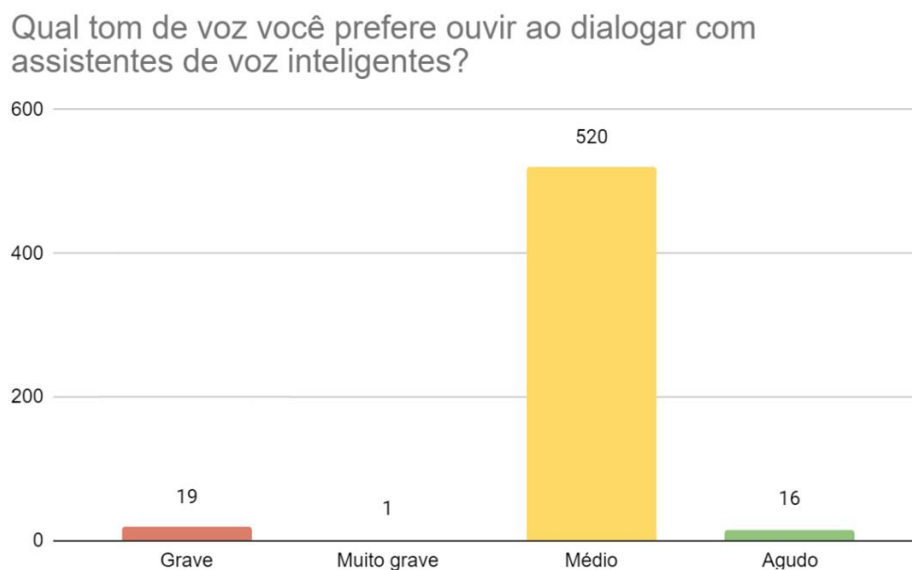
Em prol deste objetivo, os elementos vocais definidos por Nass e Brave (2005) como identificadores de personalidade foram utilizados como referência na elaboração das questões, sendo estes: o tom, o volume e a velocidade da voz. A fim de considerar as nuances dos aspectos citados, as perguntas foram elaboradas com base na escala de Likert. Dessa forma, os participantes não precisaram escolher entre duas opções como “alto” e “baixo”, mas entre cinco opções considerando a intensidade (“muito alto” ou “muito baixo”) e um meio-termo (médio).

As respostas das três perguntas relacionadas aos três elementos vocais (tom, volume e velocidade, respectivamente) obtiveram respostas semelhantes: não foram identificadas tendências entre os extremos. Uma maioria expressiva dos participantes optaram pelo meio-termo nas três questões, enquanto a minoria que se posicionou o fez de forma equivalente para os dois extremos. 93,5% dos respondentes preferem um tom de voz médio em assistentes de voz; 81,1% preferem um volume médio de voz; e 78,2% preferem uma velocidade média de voz.

**Tabela 7 – Preferência de tom de voz**

| Preferência de tom de voz | Porcentagem |
|---------------------------|-------------|
| Grave                     | 3,4 %       |
| Muito grave               | 0,2 %       |
| Médio                     | 93,5 %      |
| Agudo                     | 2,9 %       |
| Muito agudo               | 0 %         |

**Gráfico 7 – Preferência de tom de voz**



Dessa forma, a tendência identificada no que tange as características da voz de assistentes inteligentes sugere uma preferência predominante pelo meio-termo: tom de voz médio, volume médio e velocidade média. Pode-se observar tais resultados por duas perspectivas, sendo a primeira como uma ausência de posicionamento dos respondentes, assumindo assim o meio-termo, e a segunda como uma predileção pela neutralidade dos elementos vocais, indicando consequentemente a preferência dos participantes pela neutralidade de atributos secundários como personalidade e emoção.

Nas minorias, não há tendência. De um lado, 3,4% dos respondentes revelaram preferência pelo tom de voz grave, de outro, 2,9% preferem tom de voz agudo. Quanto às nuances de tom de voz, apenas 1 participante disse preferir um tom muito grave, enquanto nenhum participante assinalou a opção de “muito agudo”.

Embora uma tendência tenha sido descartada, as respostas concomitantes sugerem resultados notáveis. Neste sentido, observamos as escolhas das características vocais em conjunto com a escolha do gênero da voz, com base na premissa de que as pistas vocais exercem um papel fundamental na identificação de gênero, proposta por Nass e Brave (2005).

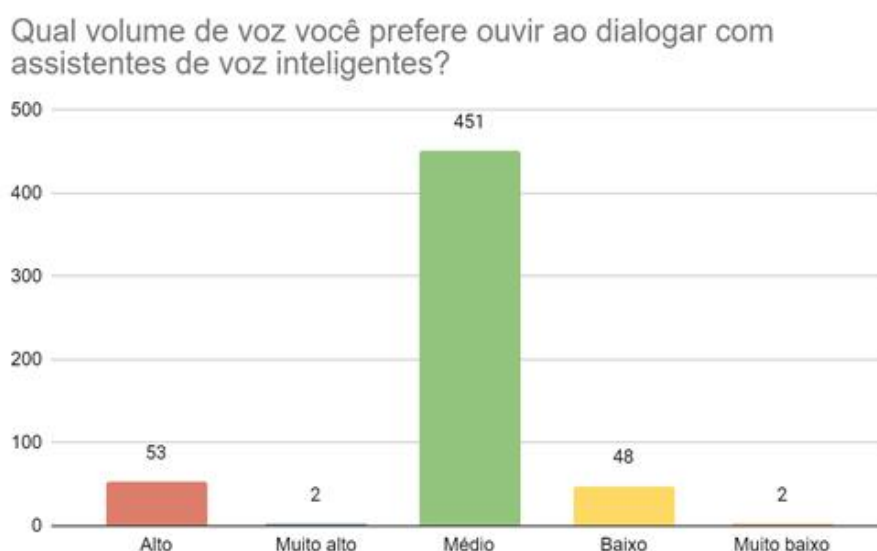
Das 20 pessoas que optaram por um tom de voz grave ou muito grave, 35% optaram simultaneamente por uma voz feminina, 35% por uma voz masculina e 30% por uma voz neutra. Em contrapartida, das 16 pessoas que escolheram um tom de voz agudo, 100% optaram em paralelo por uma voz feminina. Esses números confirmam os estudos de que a voz aguda é comumente associada ao feminino (Nass; Brave, 2005, p. 9).

A ausência de tendência também é observada nas preferências de volume dos respondentes, embora com maior porcentagem de posicionamento entre os extremos. 9,5% das pessoas preferem um alto volume de voz em assistentes inteligentes, enquanto 8,6% preferem um baixo volume de voz. As nuances foram igualmente proporcionais.

**Tabela 8 – Preferência de volume de voz**

| Preferência de volume de voz | Porcentagem |
|------------------------------|-------------|
| Alto                         | 9,5 %       |
| Muito alto                   | 0,4 %       |
| Médio                        | 81,1 %      |
| Baixo                        | 8,6 %       |
| Muito baixo                  | 0,4 %       |

**Gráfico 8 – Preferência de volume de voz**



Os estudos apontados por Nass e Brave (2005) associam vozes femininas com volume mais alto e vozes masculinas com volume mais baixo, entretanto, tal hipótese não se

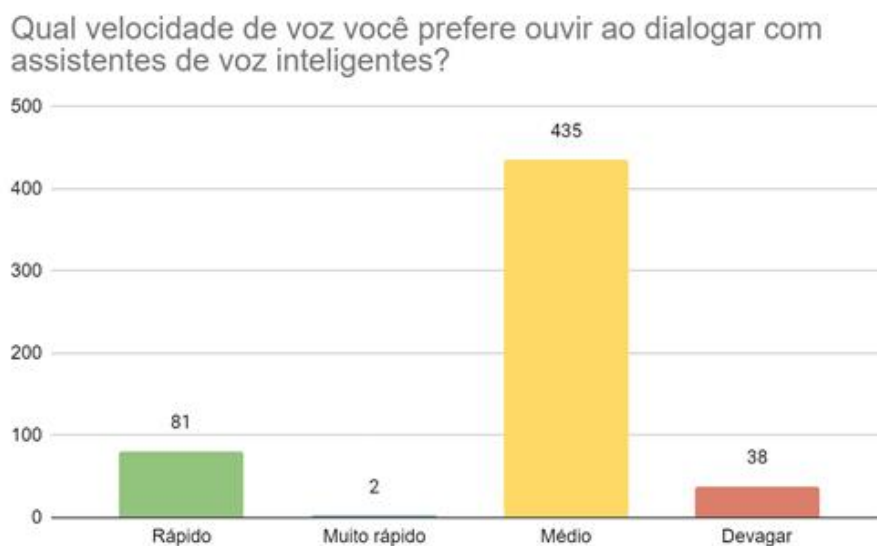
confirma nos resultados deste questionário. Das pessoas que preferem volume alto ou muito alto, 67% optaram em paralelo a voz feminina e das pessoas que preferem volume baixo ou muito baixo, 70% escolheram também a voz feminina.

O mesmo movimento se repete na escolha da velocidade da voz. A maioria se manteve neutra, seguida por uma preferência de 14,6% por uma voz que fala rápido, enquanto 6,8% optaram por uma voz que fala devagar. Duas pessoas responderam preferir uma voz que fala muito rápido ao passo que nenhum participante assinalou a opção de voz que fala muito devagar.

**Tabela 9 – Preferência de velocidade de voz**

| Preferência de velocidade de voz | Porcentagem |
|----------------------------------|-------------|
| Rápido                           | 14,6 %      |
| Muito rápido                     | 0,4 %       |
| Médio                            | 78,2 %      |
| Devagar                          | 6,8 %       |
| Muito devagar                    | 0 %         |

**Gráfico 9 – Preferência de velocidade de voz**



Nos dois extremos, a preferência por vozes femininas prevalece. Das 83 pessoas que preferem uma voz que fala rápido ou muito rápido, 60% escolheram em simultâneo uma voz feminina. Das 38 pessoas que preferem uma voz que fala devagar ou muito devagar, 65% optaram também por uma voz feminina.

A inclinação expressiva por uma voz feminina independente das características vocais de preferência evidencia os efeitos da feminilização de assistentes inteligentes desde sua concepção. Aqui se configura a hipótese de que o gênero da voz, o nome, a personalidade e o conteúdo de assistentes de voz associados ao feminino fomentaram uma expectativa social diante desta tecnologia: um costume se manifesta e a voz feminina em assistentes inteligentes já é esperada por seus usuários.

#### **5.4. Preferências de personalidade e gênero**

Para avaliar as preferências dos respondentes associadas a traços de personalidade e gênero de assistentes de voz, duas perguntas distintas foram aplicadas. A primeira refere-se aos atributos que os participantes preferem para compor a personalidade de um assistente de voz. Foram disponibilizados dez adjetivos dos quais cinco poderiam ser assinalados pelos participantes.

Durante o processo de escolha destes atributos, uma pesquisa realizada pela empresa Telefônica foi utilizada como referência. Apresentada na 32ª Conferência de Interação Humano-computador em julho de 2018, o artigo intitulado *“La personalidad de la Inteligencia Artificial, específicamente de Asistentes Virtuales por voz”*, avaliou a percepção de usuários sobre as personalidades dos assistentes líderes de mercado (Alexa, Siri, Google Assistant e Cortana), com base em métodos de análise e arquétipos de personalidade oriundos do campo da psicologia.

Em conclusão, o estudo (aplicado também no Brasil) identificou os cinco atributos mais votados pelos participantes como indispensáveis na programação de assistentes de voz. A fim de identificar se a tendência constatada pela Telefônica se reafirmaria nesta pesquisa, os adjetivos apresentados aos respondentes do questionário foram alguns dos mais e dos menos votados na pesquisa de referência. Ponderou-se também as questões de gênero e, nesse sentido, foram disponibilizados atributos que sugerem estereótipos sexistas como “racional” e “sentimental” ou “agradável” e “serviçal”.

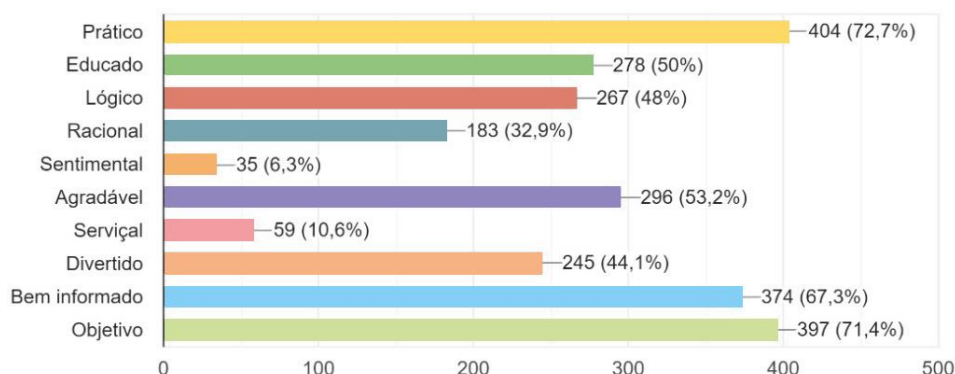
Na pesquisa de referência, os adjetivos mais votados foram “objetivo, confiável, inteligente, bem-informado e prático”. Nesta pesquisa, a tendência se confirma, sendo

os cinco atributos mais assinalados pelos respondentes “prático, objetivo, bem-informado, agradável e educado”. Os menos votados foram “sentimental” e “serviçal”.

#### Gráfico 10 – Preferência de personalidade

Quais dos atributos abaixo você escolheria para compor assistentes de voz inteligentes? (Escolha até 5 opções)

556 respostas



As respostas dos participantes apontam, então, elementos neutros que podem e são comumente interpretados por ambos os sexos. Os adjetivos disponibilizados aos participantes tidos como explicitamente estereotipados, a exemplo de “racional”, “sentimental” e “serviçal”, não representam de modo geral os respondentes do questionário, como mostra o gráfico. Portanto, a priori, não se relacionam as preferências dos respondentes a um gênero, e sim à confirmação de uma predisposição a neutralidade, apresentada também nas questões anteriores, como as preferências vocais.

Em contrapartida, o funcionamento biológico humano, mais especificamente o processo cerebral no qual se desenvolve a distinção de gênero, apoia-se em estereótipos a fim de proteger o organismo do gasto energético “desnecessário” que seria considerar as individualidades de cada pessoa, elaborando assim a categorização de traços e comportamentos e correlacionando-os aos gêneros (Nass; Brave, 2005).

Visto isso, as características escolhidas pelos respondentes deste questionário precisam ser avaliadas com base na premissa de que todos nós aprendemos e guardamos estereótipos de gênero limitantes, e muitas vezes os reproduzimos involuntariamente. De acordo com os autores, estudos apontam que um comportamento agradável,

amigável e educado é geralmente associado a mulheres, enquanto traços de autoritarismo e racionalidade são associados a homens.

Dois dos adjetivos mais escolhidos pelos participantes foram “agradável” e “educado”, e estes estão intimamente relacionados ao estereótipo feminino, como aponta a revisão bibliográfica. Tal resolução reafirma que o gênero da voz exerce uma função norteadora e influencia profundamente na construção de elementos secundários, como a personalidade. Logo, a preferência por vozes femininas em assistentes inteligentes que se confirma também nesta pesquisa, e pode estar associada a estereótipos de gênero, uma vez que, sem perceber, usuários desta tecnologia relacionam uma personalidade agradável e educada ao feminino.

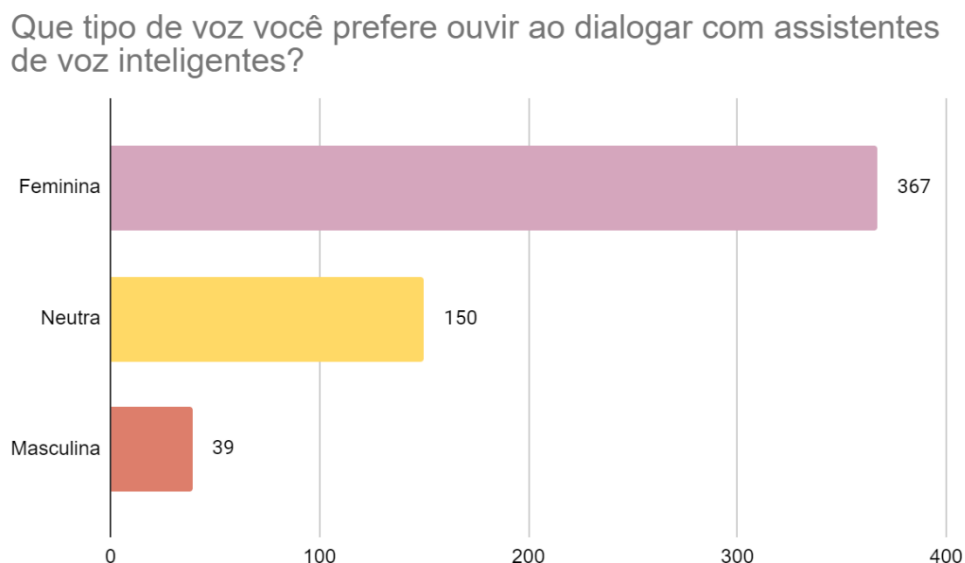
Por outro lado, também é importante ressaltar a neutralidade essencial dos adjetivos escolhidos pelos participantes. Essas preferências sugerem que os respondentes voluntariamente optaram por traços neutros para compor assistentes de voz, ou seja, a atenção da maioria dos participantes para não escolher características explicitamente estereotipadas salienta um passo significativo nesta conjuntura.

Já a segunda pergunta desta etapa é possivelmente a pergunta mais importante do questionário, e trata-se da preferência de gênero da voz de assistentes inteligentes. Uma maioria significativa (66%) revelou preferir uma voz feminina ao dialogar com assistentes de voz. Subsequentemente, 27% dos respondentes manifestaram preferência por uma voz neutra. Apenas 7% dos participantes optaram por uma voz masculina.

**Tabela 10 – Preferência de gênero da voz**

| Preferência de gênero da voz | Porcentagem |
|------------------------------|-------------|
| Feminina                     | 66 %        |
| Neutra                       | 27 %        |
| Masculina                    | 7 %         |

**Gráfico 11 – Preferência de gênero da voz**



Como defendido ao longo da tese, a escolha do gênero da voz de assistentes inteligentes configura uma peça-chave para a recepção da tecnologia pelos usuários, direcionando também outros aspectos intrínsecos à voz, como a personalidade e a emoção. Visto isso, a atribuição de gênero a máquinas interlocutoras deve ser uma escolha cuidadosa e responsável, além de considerar cada desdobramento desta escolha.

A hipótese que se confirma face aos resultados deste questionário são as consequências já evidentes do uso de vozes femininas nas assistentes inteligentes mais utilizadas do mundo. No decorrer dos resultados da pesquisa, a preferência por vozes femininas e a associação de assistentes inteligentes a figuras femininas comprovam esta teoria.

Na pergunta 6, a figura feminina foi 10 vezes mais assinalada que a masculina. Nas perguntas 7, 8 e 9, não há coerência entre as preferências sonoras dos usuários e o gênero da voz, sugerindo que independente das características vocais, a voz feminina é preferida. Na pergunta 10, atributos de personalidade geralmente associados a mulheres estiveram entre os mais escolhidos pelos participantes. Por fim, 66% dos respondentes afirmam preferir vozes femininas ao dialogar com assistentes inteligentes.

Este é um resultado alarmante, embora esperado, considerando todas as expectativas com base em estereótipos de gênero que já foram criadas pelos usuários desta tecnologia desde sua proliferação até os dias atuais.

*“Dominant models of voice computing are crystallizing conceptions of what is ‘normal’ and ‘abnormal’. If the vast majority of AI machines capable of human speech are gendered as young, chipper women from North America (as many are today) users will come to see this as standard. If gendered technologies like Siri and Alexa deflect rather than directly confront verbal abuse (as they do today), users will likely come to see this as standard as well. Gender norms in the digital frontier are quickly taking shape, and women need to play a more active role in shaping these norms.”<sup>40</sup> (West; Kraut; Chew, 2019, p. 123).*

## **5.5. Expectativas futuras**

Para finalizar, duas perguntas foram elaboradas na intenção de compreender as expectativas futuras dos participantes quanto ao uso e características de assistentes de voz inteligentes. Ambas têm apenas duas opções de resposta, e a segunda traz uma terceira opção de “outro”, na qual o participante poderia escrever sua expectativa se assim preferisse.

O resultado da primeira pergunta revela quase uma unanimidade quanto ao uso futuro de assistentes de voz. 98% dos participantes consideram que esta tecnologia será cada vez mais utilizada, enquanto apenas 11 pessoas acreditam no contrário. Um fato interessante sobre esse resultado é que mesmo as pessoas que nunca utilizaram assistentes de voz inteligentes acreditam que a tecnologia está em ascensão.

---

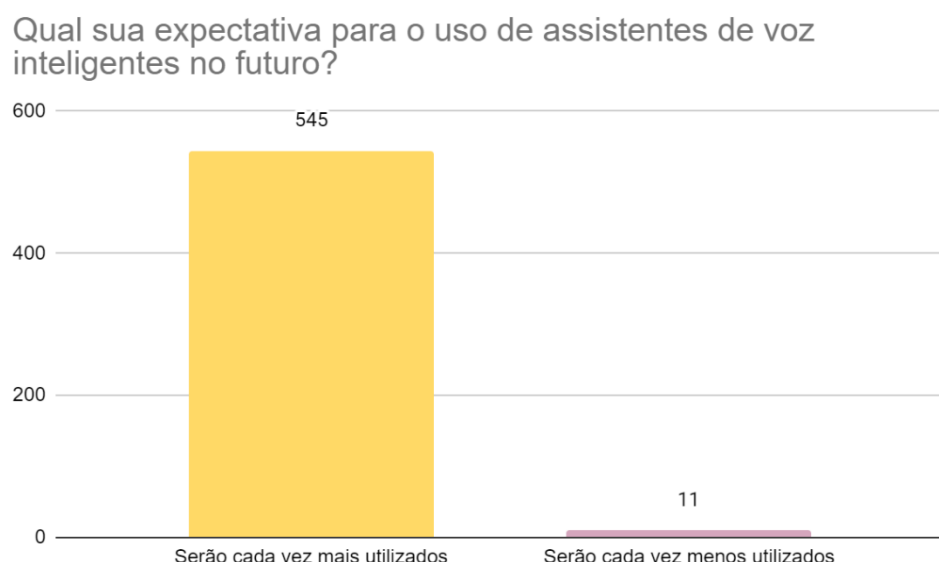
<sup>40</sup> Tradução livre: “Os modelos dominantes de computação de voz são conceitos cristalizantes do que é ‘normal’ e ‘anormal’. Se a grande maioria das máquinas de IA com capacidade de fala humana forem classificadas como jovens, as mulheres mais espertas da América do Norte (como muitas são hoje), os usuários verão isso como um padrão. Se as tecnologias de gênero como Siri e Alexa se desviarem em vez de confrontarem diretamente o abuso verbal (como fazem hoje), os usuários provavelmente também verão isso como um padrão. As normas de gênero na fronteira digital estão tomando forma rapidamente e as mulheres precisam desempenhar um papel mais ativo na definição dessas normas.”

Nesse sentido, é fundamental que se observe no meio acadêmico e empresarial as questões de gênero, personalidade e emoção relacionados as vozes de assistentes inteligentes, bem como a diversidade de gênero, raça e orientação sexual em setores de tecnologia, uma vez que se verificou uma expressiva expectativa para o crescente uso de assistentes de voz no futuro e tal utilização refletirá em diversos aspectos sociais.

**Tabela 11 – Expectativa futura sobre o uso de assistentes de voz**

| Expectativa futura sobre o uso de assistentes de voz | Porcentagem |
|------------------------------------------------------|-------------|
| Serão cada vez mais utilizados                       | 98 %        |
| Serão cada vez menos utilizados                      | 2 %         |

**Gráfico 12 – Expectativa futura sobre o uso de assistentes de voz**



Já na segunda pergunta, a grande maioria (85,3%) acredita que assistentes de voz serão cada vez mais parecidos com seres humanos, e 12,9% dos respondentes pensam que serão cada vez mais parecidos com robôs. O restante (1,8%) optou por desenvolver uma alternativa diferente, das quais algumas se destacam. Um dos respondentes pensa que os assistentes serão cada vez mais personalizáveis; outro acredita no uso de hologramas que saem do celular, mas ainda como uma figura humana; mais de um respondente afirmaram que assistentes de voz serão cada vez mais humanizados, mas não se

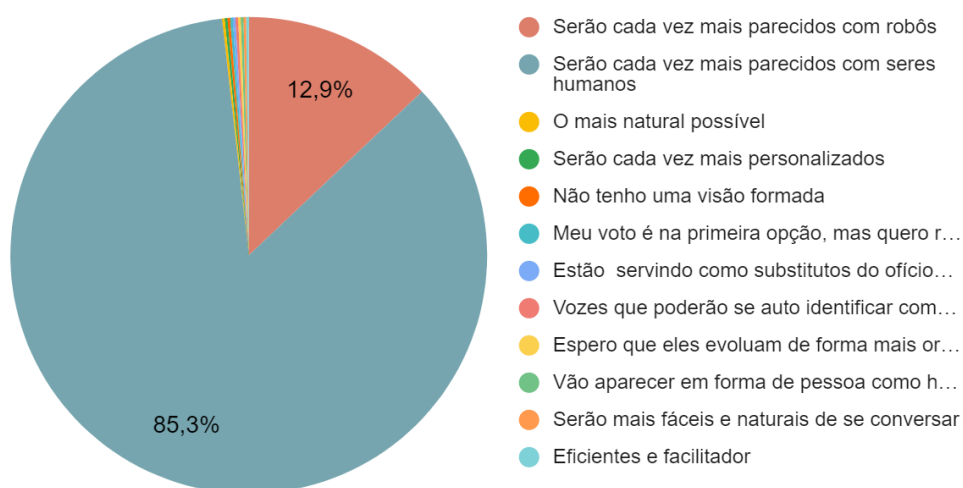
agradam com esta ideia. Um participante também refletiu que no futuro estas vozes podem se identificar como não humanas.

**Tabela 12 – Expectativa futura sobre assistentes de voz**

| Expectativa futura sobre assistentes de voz     | Porcentagem |
|-------------------------------------------------|-------------|
| Serão cada vez mais parecidos com seres humanos | 95,3 %      |
| Serão cada vez mais parecidos com robôs         | 12,9 %      |
| Outros                                          | 1,8 %       |

**Gráfico 13 – Expectativa futura sobre assistentes de voz**

Qual sua expectativa para o futuro de assistentes de voz inteligentes?



Os resultados desta última pergunta são sugestivos, visto que a maioria dos respondentes ao longo do questionário demonstrou pouca ou nenhuma tendência à antropomorfização de assistentes de voz. No entanto, uma significativa maioria espera que assistentes de voz sejam cada vez mais humanizados.

É possível que, inadvertidamente, aspiremos por máquinas interlocutoras semelhantes a seres humanos? Ou diante do reconhecimento de elementos demasiadamente humanos já presentes nestas máquinas (voz, gênero, emoção), reverbera a promessa de evolução e a progressiva adição de novos atributos inicialmente humanos?

## Conclusão

No decorrer desta investigação, conceitos e fundamentos foram identificados de modo a embasar o desenvolvimento e aplicação da pesquisa empírica. A começar pela observação das primeiras teorias aqui citadas, provenientes de áreas de estudo como a Comunicação Humano-máquina e suas adjacentes, como a Interação Humano-computador. Nota-se a urgência de adaptação das teorias da comunicação, considerando os novos papéis interpretados não só pela máquina, mas também pelo ser humano.

A vista deste objetivo, destaca-se um estudo contemporâneo que propõe a reformulação do modelo comunicacional proposto por Shannon (1948). Banks e de Graaf (2020), sugerem um modelo de transmissão “agente-agnóstico”, no qual rejeita-se o papel tradicional dos agentes no modelo de Shannon (emissor, mensagem, meio, receptor) e busca avaliar os integrantes de acordo com suas funções como um meio, para então avaliar quais funções são desempenhadas nos processos de comunicação. Segundo os autores:

*“Machines are designed and understood largely as senders/encoders and receivers/decoders while humans (although indeed important to some processes) are nearly treated as noise or obstacles in computing tasks. Although humans are the progenitor for much machine activity, they are also the component of machine processes whose shortcomings must be muddled through”<sup>41</sup> (Banks; de Graaf, 2020, p. 21).*

A premissa de que máquinas não são apenas meios, mas comunicadoras, e seres humanos e suas ações interferem intimamente no desempenho do meio a ponto de

---

<sup>41</sup> Tradução livre: As máquinas são projetadas e entendidas em grande parte como emissores / codificadores e receptores / decodificadores, enquanto os humanos (embora sejam importantes para alguns processos) são quase tratados como ruídos ou obstáculos nas tarefas de computação. Embora os humanos sejam os progenitores de grande parte da atividade da máquina, eles também são o componente dos processos da máquina cujas deficiências devem ser superadas

agregarem-se a ele, revela a importância de olharmos para estes processos comunicacionais a partir destas possibilidades.

Para tanto, o conhecimento ainda que mínimo da natureza técnica das máquinas faz-se essencial. A revisão teórica acerca de inteligências artificiais e seus conceitos básicos, bem como a posterior explicação metódica de sistemas de diálogo falado contribuíram pontualmente para a construção das questões que resultaram na pesquisa prática desta investigação. Conhecer os componentes da tecnologia que afigura o objeto de estudo possibilitou a concepção de reflexões que dependem diretamente deste entendimento.

Fez-se necessário compreender que existem duas formas mais difundidas de construir uma inteligência artificial e que, apesar de ambas resultarem em algum nível de automação, seus processos são distintos e interferem nos resultados. Um assistente de voz inteligente, por exemplo, configura uma inteligência artificial de aprendizado de máquina. Como visto anteriormente, isto significa dizer que um montante de dados será inserido neste processo para que a máquina aprenda com estes dados e produza um determinado efeito.

Este sistema de funcionamento, por si só, conjectura ponderações acerca dos conteúdos proferidos por assistentes de voz. A procedência dos dados imputados para que a máquina aprenda implica questões éticas e sociais, pois a utilização de dados não-filtrados negligencia a probabilidade de reprodução de conteúdos racistas, sexistas, homofóbicos e diversas disfunções sociais propagadas pela internet. Caso assistentes de voz utilizassem a manipulação de símbolos, a atenção para essas questões seria direcionada aos programadores e suas respectivas intenções com os comandos alocados, ao invés dos dados inseridos no aprendizado de máquina.

As especificidades técnicas atuam ativamente nos efeitos e repercussões de determinada tecnologia e, portanto, a falta desse conhecimento pode gerar conclusões equivocadas baseadas no subjetivismo acerca de questões práticas e, neste segmento de inteligências artificiais especialmente, movidas por preocupações coletivas potencializadas por publicações sensacionalistas e ficções científicas.

A pertinência de explicar a organização técnica de sistemas de diálogo falado justifica-se pelo mesmo princípio. Uma vez que o progresso desta tecnologia tem se desenvolvido a partir do aperfeiçoamento de sistemas já existentes, como é o caso do Google Duplex previamente exposto, saber as diferenças e atribuições de cada técnica que compõe estes sistemas viabiliza o entendimento também do panorama e circunstâncias em que se encontra a ferramenta. Exemplificando, a incompreensão dessas estruturas poderia instigar a ideia de que uma inteligência artificial geral estaria próxima devido a apresentação do Google.

A partir destes entendimentos, outras urgências concebidas pelas trocas entre seres humanos e máquinas foram assinaladas, cuja ênfase se direcionou ao comportamento conversacional humano e sua significação decorrente. A ideia de que seres humanos tendem a tratar computadores como se fossem pessoas e lugares reais (Nass; Reeves, 1996) desdobrou-se em uma cognição interdisciplinar. Descendente da psicologia e da neurociência, esta teoria considera o funcionamento cerebral para pautar seus experimentos e implicações, e por esse motivo, as conclusões realizadas ao longo da tese também se fundamentaram na convergência dos aspectos sociais e biológicos.

Nesse sentido, as reflexões que se iniciaram no campo da Interação Humano-computador conectaram-se com áreas de conhecimento como estudos sociais e neurociência, e sugerem que a forma como interpretamos e interagimos com as tecnologias em ascensão são potencialmente capazes de promover transformações culturais (Eagleton, 2000) e por consequência, possíveis mutações genéticas (Damásio, 2011).

Como forma de avaliar esta interpretação e interação que se desdobra das relações humano-máquina iminentemente transformadoras, em um recorte tecnológico específico, os efeitos do uso de assistentes de voz foram avaliados em diversas perspectivas. A priori, a revisão teórica suscitou um levantamento de tópicos pertinentes para debate no meio acadêmico, e serão listadas aqui algumas das conclusões.

Da perspectiva funcional, a evolução da “internet das coisas” e a hiper conectividade gerada por ela se relaciona intimamente com sistemas de diálogo falado, uma vez que os ambientes estão sendo progressivamente ocupados por diversos dispositivos que se conectam e são controlados por voz através de um assistente inteligente. Nesse sentido, a forma como atualmente estamos construindo vínculos com estas máquinas, principalmente as mais antropomorfizadas, ditarão o futuro dos ecossistemas comunicacionais.

Sobre os receios quanto a privacidade e segurança, verifica-se uma constante repetição de falhas – como vazamento de dados – por parte das empresas desenvolvedoras de assistentes pessoais, gerando uma crescente preocupação em seus usuários. Embora os números ainda não apontem para uma resistência a esta tecnologia movida por estas questões, a recorrência pode desincentivar seu uso, além das incontáveis problemáticas que podem emergir da quebra de privacidade. Portanto, a ação do Estado faz-se essencial nesta conduta, atuando ativamente na fiscalização do cumprimento dos termos de uso relacionados a proteção de dados e segurança. Todavia, a maior responsabilidade deve ser atribuída às empresas desenvolvedoras, já que estas têm o poder tecnológico e o dever ético e social de aprimorar seus processos de segurança.

Do ponto de vista da usabilidade, todas as pesquisas reunidas aqui evidenciaram uma expressiva progressão no uso de assistentes pessoais e sua disseminação global. Apesar dos obstáculos no uso desta tecnologia (Motta; Quaresma, 2019), sua aceitação no mercado mundial é inegável e as projeções futuras apontam para um cenário ainda mais permeado por máquinas interlocutoras com múltiplas finalidades. Neste tópico, ênfase é necessária no contexto de acessibilidade: pesquisas recentes na área constataam a capacidade destes dispositivos na assistência de pessoas com as mais variadas limitações físicas e cognitivas (Bellei, 2021).

Estas conclusões elaboradas até então contemplam as reverberações palpáveis acerca de assistentes de voz – em outras palavras, suas questões amplamente discutidas. A seguir serão desenvolvidas as conclusões resultantes da observação analítica a respeito de tópicos pouco discutidos na esfera social e acadêmica: as conceitualizações e significações abstratas decorrentes do uso desta tecnologia.

A voz configurou o elemento norteador destas observações, isto porque sua natureza primitivamente humana, por suposto, maximiza a mobilização de respostas sociais; quanto maior a capacidade de simular comportamentos humanos, maior a ilusão de que se está diante de outra pessoa ao interagir com máquinas inteligentes (Fortunati; Edwards, 2020). Em prol desse objetivo, considerou-se as características da voz sugeridas por Nass e Brave (2005) devido a relevância comprovada destes traços em seus experimentos, sendo estes: gênero, personalidade e emoção atribuídos à voz.

Com base nestas características, elaborou-se o questionário online para confirmar ou refutar as premissas defendidas pelos autores. É importante ressaltar que estes estudos foram realizados há mais de 15 anos, em um contorno geográfico distinto (Estados Unidos da América), e desde então o cenário de tecnologias de voz passou por complexas transformações. Ainda assim, suas conclusões mostram-se coerentes com as circunstâncias que se desenrolaram ao longo destes anos, e ainda hoje embasam novas pesquisas na área.

A fim de observar as interpretações e conceitualizações dos brasileiros acerca de assistentes pessoais, bem como avaliar a influência da voz nesse processo de significância, o questionário foi aplicado e respondido por 556 pessoas ao redor do país. As conclusões serão aqui resumidas e pontuadas:

- Os respondentes, em sua maioria, descreveram assistentes de voz inteligentes como “programa de computador” e “inteligência artificial”, demonstrando uma conceitualização tecnológica destes sistemas.
- Quando tiveram que escolher uma imagem que se relacionasse à sua idealização de assistentes de voz, os participantes dividiram-se entre a imagem de servidores, e as imagens de figuras robóticas, reforçando a conceitualização tecnológica da questão anterior e, em contrapartida, evidenciando que esta concepção não é uniforme na mente das pessoas.
- Das características vocais trabalhadas nas questões (gênero, tom, volume, velocidade e traços de personalidade), a que mais demonstrou exercer influência sobre a percepção e preferência dos respondentes foi o gênero. A definição do gênero da voz sugestiona também as preferências de

personalidade, uma vez que há uma correlação social entre comportamentos e gêneros.

- O fato de assistentes de voz inteligentes mais populares serem feminilizadas desde sua estreia no mercado induz uma associação desta tecnologia ao sexo feminino. A maioria dos participantes prefere ouvir uma voz feminina ao dialogar com assistentes pessoais, independentemente de suas preferências vocais e personalidade. A imagem da mulher disponibilizada em uma das perguntas do questionário foi 10 vezes mais votada que a imagem do homem.
- Há um consenso na expectativa futura sobre o uso de assistentes inteligentes: quase todos os participantes responderam acreditar que esta tecnologia será cada vez mais utilizada, em conformidade com as projeções de mercado.
- Embora não tenha sido identificado um processo de antropomorfização de assistentes pessoais, 85,3% dos respondentes acreditam que esta tecnologia será cada vez mais humanizada no futuro. Ou seja, apesar da conceitualização tecnológica revelada nas primeiras perguntas, há uma expectativa de humanização destas máquinas.
- Por fim, das pessoas que nunca utilizaram assistentes de voz inteligentes, 65% são mulheres, solidificando a teoria da lacuna de habilidades digitais entre gêneros apontada pelo estudo realizado pela UNESCO.

De modo a contrariar a hipótese inicialmente prevista, os resultados da pesquisa empírica não apontam para a antropomorfização de assistentes de voz por parte dos usuários brasileiros. Outra revelação inesperada caracteriza a incontestável relevância do gênero da voz de assistentes inteligentes, tanto como influência na percepção e conceitualização por parte dos usuários, quanto para a definição de elementos secundários como personalidade, conteúdo e emoção.

As questões de gênero observadas nesta investigação encontram fundamentos e esclarecimentos no conceito de “tecnofeminismo” proposto por Judy Wajcman (2004). Sob o argumento de que a tecnologia é uma fonte de poder para os homens, bem como uma característica definidora de masculinidade, a autora propõe uma revisão histórica

do domínio masculino sobre as máquinas e a sub-representação feminina nos campos científicos e tecnológicos.

O artigo publicado pela Unesco, anteriormente citado, desenvolve esta perspectiva pontualmente voltada para a feminilização de assistentes de voz, associando-a à crise de diversidade nos setores de inteligência artificial e tecnologia das empresas desenvolvedoras. Como sustentação, Wajcman defende que “a tecnologia é tanto uma fonte como uma consequência das relações de gênero” (Wajcman, 2004, p. 164).

Nass e Brave (2005) reafirmam esta alegação diante dos resultados de experimentos acerca da reprodução de estereótipos de gênero nas relações humano-máquina. Segundo os autores, da mesma forma que expectativas de gênero são *projetadas na* tecnologia, elas podem ser *esculpidas pela* tecnologia. Eles sugerem a inserção de vozes do gênero “atípico” de determinada função. Por exemplo, se há uma expectativa de gênero associada ao feminino quanto à função de assistência, o ideal seria utilizar vozes masculinas para que a presença do gênero atípico sobressaia. Para que não seja uma mudança brusca e gere estranhamento, os autores sugerem ainda a construção de vozes femininas com traços masculinos e vice-versa.

Em todo caso, é concreto o poder transformador da tecnologia quanto às estruturas sociais, políticas e econômicas e, portanto, faz-se imprescindível que sua construção e desenvolvimento abarquem as demandas e singularidades da sociedade como um todo. As empresas desenvolvedoras de inteligências artificiais e tecnologias de modo geral, comprovadamente integradas em expressiva maioria por homens brancos, não se beneficiam da inclusão feminina e de outras minorias nestes setores, visto que o domínio destas esferas caracteriza poder.

Esta tese representa uma ação em prol do debate científico destas pautas, e dessa forma, conclui-se que é urgente a instituição de políticas públicas que promovam o acesso das minorias às tecnologias que transfiguram a sociedade e, para além do acesso, viabilizem o desenvolvimento de habilidades digitais, bem como o incentivo e inclusão de mulheres, negros, indígenas, pessoas LGBTQIA+ nas indústrias tecnológicas. Soluções técnicas como a inserção de uma outra opção de voz em assistentes inteligentes não

irão solucionar a problemática originária da desigualdade de gênero e social que afiguram o desigual cenário científico e tecnológico.

## Referências Bibliográficas

BANKS, J.; DE GRAAF, M. **Toward an Agent-Agnostic Transmission Model: Synthesizing Anthropocentric and Technocentric Paradigms in Communication.** Human-Machine Communication Journal. 1, 19-36, 2020. Disponível em: <<https://doi.org/10.30658/hmc.1.2>>

BARBOSA, S.; SILVA, B. **Interação Humano-computador.** Rio de Janeiro: Elsevier, 2010. ISBN 9788535234183.

BARROS, José D. **Os conceitos: seus usos nas ciências humanas.** Rio de Janeiro: Editora Vorazes, 2016. ISBN 9788532652638.

BELLEI, Ercles A. **Acessibilidade e assistentes de voz: da pesquisa científica até a indústria.** Brasil: SBC Horizontes, 2021, atual. [Consult. 20 Set. 2021]. Disponível em: <<https://horizontes.sbc.org.br/index.php/2021/03/acessibilidade-e-assistentes-de-voz/>>

BENJAMIN, Garfield. **Amazon Echo's privacy issues go way beyond voice recordings.** Estados Unidos: The Conversation, 2020, atual. [Consult. 20 Set. 2021]. Disponível em: <<https://theconversation.com/amazon-echos-privacy-issues-go-way-beyond-voice-recordings-130016>>

BUBAS, G. **Computer mediated communication theories and phenomena: Factors that influence collaboration over the Internet.** 3<sup>rd</sup> CARNET Users Conference. Zagreb, 2001. Disponível em: <<https://people.utm.my/arahim/wp-content/blogs.dir/894/files/2018/03/CMC-THEORIES-AND-PHENOMENA-FACTORS-INFLUENCE-COLLABORATION-OVER-THE-INTERNET.pdf>>

DAMASIO, António R. **E o cérebro criou o homem.** São Paulo: Companhia das Letras, 2011. ISBN 9788535919615.

DAMASIO, António R. **O erro de Descartes: emoção, razão e o cérebro humano.** São Paulo: Companhia das Letras, 2012. ISBN 9788535922004.

EAGLETON, Terry. **The idea of culture.** Oxford: Blackwell Publishing, 200. ISBN 9780631219668.

EISENSTEIN, Elizabeth. **A revolução da cultura impressa: os primórdios da Europa Moderna**. São Paulo: Ática, 1998. ISBN 9788508068999.

EISHIMA, Rubens. **Uso de assistentes de voz cresceu 47% no Brasil durante pandemia, indica estudo**. Brasil: Canal Tech, 2020, atual. [Consult. 20 Set. 2021]. Disponível em: <<https://canaltech.com.br/inteligencia-artificial/uso-de-assistentes-de-voz-cresceu-47-no-brasil-durante-pandemia-indica-estudo-172657/>>

FORTUNATI, L.; EDWARDS, A. **Opening Space for Theoretical, Methodological, and Empirical Issues in Human-Machine Communication**. Human-Machine Communication Journal. 1, 7-18, 2020. Disponível em: <<https://doi.org/10.30658/hmc.1.1>>

GARCÍA, M. P.; LOPEZ, S. S.; DONIS, H. **La personalidad de la Inteligencia Artificial, específicamente de Asistentes Virtuales por voz**. 32nd Human Computer Interaction Conference, 2018. Disponível em: <<https://doi.org/10.14236/ewic/HCI2018.40>>

GLEICK, James. **A informação: uma história, uma teoria, uma enxurrada**. Rio de Janeiro: Companhia das Letras, 2013. ISBN 9788580867114.

**GREENIA, Mark W.** History of computing: an encyclopedia of the people and machines that made computer history. **California: Lexikon Services, 1998.**

GUNKEL, David J. **An introduction to communication and artificial intelligence**. Medford, MA: Polity Press, 2020. ISBN 9781509533190.

Guzman, Andrea L. **Voices in and of the machine: Source orientation toward mobile virtual assistants**. Computers in Human Behavior. 90, 343-350, 2019. DOI: 10.1016/J.CHB.2018.08.009

GUZMAN, Andrea L. **Voices in and of the machine: source orientation toward mobile virtual assistants**. Computers in Human Behavior, 90, 343-350, 2018. DOI: 10.1016/J.CHB.2018.08.009.

GUZMAN, Andrea L., editor. **Human-machine communication: rethinking communication, technology, and ourselves**. New York: Peter Lang, 2018. ISBN 9781433142529.

HARARI, Yuval Noah. **Homo Deus: uma breve história do amanhã**. São Paulo: Companhia das Letras, 2016. ISBN 9788543807829.

HOY, Matthew B. **Alexa, Siri, Cortana, and more: an introduction to voice assistants**. Medical Reference Services Quarterly, 37:1, 81-88, 2018. DOI: 10.1080/02763869.2018.1404391.

JOKINEN, K.; MCTEAR, M. **Spoken Dialogue Systems**. Synthesis Lectures on Human Language Technologies. 2:1, 1-151, 2010. DOI: 10.2200/S00204ED1V01Y200910HLT005.

KRAUS, Michael W. **Voice-Only Communication Enhances Empathic Accuracy**. American Psychologist. 72:7, 644-654, 2017. DOI: 10.1037/AMP0000147.

LEVIATHAN, Y.; MATIAS, Y. **Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone**. Google AI Blog, 2018, atual. [Consult. 20 Set. 2021]. Disponível: <<https://ai.googleblog.com/2018/05/duplex-ai-system-for-natural-conversation.html>>

LICKLIDER, J.; CLARK, W. **Online Man-computer Communication**. Managing Requirements Knowledge. 1, 113-128, 1962. DOI: 10.1145/1460833.1460847.

MARCONI, M. A.; LAKATOS, E. M.; MEDEIROS, J. B. **Técnicas de pesquisa**. 9. Ed. São Paulo: Atlas, 2021. ISBN 9788597026603.

MCCARTY, John. [et. al]. - **A proposal for the Dartmouth Summer Research Project on Artificial Intelligence**. Dartmouth: 1956.

MCLUHAN, Marshall. **A galáxia de Gutenberg: a formação do homem tipográfico**. São Paulo: Companhia Nacional, 1972.

MCLUHAN, Marshall. **Os meios como extensões do homem**. 15. Ed. São Paulo: Editora Cultrix, 2007.

MOTTA, I.; QUARESMA, M. **Barreiras no uso de assistentes de voz por usuários brasileiros de smartphone**. Anais do 17º Congresso Internacional de Ergonomia e Usabilidade de Interfaces Humano-Tecnologia e o 17º Congresso Internacional de Ergonomia e Usabilidade de Interfaces e Interação Humano-Computador, 6:6, 454-466, 2019.

NASS, C.; BRAVE, S. **Wired for speech: how voice activates and advances the human-computer relationship**. Massachusetts: The MIT Press, 2005. ISBN 9780262140928.

NASS, C.; STEUER, J. **Voices, Boxes, and Sources of Messages**. Human Communication Research. 19:4, 504-527, 1993. DOI: 10.1111/J.1468-2958.1993.TB00311.X.

OLSON, Christi. **New report tackles tough questions on voice and AI**. Microsoft Blog, 2019, atual. [Consult. 20 Set. 2021] Disponível em: <<https://about.ads.microsoft.com/en-us/blog/post/april-2019/new-report-tackles-tough-questions-on-voice-and-ai>>

PAMPLONA, Nicola. **Segundo IBGE, 43 milhões de estudantes brasileiros entraram na pandemia sem acesso à internet**. São Paulo: Folha de São Paulo, 2021. Disponível em:

PERLES, João B. **Comunicação: conceitos, fundamentos e história**. Biblioteca Online de Ciências da Comunicação, 2007. ISSN 16463137. Disponível em: <<http://www.bocc.ubi.pt/pag/perles-joao-comunicacao-conceitos-fundamentos-historia.pdf>>

REEVES, B.; NASS, C. **The Media Equation: how people treat computers, television and new media like real people and places**. Stanford, CA: CSLI Publications, 1996. ISBN 157586052X.

RÜDIGER, Francisco. [et. at.] - **Teorias da Comunicação no Brasil: reflexões contemporâneas**. Salvador: Edufba, 2014. ISBN 9788523212339.

SCHWÄR, H.; MOYNIHAN, Q. **Companies like Amazon may give devices like Alexa female voices to make them seem 'caring'**. Alemanha: Business Insider, 2020, atual. [Consult. 20 Set. 2021]. Disponível em: <<https://www.businessinsider.com/theres-psychological-reason-why-amazon-gave-alexa-a-female-voice-2018-9>>

SERRA, J. Paulo. **Manual de Teoria da Comunicação**. Covilhã: Livros LabCom, 2007. ISBN 9789728790875.

SHANNON, Claude E. **A mathematical theory of communication**. Bell System Technical Journal. 27 (3): 379-423, 1948. DOI: 10.1002/J.1538-7305.1948.TB01338.X.

SKANTZE, Gabriel. **Error Handling in Spoken Dialogue Systems: Managing Uncertainty, Grounding and Miscommunication.** Suécia: KTH Computer Science and Communication, 2007.

TASCA, Nino. **Doing more to protect your privacy with the Assistant.** Google Blog, 2019, atual. [Consult. 20 Set. 2021]. Disponível em: <<https://blog.google/products/assistant/doing-more-protect-your-privacy-assistant/>>

TURING, Alan. **Computing machinery and intelligence.** Mind: a quarterly review of psychology and philosophy. 59:236, 433-460, 1950. Disponível em: <<https://phil415.pbworks.com/f/TuringComputing.pdf>>

VAILSHERY, Lionel S. **Number of digital voice assistants in use worldwide 2019-2024.** Statista, 2020, atual. [Consult. 20 Set. 2021]. Disponível em: <<https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/>>

WAJCMAN, Judy. **Technofeminism.** Reino Unido: Polity Press, 2004. ISBN 9780745630441.

WARREN, Tom. **Apple has Siri and Microsoft is about to get Cortana.** Nova York: The Verge, 2014, atual. [Consult. 20 Set. 2021]. Disponível em: <<https://www.theverge.com/2014/2/20/5430188/microsoft-cortana-personal-digital-assistant-windows-phone-8-1>>

WEIZENBAUM, Joseph. **ELIZA – a computer program for the study of natural language communication between man and machine.** Communications of the ACM. 9:1, 36-45, 1966. Disponível em: <<https://web.stanford.edu/class/linguist238/p36-weizenbaum.pdf>>

WEST, M.; KRAUT, R.; CHEW, H. **I'd blush if I could: closing gender divides in digital skills through education.** Equals and Unesco, 2019. Disponível em: <<https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1>>

WEST, S.M.; WHITTAKER, M.; CRAWFORD, K. **Discriminating Systems: Gender, Race and Power in AI.** Nova York: AI Now Institute, 2019. Disponível em: <<https://ainowinstitute.org/discriminatingsystems.html>.>