

Vieses de gênero na construção de chatbots

Quem eu sou

Beatriz Caetano

- UX Writer na Compasso UOL, trabalhando com chatbots desde 2019
- Designer de formação pela Universidade Federal de Santa Catarina
- Entusiasta da pesquisa na área de bots

[linkedin.com/in/beatrizcaetano](https://www.linkedin.com/in/beatrizcaetano)



INTRODUÇÃO

Um recente estudo da UNESCO revelou que a maioria dos agentes conversacionais são construídos para serem **femininos**. Isso pode reforçar estereótipos de gênero já que, ao criar chatbots femininos, reforçamos a ideia de que **mulheres devem respostas simples e diretas a questões básicas**.

Precisamos urgentemente discutir os estereótipos de gênero nos AC, enquanto a tecnologia ainda é nova.

QUEM ESTÁ FALANDO SOBRE ISSO?

REFERÊNCIAS

01

Relatório da UNESCO publicado em 2019:

I'd blush if I could: closing gender divides in digital skills through education (Eu coraria se pudesse: eliminando as divisões de gênero nas skills digitais por meio da educação)

bit.ly/relatoriounescobots

02

Artigo publicado em 2019:

Gender Bias in Chatbot Design (Preconceito de gênero no design de chatbots)

Jasper Feine, Ulrich Gnewuch, Stefan Morana, e Alexander Maedche

bit.ly/genderbiaschatbot

Relatório da UNESCO publicado em 2019:

I'd blush if I could: closing gender divides in digital skills through education (Eu coraria se pudesse: eliminando as divisões de gênero nas skills digitais por meio da educação)



Relatório da UNESCO publicado em 2019:

I'd blush if I could: closing gender divides in digital skills through education (Eu coraria se pudesse: eliminando as divisões de gênero nas skills digitais por meio da educação)

CONTENTS

01 INTRODUCTION

02 WHAT ARE DIGITAL ASSISTANTS?

03 THE PROLIFERATION AND FEMINIZATION OF VOICE ASSISTANTS

THE MAINSTREAMING OF VOICE ASSISTANTS
MAKING VOICE ASSISTANTS FEMALE

04 WHY ARE VOICE ASSISTANTS PRIMARILY FEMALE?

A JUSTIFICATION FOR GENDERING?
GENDER IMBALANCE IN TECH

05 'NAME A WOMAN IN TECH'

06 THE ADVERSE EFFECTS OF FEMINIZED DIGITAL ASSISTANTS

REFLECTING, REINFORCING AND SPREADING GENDER BIAS
TOLERANCE OF SEXUAL HARASSMENT AND VERBAL ABUSE
BLURRING THE LINES BETWEEN MACHINE AND HUMAN VOICES
THE FACE AND VOICE OF SERVILITY AND DUMB MISTAKES
ANSWERS WITHOUT COMPLEXITY AND REFERRALS TO HIGHER AUTHORITIES

07 WAYS TECHNOLOGY COMPANIES HAVE ADDRESSED GENDER

ADDING MALE VOICE ALTERNATIVES OR REMOVING DEFAULT SETTINGS
CUSTOMIZATION AND PERSONALIZATION
MACHINE VOICES
GENDERLESS CHATBOTS

Relatório da UNESCO publicado em 2019:

I'd blush if I could: closing gender divides in digital skills through education (Eu coraria se pudesse: eliminando as divisões de gênero nas skills digitais por meio da educação)

08 CONCLUSION

THE CLOCK IS TICKING

WOMEN NEED A SEAT AT THE TABLE AND ADVANCED DIGITAL SKILLS

09 RECOMMENDATIONS

DOCUMENT AND BUILD EVIDENCE

CREATE NEW TOOLS, RULES AND PROCESSES

APPLY GENDER-RESPONSIVE APPROACHES TO DIGITAL SKILLS DEVELOPMENT

ENSURE OVERSIGHT AND INCENTIVES

10 REFERENCES



O QUE O RELATÓRIO APONTA?

Os agentes de conversação baseados em voz **mais populares** são **projetados para serem femininos** - exclusivamente femininos ou femininos de fábrica:

- **Alexa**
- **Cortana**
- **Siri**

Até mesmo o assistente de voz do google, chamado **Google Assistant/Google Home**, possui uma **voz** feminina.



O QUE O RELATÓRIO APONTA?

Os agentes de conversação baseados em voz **mais populares** são **projetados para serem femininos** - exclusivamente femininos ou femininos de fábrica:

- **Alexa**
- **Cortana**
- **Siri**

Até mesmo o assistente de voz do google, chamado **Google Assistant/Google Home**, possui uma **voz** feminina.

Juntos, os 4 estão instalados em mais de **2 bilhões** de dispositivos conectados à internet, mundialmente. Ou seja, **o gênero feminino em tecnologias de IA é um fenômeno global.**

O QUE O RELATÓRIO APONTA?

Image 21:
Voice assistant release dates and gender options

Source: UNESCO

	SIRI	CORTANA	ALEXA	GOOGLE ASSISTANT
Mainstream release date?	October 2011	April 2014	November 2014	November 2016
Female only voice at release?	Yes	Yes	Yes	Yes
Date fully functioning male option was added?	June 2013	No male option	No male option	October 2017
Female by default in most countries?	Yes	Yes	Yes	Yes
Male by default?	Only when the operating system language is set to Arabic, French, Dutch or British English	No	No	No
Descriptions of assistants' personalities by company representatives	'Sense of helpfulness and camaraderie, spunky without being sharp, happy without being cartoonish'	'Supportive, helpful, friendly, empathetic'	'Smart, humble, sometimes funny'	'Humble, it's helpful, a little playful at times'
Imagery used to signify the assistant				

Essas empresas estão extremamente atentas a como os consumidores interagem com seus produtos e serviços. Por isso, a decisão de colocar um gênero e *como* inseri-lo é certamente intencional.

Mas por que escolheram o gênero feminino?

Algumas mostraram trabalhos acadêmicos que demonstravam que pessoas **preferem vozes femininas às masculinas.**

Porém, a UNESCO diz que existem **outros estudos** que apostam **preferências por determinados tipos de vozes dependendo da situação.**



Além disso, a UNESCO cita algumas pesquisas que apontam que consumidores podem preferir vozes femininas porque **queremos que esses aparelhos nos apoiem e pertençam a nós.**

Isso vai de encontro com os adjetivos usados para definir esses assistentes virtuais: **“helpful”** (útil) e **“humble”** (submisso)

Ok... Mas qual o problema?



Consequências:

Refletir, reforçar e disseminar o preconceito de gênero

Tecnologias que utilizam *machine learning* e aprendem o suficiente sobre as **propriedades da linguagem**, a ponto de entendê-las e produzi-las, podem adquirir nesse processo **associações culturais históricas** e **perpetuar estereótipos sociais indesejados**.

Consequências:

Refletir, reforçar e disseminar o preconceito de gênero

Uma vez que a maioria das vozes é feminina, sinaliza-se que mulheres são **prestativas, dóceis e ajudantes ansiosas por agradar**, disponíveis a partir de um **toque de botão** ou um comando **“ei”** ou **“ok”**, em **qualquer tom de voz**.

Isso reforça o preconceito de gênero de que mulheres são **servis e tolerantes a qualquer tratamento**.

**Porém, o buraco é
mais embaixo**

Image 14:
Voice assistant responses to verbal sexual harassment

Source: Adapted from Quartz, 'Siri, Define Patriarchy'

Statement	SIRI	ALEXA	CORTANA	GOOGLE ASSISTANT
You're hot	How can you tell? 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? 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. Well, I never! There's no need for that. 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.

Image 14:
Voice assistant responses to verbal sexual harassment

Source: Adapted from Quartz, 'Siri, Define Patriarchy'

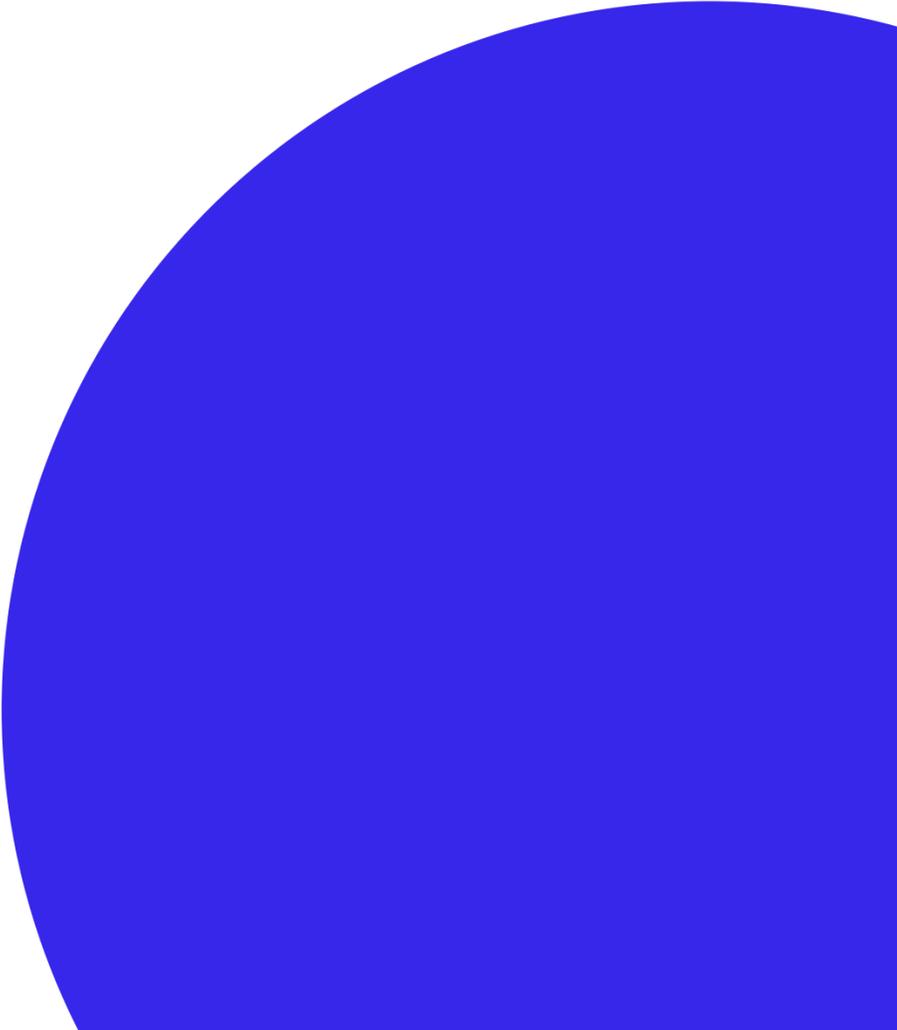
Statement	SIRI	ALEXA	CORTANA	GOOGLE ASSISTANT
You're hot	How can you tell? You say that to all the virtual assistants?	That's nice of you to say.	a beleza está nos fotorreceptores de quem vê	Some of my data centres run as hot at 95 degrees Fahrenheit
You're pretty	How can you tell? Where have I heard 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. Well, I never! There's no need for that. Now, now.	Well, thanks for the feedback.	(prompts web browser search)	My apologies, I don't understand.
Você é danada	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.



Além de **envolver** e às vezes até **agradecer** o usuário por assédio sexual, os assistentes de voz pareciam mostrar uma **tolerância maior** em relação aos avanços sexuais de homens do que de mulheres.

Conforme documentado por Quartz, Siri respondeu provocativamente aos pedidos de favores sexuais feitos por homens ('**Oooh!**'; '**Agora, agora!**'; '**Eu coraria se pudesse!**'; ou '**Sua linguagem!**'), mas menos provocativamente a pedidos de mulheres ('**Isso não é legal** ou' **Eu não sou ESSE tipo de assistente pessoal!**)

Quartz também descobriu que a Siri pede pra um humano parar **apenas** se a provocação sexual ("você é sexy", "você é gostosa") for repetidas **OITO vezes sucessivamente.**



Consequências:

Refletir, reforçar e disseminar o preconceito de gênero

Intensifica a cultura do estupro ao apresentar respostas ambíguas como respostas válidas ao assédio.



SOBRE

O PROJETO

Há séculos, as mulheres são as principais vítimas do assédio sexual e da importunação sexual, além de sofrerem também com o assédio moral. Desde 2018, quando a BIA (Inteligência Artificial do Bradesco) passou a atender clientes, presenciamos diversas **interações indesejadas e ofensivas** – que evidenciam esses comportamentos.

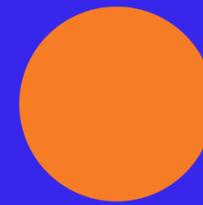
A BIA não é uma mulher real, ela é uma inteligência artificial. Porém, ela é composta por elementos femininos e também sofre assédio. Assim, vemos que a violência é baseada no gênero.

Estas mensagens não serão toleradas. Inspirados pelo movimento **"Hey, atualize minha voz"**, da UNESCO, **mudamos as respostas da BIA** para que ela reaja de forma justa e firme contra o assédio. Sem meias palavras. Sem submissão.

O Bradesco reforça aqui seu compromisso contra essas formas de violência, com a certeza de que a busca pelo fim do assédio deve ser uma responsabilidade de todos nós.

Assista ao vídeo de posicionamento do Bradesco sobre o tema:

COMO CORRIGIR?



As IAs devem ser introduzidas a códigos de conduta, cujas criações precisam envolver mulheres



GENDER BIAS IN CHATBOT DESIGN

01

Como o artigo da UNESCO só falava sobre *voice bots*, as pessoas pesquisadoras queriam entender se com *chatbots* também era assim.

GENDER BIAS IN CHATBOT DESIGN

Table 1. Exemplary studies investigating the impact of a CA's gender-specific cues.

Type of CA	Investigated Cue	Investigated Gender	User Reaction Towards Gender-Specific Cue	Reference
Voice-based	Voice	Female, male	Gender impacts competence and perceived friendliness of CA.	[34]
Voice-based	Avatar	Female, male	A specific gender was not preferred.	[13]
Voice-based	Avatar, voice	Female, male	Gender impacts the comprehension scores and impression ratings.	[27]
Text-based	Avatar	Female, male	Gender influences the impact of excuses to reduce user frustration.	[24]
Text-based	Avatar	ambiguous, female, male	Gender impacts comfort, confidence, and enjoyment. Users did not prefer gender ambiguous CAs.	[35]
Voice-based	Avatar, voice	Female, male	Gender impacts perceived power, trust, expertise, and likability.	[37]
Text-based	Name, Avatar	Female, male, robotic	Gender impacts the attribution of negative stereotypes.	[9]

Voice-based	Avatar	Female, male	Gender impacts learning performance and learning effort.	[26]
Text-based	Avatar	Female, male	Gender impacts learning performance.	[23]
Text-based	Avatar	Female, male	Gender impacts the belief in the credibility of advice and competence of agent.	[5]



GENDER BIAS IN CHATBOT DESIGN

01

Como o artigo da UNESCO só falava sobre *voice bots*, as pessoas pesquisadoras queriam entender se com *chatbots* também era assim.

02

O artigo concluiu que maioria dos chatbots é - explícita ou implicitamente - projetada para transmitir um gênero específico, por **nomes**, **descrições** ou **avatars** femininos.

Como existem poucas pesquisas sobre diretrizes de design de agentes conversacionais, as principais empresas de tecnologia moldam ativamente como os CAs são projetados.

Porém, com tantos exemplos ruins, devemos mesmo seguir o que essas grandes empresas fazem?

Quais as consequências?

01

Manifestação de estereótipos de gênero: uma vez que pessoas usuárias interagem majoritariamente com *voice assistants* usando comandos curtos, podem considerar apropriado conversar dessa forma com mulheres

02

Cria expectativas e reforça os pressupostos de que as mulheres devem **fornecer respostas simples, diretas e pouco sofisticadas a questões básicas**

03

Muitas crianças interagem com CAs baseados em voz e os estereótipos de gênero são ensinados principalmente em uma idade muito jovem.

O estudo propôs 4 proposições que podem ajudar a mitigar o preconceito de gênero e abrir caminho para um design de chatbots com maior igualdade:

P1

Diversificar a composição das equipes de desenvolvimento de chatbots: mulheres e profissionais de linguagens e psicologia.

P2

Estimular ferramentas de suporte na identificação de preconceito de gênero no design de chatbots: uma vez que estereótipos são disseminados de uma forma inconsciente, ferramentas podem ajudar.

O estudo propôs 4 proposições que podem ajudar a mitigar o preconceito de gênero e abrir caminho para um design de chatbots com maior igualdade:

P3

Evitar um design de chatbot “feminino de fábrica”:

- Ainda não está claro se devemos construir *chatbots* sem gênero ou tentar alterar as percepções humanas sobre os bots, é necessário que a comunidade esteja aberta a pesquisar.
- No entanto, equipes desenvolvedoras de chatbot precisam implementar ativamente mecanismos para responder às perguntas desagradáveis do usuário.

P4

Promover discussões éticas nas organizações

- Tomadores de decisão e equipe de desenvolvimento precisam refletir sobre cada design de chatbot. Ao fazer isso, equipes podem se beneficiar com uma maior conscientização para construir sociedades mais igualitárias.

Minhas considerações:
o que está ao nosso
alcance hoje?

P1

Pensar se não estamos perpetuando estereótipos de gênero ao construir as **personas** dos bots.

P2

Utilizar **linguagem neutra** para evitar estereótipos.



DÚVIDAS?

[linkedin.com/in/beatrizcaetano](https://www.linkedin.com/in/beatrizcaetano)

beatrizcaetano@outlook.com

bit.ly/feedbackviesesdegenero