

**CESAR SCHOOL**

**FERNANDA MARIA DE SOUZA ALVES**

**ANÁLISE DA ABORDAGEM ANTIRRACISTA COMO PILAR FUNDAMENTAL NO  
DESIGN DE INTERVENÇÕES DE INTELIGÊNCIA ARTIFICIAL NA ÁREA DE SAÚDE**

**RECIFE**

**2024**

**FERNANDA MARIA DE SOUZA ALVES**

**ANÁLISE DA ABORDAGEM ANTIRRACISTA COMO PILAR FUNDAMENTAL NO  
DESIGN DE INTERVENÇÕES DE INTELIGÊNCIA ARTIFICIAL NA ÁREA DE SAÚDE**

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Design do Centro de Estudos e Sistemas Avançados do Recife - CESAR School como requisito para a obtenção do título de Mestre em Design.

Orientadora: Profa. Dra. Helda Oliveira Barros.

**RECIFE**

**2024**

Catálogo da Publicação na Fonte (CIP)  
Centro de Estudos e Sistemas Avançados do Recife – CESAR School Recife  
Bibliotecária Ana Clara Amorim, CRB-2145

A474a Alves, Fernanda Maria de Souza

Análise da abordagem antirracista como pilar fundamental no design de interações artificial na área de saúde / Fernanda Maria de Souza Alves. – Recife: A Autora, 2024.

83p. : il.

Orientação: Profa. Helda Oliveira Barros.

Dissertação (Mestrado) - Centro de Estudos e Sistemas Avançados do Recife – CESAR School Recife. Programa de Pós-graduação em Design.

Inclui referências.

1. Inteligência artificial. 2. Racismo estrutural. 3. Equidade na saúde. 4. Antirracismo. I. Barros, Helda Oliveira (Orientadora). II. Título.

CDD 745.2

CESAR School (BiblioBrum2024)



**FERNANDA MARIA DE SOUZA ALVES**

**ANÁLISE DA ABORDAGEM ANTIRRACISTA COMO PILAR FUNDAMENTAL NO  
DESIGN DE INTERVENÇÕES DE INTELIGÊNCIA ARTIFICIAL NA ÁREA DE SAÚDE**

Trabalho aprovado em Recife: 27/03/2024.

**Professora:** Helda Oliveira Barros  
**(CESAR SCHOOL)**  
Orientadora

**Professor:** José Carlos Porto Arcoverde Júnior  
**(CESAR SCHOOL)**  
Avaliador Interno

**Professora:** Tatalina Cristina Silva de Oliveira  
**(CESAR SCHOOL)**  
Avaliadora Interna

**Professor:** Renan Costa Alencar  
**(FACULDADE ESUDA)**  
Avaliador Externo

**RECIFE  
2024**

## **Dedicatória**

Dedico esta dissertação aos meus pais, que sempre me apoiaram e incentivaram, se dedicaram ao máximo ao longo de suas vidas para me oferecer as melhores oportunidades e condições para que eu pudesse alcançar meus objetivos, mesmo quando estes divergiam de seus planos.

## Agradecimentos

Ao meu irmão, Emannuel, ao meu primo-irmão Lucas por todo apoio e amor e minha amiga-irmã, Lu, sua capacidade de superação me incentivam a nunca desistir e sua paixão pelos estudos me inspira a ser uma aluna cada dia melhor. Mannu, Lucas e Lu, nossas conversas sobre a vida e mundo enriquecem minha percepção e me ajudam a crescer como pessoa, sou muito grata por tê-los ao meu lado.

Agradeço a Ada (*in memoriam*), que foi como uma segunda mãe para mim, sempre me encorajando e acreditando em meu potencial. Seu abraço e alegria ao saber que fui selecionada para este mestrado foi um momento inesquecível.

Sou grata também a Potyra, meu amor e companheira de vida, por seu apoio fundamental em cada fase deste trabalho. Como pesquisadora, seu suporte foi crucial, especialmente diante das inquietações sobre injustiça racial. Nossas conversas diárias são minha fonte inspiradora para lutar por um mundo mais justo e equitativo.

Aos meus amigos, que compreenderam minhas ausências devido aos compromissos com o mestrado, especialmente a Davi, que sempre esteve presente para me apoiar, mesmo quando ocupado com sua própria dissertação.

Ao meu amigo, Luiz, pelo incentivo constante, sempre disponível para oferecer feedback e colo quando necessário.

Agradeço imensamente ao SiDi, e às minhas líderes, Fernanda e Angélica, pelo suporte e incentivo que me proporcionaram ao longo deste trabalho.

Aos meus amigos Designers do CIN/SiDi, em especial a Túlio, Verinha, Sophia e Mano, que me inspiram a sempre buscar conhecimento e ser uma profissional melhor.

À minha orientadora e amiga, Helda Barros, pela orientação cuidadosa, serena e generosa ao longo deste processo de pesquisa.

À turma MPD 2021.2, aos amigos do grupo Vermelho "Axé Poco", Alice, Adalberto, Davi, Felipe, Ramon e Renata, obrigada pelas risadas, pela colaboração e apoio que tornaram esta jornada mais leve.

Por fim, gostaria de expressar minha gratidão a todas as pessoas que estiveram ao meu lado durante esta jornada.

## Epígrafe

*"Um passo à frente e você não está mais no mesmo lugar."*

*Chico Science, 1996.*

## Prólogo

Como mulher não-branca, Designer e Pesquisadora, olho para o mundo por meio de lentes que filtram não apenas a realidade que me cerca, mas também a herança de um passado impregnado de opressão, exploração e discriminação. Minha pele, embora não-branca, não me absolve das amarras do racismo estrutural, mas sim me coloca em uma posição peculiar dentro do complexo sistema de privilégios e desigualdades que permeiam a sociedade. Sou beneficiada e, ao mesmo tempo, responsável pelo pacto da branquitude, conforme conceituado por Cida Bento (2022), uma pactuação que se estende não apenas aos acordos formais, mas às sutilezas das relações de poder que moldam as estruturas institucionais e sociais do Brasil contemporâneo.

O pacto da branquitude, embora não verbalizado, mantém um segmento homogêneo, predominantemente masculino e branco, nos espaços de poder do país. Este pacto não é um acordo formal, mas sim um conjunto de práticas e valores que sustentam as desigualdades raciais, relegando aqueles que não fazem parte desse grupo ao ostracismo e à marginalização (Bento, 2022).

Ao encarar o pacto da branquitude, é inegável percebê-lo também como parte do racismo estrutural. O racismo estrutural, no Brasil, é reforçado por uma ideologia oriunda do período colonial e mantida na realidade atual como tecnologia da colonialidade, visando manter o *status* de superioridade do branco entre as instituições sociais. É importante ressaltarmos que essa noção de superioridade de uma suposta raça branca sobre a negra é construída histórica, cultural e politicamente e não se dá somente em função de cor, mas das circunstâncias que conduziram essa população branca a acessar bens materiais e simbólicos (Almeida, 2019).

Dado o enraizamento do racismo estrutural na sociedade, torna-se urgentemente necessário promover a sensibilização e estimular a reflexão sobre esses temas. Nesse contexto, os materiais audiovisuais, como as séries de TV, desempenham um papel crucial. O segundo episódio da temporada dezoito da série de TV *Grey's Anatomy*, que

aborda o tema de vieses raciais na medicina, foi uma inspiração adicional para aprofundar meu trabalho na área da saúde. Esta é uma área em que, mesmo quando não existe a presença da tecnologia, persistem disparidades raciais intrínsecas nos protocolos, tratamentos e diagnósticos realizados diariamente em todo o mundo.

Nesse cenário, o *Transition Design* emerge como uma inspiração filosófica vital, principalmente durante o desenvolvimento de ferramentas de Inteligência Artificial. Como afirmado por Tonkinwise (2015), o *Transition Design* reconhece a extensão das crises sociais contemporâneas e promove práticas de design social e sustentável. É uma abordagem que busca inovação orientada pela prática em várias etapas, visando não apenas resolver problemas imediatos, mas também transformar as estruturas e sistemas que perpetuam as desigualdades sociais, raciais, econômicas e ambientais.

A partir desta perspectiva, exploraremos a interseção entre a abordagem antirracista e o design de intervenções de inteligência artificial na área da saúde. Este trabalho é mais do que uma dissertação acadêmica; é uma resposta pessoal e coletiva ao pacto de branquitude que, em muitos momentos, me calei e compactuei. É um lembrete constante para mim mesma e para a comunidade sobre a necessidade de coragem, persistência e compromisso no combate às injustiças raciais. É uma convocação à ação, uma convocação para a quebra desse pacto por meio do design, do compartilhamento dos saberes de autores não brancos e eurocêntricos, da autorresponsabilidade e de práticas antirracistas.

## Resumo

O crescente uso de inteligência artificial (IA) no setor da saúde no Brasil enfrenta o desafio crítico de vieses raciais, que ameaçam amplificar as desigualdades existentes e comprometer a qualidade do atendimento em um país marcado por sua significativa diversidade racial. Esta pesquisa visa investigar como a inclusão de práticas antirracistas no design de soluções de IA pode contribuir para a equidade em soluções de inteligência artificial na área de saúde. Busca também entender os mecanismos pelos quais vieses raciais afetam as soluções de IA, além de propor estratégias para mitigar esses vieses de forma eficaz. A partir de uma abordagem metodológica exploratória, indutiva e construtivista, baseada na análise de entrevistas com profissionais da saúde, designers e desenvolvedores, o estudo destaca a escassez de iniciativas antirracistas no ciclo de desenvolvimento de soluções de IA, evidenciando a necessidade de reformular as práticas atuais. Assim, revela-se a necessidade de uma abordagem holística para alcançar equidade em soluções de IA na saúde, enfatizando a adoção de práticas antirracistas, a promoção de diversidade nas equipes de desenvolvimento, e a importância de colaborações multidisciplinares. Essas medidas buscam garantir que as soluções de IA atendam às necessidades variadas da população, com a ética, a equidade racial e as práticas explicitamente antirracistas se tornando elementos centrais da inovação em IA na saúde.

**Palavras-chave:** Inteligência Artificial; Racismo estrutural; Equidade na saúde; Antirracismo.

## **Abstract**

The increasing use of artificial intelligence (AI) in the healthcare sector in Brazil faces the critical challenge of racial biases, which have the potential to amplify existing inequalities and compromise the quality of care in a country marked by its significant racial diversity. This research aims to investigate how the inclusion of anti-racist practices in the design of AI solutions can contribute to equity in artificial intelligence solutions in healthcare. It also seeks to understand the mechanisms by which racial biases affect AI solutions, as well as proposing strategies to effectively mitigate these biases. Using an exploratory, inductive, and constructivist methodological approach, based on the analysis of interviews with healthcare professionals, designers, and developers, the study highlights the scarcity of anti-racist initiatives in the AI solutions development cycle, showing the need to reformulate current practices. In conclusion, it reveals the need for a holistic approach to achieve equity in AI solutions in healthcare, emphasizing the adoption of anti-racist practices, the promotion of diversity in development teams, and the importance of multidisciplinary collaborations. These strategies aim to ensure that AI solutions meet the varied needs of the population, with ethics, racial equity, and explicitly anti-racist practices becoming central elements of innovation in healthcare AI.

**Keywords:** Artificial Intelligence; Systemic Racism; Health Equity; Antiracism.

## Lista de ilustrações

Figura 1: Mapeamento das 702 startups de IA no Brasil.	21
Figura 2: Desenho da Pesquisa.	30
Quadro 1: Objetivos específicos da pesquisa em tópicos	34
Quadro 2: Perguntas da entrevista, de acordo com hipótese e tópicos	35
Figura 3: Ficha de planejamento das entrevistas	37
Quadro 3: Frameworks selecionados para o benchmarking	41
Figura 4: Processo de análise dos frameworks selecionados	42
Figura 5: Análise individual do framework 1.	44
Figura 6: Análise individual do framework 2.	46
Figura 7: Análise individual do framework 3.	48
Figura 8: Análise individual do framework 4.	50
Figura 9: Análise individual do framework 5.	52
Figura 10: Análise individual do framework 6.	54
Quadro 4: Resumo de memorandos criados durante as entrevistas	56
Figura 11: Mapa mental para organização visual dos dados coletados nas entrevistas.	59
Quadro 5: Codificação inicial	59
Quadro 6: Codificação focalizada	63
Quadro 7: Categorias finais	64
Figura 12: Processo de codificação e categorização	65

## Lista de Tabelas

Tabela 1: Perfil dos participantes

34

## **Lista de Abreviaturas e Siglas**

Chat - Bate-papo.

ECCD - Design de Equidade Centrado na Comunidade.

ESG - Ambiental, Social e Governança.

GPT - Generative Pre-Trained Transformer – Chat gerador transformador pré-treinado.

IA - Inteligência Artificial

LLM - Large Language Model

ML - Machine Learning

ONU - Organização das Nações Unidas

PIB - Produto Interno Bruto

TFD - Teoria fundamentada nos dados

USP - Universidade de São Paulo

## Sumário

<b>1. INTRODUÇÃO</b>	<b>13</b>
1.1 Tema	13
1.2 Problematização	16
1.3 Justificativa	16
1.4 Objeto de estudo	18
1.5 Objetivos	18
1.5.1 Objetivo Geral	18
1.5.2 Objetivos Específicos	18
<b>2 REFERENCIAL TEÓRICO</b>	<b>19</b>
2.1. Inteligência Artificial (IA)	19
2.2 Vieses e desafios da inteligência artificial	22
2.3 Racismo na área de saúde	25
2.4 Tecnologia para a Equidade: Design Antirracista em IA	27
<b>3 METODOLOGIA</b>	<b>29</b>
3.1 Desenho da Pesquisa	30
3.2 Etapas da Pesquisa	30
3.2.1 Referencial teórico	30
3.2.2 Identificação de práticas e premissas	31
3.2.3 Levantamento do cenário atual	31
3.2.4 Criação das teorias	31
3.3 Amostra de Participantes	32
3.5 Recrutamento dos Participantes:	32
3.6 Instrumentos de Coleta de Dados	32
3.7 Análise e interpretação dos dados	32
<b>4 RESULTADOS</b>	<b>33</b>
4. 1 Identificação de práticas e premissas	33
4.2 Principais resultados a partir das entrevistas	46
4.3 Descrição de fenômenos	52
<b>5 DISCUSSÃO</b>	<b>59</b>
<b>6 CONSIDERAÇÕES FINAIS</b>	<b>63</b>
6.1 Conclusão	63
6.2 Trabalhos futuros e direções estratégicas	64
<b>REFERÊNCIAS</b>	<b>67</b>

# 1. INTRODUÇÃO

## 1.1 Tema

O avanço da Inteligência Artificial (IA) no setor da saúde é capaz de transformações significativas na maneira como os cuidados são oferecidos e acessados pelas pessoas. No entanto, a implementação dessa tecnologia vem acompanhada de importantes questionamentos éticos e morais, sobretudo diante da persistência do racismo estrutural nas sociedades.

O racismo é sempre estrutural, ou seja, de que ele é um elemento que integra a organização econômica e política da sociedade. Em suma, o que queremos explicitar é que o racismo é a manifestação normal de uma sociedade, e não um fenômeno patológico ou que expressa algum tipo de anormalidade. O racismo fornece o sentido, a lógica e a tecnologia para a reprodução das formas de desigualdade e violência que moldam a vida social contemporânea (Almeida, 2019, p. 19).

Também de acordo com Almeida (2019), o entendimento do racismo como parte integrante da estrutura social não exime os indivíduos de responsabilidade perante atitudes racistas, mas reforça a necessidade de uma postura ativa contra o racismo. Reconhecendo que o racismo transcende a necessidade de intenções deliberadas para se manifestar, é imperativo que o combate ao mesmo inclua ações antirracistas efetivas, pois o silêncio ou a omissão contribuem para sua perpetuação. Almeida (2019) destaca que a noção de raça é intrinsecamente uma relação social que se expressa por meio de conflitos e antagonismos concretos em uma determinada estrutura social. Assim, o racismo é visto não apenas como uma série de eventos isolados, mas como um fenômeno enraizado tanto na história quanto na política.

A centralidade da ética e da perspectiva antirracista no processo de design é fundamental para assegurar que o desenvolvimento de produtos e serviços de IA na saúde não avance somente em termos tecnológicos, mas também promova equidade e a inclusão social.

No relato de experiência “Interdisciplinaridade e diversidade racial na formação do profissional de saúde”, Germani et al. (2022) abordam a experiência de uma disciplina voltada ao combate ao racismo na Universidade de São Paulo (USP), reforçando a importância da ética e da perspectiva antirracista no processo de design em IA na saúde. Especificamente, quando descrevem como a disciplina buscou promover uma formação antirracista, integrando saberes interdisciplinares e valorizando a diversidade racial, para preparar profissionais de saúde capazes de enfrentar as desigualdades raciais dentro do setor. Esse enfoque é um exemplo concreto de como a educação e a formação profissional, quando alinhadas com princípios éticos e antirracistas, podem influenciar positivamente o desenvolvimento de práticas inclusivas e equitativas, inclusive na área da tecnologia e da IA.

Grupos subalternizados com pouco ou nenhum acesso à educação, acometidos por deficiências socioeconômicas e pertencentes a grupos minoritários, como mulheres, negros e indígenas, enfrentam dificuldades de acesso e uso das ferramentas tecnológicas (Lima et al., 2020, p. 3).

Lima et al. (2020), realçam ainda mais as barreiras enfrentadas por grupos marginalizados no acesso e uso de tecnologias, enfatizando a importância de considerar essas barreiras na concepção de intervenções de IA. Portanto, para uma concepção de intervenções de IA de forma responsável, é exigido uma reflexão profunda sobre como essas soluções podem servir equitativamente a toda a população, sem reforçar preconceitos existentes.

No campo do Design, Tonkinwise (2015) defende a necessidade de uma teoria de mudança mais explícita, que vem se expandindo em alcance e influência desde o século XX. O autor observa uma evolução do design, de uma atividade de resolução de problemas voltada para a criação de novas tecnologias e produtos, para uma prática de pensamento de design aplicada a questões sociais e gerenciais. Ele destaca que o design moderno deve transcender a visão simplista do ser humano apenas como consumidor e considerar a complexidade das interações sociais e éticas. Tonkinwise (2015) ressalta a importância do Design Social em promover mudanças sociais profundas e de longo prazo, utilizando conhecimento sobre mudanças individuais,

sociais e a difusão de inovações. Esse enfoque capacita os designers a serem agentes efetivos de mudança, mantendo sua habilidade essencial no design, e exige uma reavaliação dos papéis sociais dos designers, particularmente no que tange ao design que visa mudanças estruturais em contextos de urgência e crise.

Perez et al. (2022) afirmam ainda que o design contemporâneo enfrenta o desafio de transcender as práticas tradicionais para abordar complexidades socioambientais crescentes, reconhecendo a necessidade de uma abordagem mais sistemática e transdisciplinar. Neste contexto, emerge o design de transições para a sustentabilidade, que visa promover mudanças sistêmicas e radicais, enraizadas no contexto local, por meio do desenvolvimento colaborativo de cenários desejáveis e planos de ação estratégicos. Este segmento, ainda em formação, busca inspiração e contribuições de diversas disciplinas e abordagens de design, propondo novas formas de pensar e praticar o design em uma estrutura ética e sustentável.

A transição para práticas de design sustentáveis não apenas reflete uma expansão do campo de atuação do design, mas também uma resposta à urgente necessidade de abordagens decoloniais que reconsiderem e ressignifiquem às relações entre humanos, tecnologia e ambiente. Esse processo implica à adoção da cosmovisão do Bem Viver como um referencial ético e filosófico, desafiando os designers a repensarem o futuro de maneira coletiva, inclusiva e diversificada, alinhando-se a uma práxis decolonial que valoriza saberes locais e promove a inclusão social (Perez et al., 2022).

Embora abordagens como o *Transition Design* ou “Design de transição”, Tonkinwise (2015); Perez et al. (2022), apresentem uma sólida perspectiva de design social, e existem diversas ferramentas de design que inspiram e orientam os profissionais na concepção ética de produtos e serviços com IA, observa-se uma lacuna significativa no que tange à abordagem do racismo estrutural de forma prática. Esta omissão resulta em produtos que, inadvertidamente, podem perpetuar ou ampliar desigualdades raciais no setor de saúde.

Enfrentar o desafio de integrar a ética e a luta antirracista no design de intervenções de IA na saúde requer uma mudança de paradigmas nas práticas de desenvolvimento. Isso implica não apenas a reavaliação das ferramentas de design existentes focadas no processo de criação de produtos e serviços de IA, mas também um compromisso com a

formação de designers, pessoas desenvolvedoras e profissionais da saúde conscientes das complexas interações entre tecnologia, sociedade e racismo estrutural.

## **1.2 Problematização**

Em um contexto global onde as disparidades raciais e étnicas persistem, é crucial examinar como as práticas de design podem influenciar a equidade e a justiça na saúde.

A partir de uma abordagem especializada de design que reconhece a extensão das nossas crises sociais e promovendo as práticas de design social e sustentável por meio da incorporação de uma inovação orientada pela prática em várias etapas proposta por Tonkinwise (2015), somos levados ao seguinte questionamento:

Como o processo de design a partir de uma perspectiva ética e antirracista pode promover, de forma responsável, a criação e desenvolvimento de produtos ou serviços que utilizem IA na área de saúde?

Apesar de existirem diversas ferramentas de design para idear produtos e serviços com IA de forma responsável e ética, o racismo estrutural é pouco mencionado nessas ferramentas. Desta forma, quando o produto já está no mercado, empresas e profissionais continuam perpetuando o racismo estrutural na área de saúde, tanto pelos profissionais quanto também pelos sistemas que utilizam IA.

## **1.3 Justificativa**

Para o designer Klaus Krippendorff (2000), o design tradicional, focado no produto e em seus aspectos objetivos, passou para um design centrado no ser humano e em sua forma de perceber, interpretar e interagir com o mundo. Essa mudança representa uma evolução significativa no campo do design, onde a prioridade agora é compreender as necessidades, desejos e experiências dos usuários para criar soluções mais eficazes e

significativas. De acordo com Krippendorff (2000), o design centrado no ser humano busca colocar as pessoas no centro do processo de design, considerando suas emoções, valores e contextos sociais, a fim de desenvolver produtos e serviços que atendam verdadeiramente às suas necessidades e melhorem sua qualidade de vida. Essa abordagem enfatiza a importância da experiência do usuário, promovendo a empatia, a inclusão e a sustentabilidade no desenvolvimento de soluções inovadoras e relevantes para a sociedade.

Os profissionais de design têm papel fundamental ao projetar novos produtos e serviços. Logo, o incentivo ao desenvolvimento de produtos e serviços que utilizem I.A. por meio do processo de design de forma responsável, ética e antirracista, poderá contribuir para a redução de impactos éticos negativos na sociedade, além de mitigar o racismo estrutural na medicina por meio de abordagem consciente de raça e não baseada em raça. O estudo é também uma forma de promover a equidade social, contribuindo para o alcance do Objetivo de Desenvolvimento Sustentável 10, almejado pela Organização das Nações Unidas (ONU), que busca a redução das desigualdades.

Além disso, é essencial destacar que a incorporação de práticas de design ético e antirracista não apenas é benéfico socialmente em termos de equidade e justiça, mas também pode impactar positivamente o desempenho das empresas em relação a questões ambientais, sociais e de governança (ESG). Ao adotar uma abordagem inclusiva e consciente, as empresas têm a oportunidade não apenas de promover a equidade e a justiça, mas também de avançar em direção a um modelo de negócio mais sustentável e responsável, alinhado aos princípios do ESG.

Este estudo é focado na ética do ser humano em seu aspecto social, pois é sabido a importância de refletir e ter um olhar crítico sobre as implicações no âmbito coletivo da criação de produtos e serviços que utilizam inteligência artificial e seus impactos na sociedade. Também é esperado estimular uma reflexão investigativa sobre as implicações que a prática antirracista durante o processo de design de novos produtos e serviços pode trazer para a coletividade na área de saúde.

## **1.4 Objeto de estudo**

Guias, kits de ferramentas, frameworks ou boas práticas que podem promover a consideração de fatores éticos, antirracistas e de responsabilidade social durante a concepção de produtos de que utilizam IA na área de saúde.

## **1.5 Objetivos**

### **1.5.1 Objetivo Geral**

Analisar o uso da abordagem antirracista como um requisito essencial no design de produtos e serviços que façam uso de inteligência artificial (IA) em projetos destinados para a área da saúde.

### **1.5.2 Objetivos Específicos**

1.5.2.1 Gerar histórico sobre como o racismo é base estrutural na área da saúde;

1.5.2.2 Compreender como a ética e práticas antirracistas podem ser premissas indispensáveis, durante o processo de design de novos artefatos;

1.5.2.3 Levantar o cenário atual quanto a perspectiva ética e antirracista durante o processo de criação de novas soluções que utilizam inteligência artificial na área de saúde.

## 2 REFERENCIAL TEÓRICO

### 2.1. Inteligência Artificial (IA)

A Inteligência Artificial é um campo de estudo, cujas inovações e desenvolvimentos resultaram em computadores, máquinas e outros artefatos que possuem uma inteligência semelhante à humana, caracterizada por habilidades cognitivas, aprendizado, adaptabilidade e capacidades de tomada de decisão (Chen; Chen; Lin, 2020)

Nos últimos anos, presenciamos uma aceleração não só no desenvolvimento tecnológico, mas também no acesso a tecnologias emergentes via artefatos que utilizam modelos de linguagem como o ChatGPT (Chat - *Generative Pre-trained Transformer*), artefato que permite aos usuários sem conhecimento técnico em inteligência artificial, gerar conteúdo textual imitando uma conversa humana, ou o DALL-E (Modelo GPT), plataforma digital que permite a geração de imagens por conteúdo textual informado pelo usuário. O que estas tecnologias têm em comum é a utilização de LLM (*Large Language Model*) ou seja, modelos de linguagem grandes. Este tipo de modelo se tornou ainda mais popular após o lançamento do ChatGPT, pela empresa OpenAI, em 30 de novembro de 2022, com uma interface simples e intuitiva que possibilita o acesso facilitado à tecnologia por pessoas sem o conhecimento técnico sobre ela. Este acesso facilitado e gratuito desperta uma preocupação quanto aos aspectos relacionados a ética, já que modelos de inteligência artificial são sistemas computacionais capazes de serem treinados com dados já existentes e aprender a partir deste treinamento. Por ser um sistema computacional treinado com dados, estes modelos podem ter vieses ou limitações quando se trata de aspectos sociais.

Embora a IA possa emular muitos aspectos do comportamento humano, a IA ainda é limitada e não possui verdadeira compreensão ou consciência. Ela opera estritamente com base em dados, algoritmos e modelos pré-determinados. Por este motivo, geralmente, o processo de desenvolvimento, para tornar o modelo mais preciso, também inclui técnicas de *fine-tuning*, e podem estar relacionadas as questões de viés,

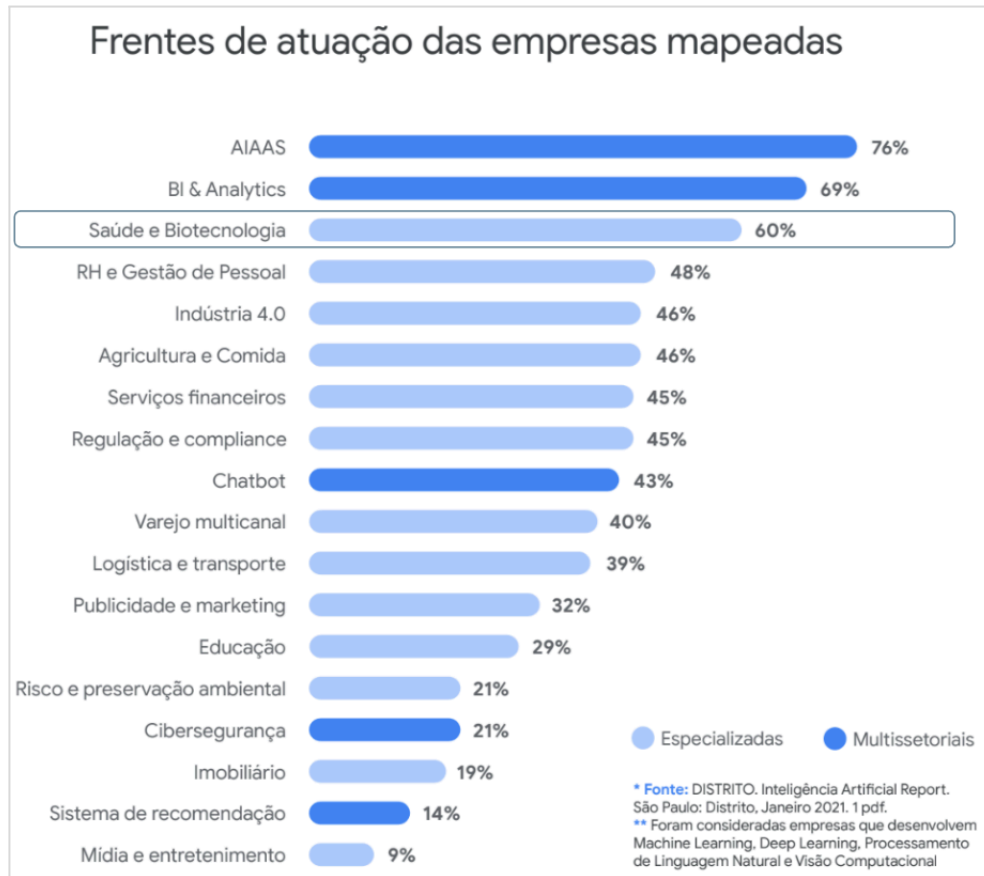
o que é um aspecto importante a ser considerado ao adaptar modelos pré-treinados para tarefas específicas.

Conforme Wiles (2023), empresas de capital de risco investiram mais de 1,7 mil milhões de dólares em soluções generativas de IA nos últimos três anos, isto é um indicativo de como o uso de inteligência artificial está em fase de crescimento exponencial nas últimas décadas, sendo que estes avanços despertam o interesse crescente de empresas para a criação de produtos e serviços que utilizem essa tecnologia.

Com essa quantidade de investimentos, a área de IA vem ganhando ainda mais popularidade no mundo com diversos sites que reúnem milhares de ferramentas de IA em diversos segmentos. Já no Brasil, como mostra o relatório “O impacto e o futuro da Inteligência Artificial no Brasil” (2022) realizado pelo Google for Startups em parceria com a empresa Box1824 e com a Abstartups, em um estudo global, estima-se que a I.A. gerará 13 trilhões de dólares no mundo até 2030 (Bughin et al., 2018). No mesmo ano, é esperado um aumento de 5% no Produto Interno Bruto (PIB) da América Latina em virtude da I.A. Há uma expectativa que o uso de aplicações com I.A. em saúde cresça 38% na região até 2027.

Ainda no relatório “O impacto e o futuro da Inteligência Artificial no Brasil” podemos notar que existe uma quantidade grande de empresas especializadas que atendem o setor de Saúde e tecnologia (Figura 1). Isso chama atenção para um ponto extremamente importante, que são os vieses e desafios enfrentados na criação e desenvolvimento dessas soluções na área de saúde.

Figura 1: Mapeamento das 702 startups de IA no Brasil.



Fonte: DISTRITO. Inteligência Artificial Report (2021).

Soluções que utilizam IA impactam também a área da saúde, com aplicações como prontuários eletrônicos, assistentes inteligentes de diagnóstico e triagem de pacientes. No entanto, à medida que a IA se torna mais integrada na área da saúde, surgem desafios significativos que precisam ser enfrentados, como a replicação de vieses que já existem nas decisões humanas, incluindo o racismo estrutural, que é um desafio crítico que precisa ser abordado, já que a IA aprende a partir de dados históricos e, se esses dados contêm vieses, o modelo pode perpetuá-los ou até ampliá-los. Isso é particularmente preocupante na área da saúde, onde a reprodução de vieses pode levar a decisões clínicas injustas, ampliar a desigualdade e prejudicar a qualidade do atendimento.

Para combater estes desafios, é fundamental adotar uma estratégia que inclua várias disciplinas, além de diversificar os dados usados para treinar modelos de

Inteligência Artificial. Assim, sendo estes mais variados, o impacto desses vieses nas decisões da IA tende a ser menor. Além disso, o método de design utilizado na criação de produtos de IA pode facilitar a adoção de métodos efetivos para identificar e atenuar vieses nos dados.

## **2.2 Vieses e desafios da inteligência artificial**

O termo Viés em modelos de IA refere-se ao potencial de reproduzir ou ampliar preconceitos, estereótipos ou desigualdades presentes nos dados de treinamento. Quando um modelo é pré-treinado em um conjunto de dados grande e diversificado, ele pode absorver implicitamente os padrões presentes nesses dados, que podem incluir vieses culturais, raciais, de gênero e outros.

O problema do viés está relacionado às decisões algorítmicas tendenciosas. Essas tendências são morais, pois envolvem a preferência por determinados valores. Porém, a questão que surge é se máquinas são capazes de valorar a realidade. Em certo sentido são, pois elas foram programadas por seres humanos, que são necessária e intrinsecamente avaliadores da realidade. (Rossetti; Angeluci, 2021)

De acordo com Rossetti e Angeluci (2021), vieses são inclinações ou tendências que podem ser morais, refletindo preferências por determinados valores. Eles surgem porque os algoritmos, embora possam parecer neutros, refletem os valores, escolhas e potencialmente, preconceitos de seus criadores humanos. Os algoritmos são programados por pessoas que, inevitavelmente, fazem avaliações morais sobre a realidade. Ainda segundo Rossetti; Angeluci (2021), essas avaliações podem levar a decisões tendenciosas quando os algoritmos são aplicados, resultando em ações que podem não são eticamente neutras. Portanto, os vieses nos algoritmos destacam a importância crítica de considerar os valores éticos na programação e na implementação de sistemas de inteligência artificial, bem como na supervisão humana desses sistemas para mitigar consequências prejudiciais e inesperadas.

Em concordância com os autores supracitados, Silva (2020) afirma que o conceito de viés se refere a uma tendência sistemática ou inclinação para uma determinada direção, que pode influenciar a tomada de decisões, interpretações ou análises de forma não intencional. No contexto de algoritmos e inteligência artificial, os vieses podem surgir devido aos dados de treinamento enviesados, preconceitos dos desenvolvedores ou falhas nos algoritmos que resultam em discriminação injusta, além de resultados distorcidos. É importante identificar e mitigar vieses em sistemas automatizados para garantir decisões justas e imparciais.

Como algoritmos e modelos de linguagem, tem como base o aprendizado a partir de dados existentes, então é comum que produtos e serviços que utilizem esse tipo de tecnologia em algum nível estejam suscetíveis a vieses históricos e sociais criados durante a construção política e de poder das sociedades.

A humanidade, ao longo de sua existência, tem passado por um contínuo processo de aprendizado, aperfeiçoamento e transformação. Esse ciclo ininterrupto está intrinsecamente ligado ao momento histórico e tecnológico em que vivemos, bem como ao meio social que nos cerca. Esses fatores podem contribuir para a formação e consolidação dos valores éticos e morais que orientam nossas ações e decisões.

Os aspectos éticos das interações entre a sociedade e os algoritmos não dependem apenas do desenho das tecnologias, mas da combinação entre esse desenho e as relações de uso que a sociedade faz desses algoritmos. Portanto, a sociedade deve desenvolver estratégias éticas e políticas para lidar com essas questões, que não podem ser resolvidas apenas pelo aspecto tecnológico, mas também envolvem decisões interacionais, culturais e jurídicas diretamente relacionadas ao desenho algorítmico e ao seu uso.

(Rossetti; Angeluci, 2021)

Como forma para materializar a ética para o desenvolvimento de produtos de IA, a IEEE Global Initiative (2019), na publicação "Design eticamente alinhado: uma visão para priorizar o bem-estar humano com sistemas autônomos e inteligentes", apresenta um conjunto de metodologias visando guiar a conduta ética em projetos de inteligência artificial (IA). Essas metodologias enfatizam a importância de aderir aos valores

humanos estipulados pela Declaração Universal dos Direitos Humanos das Nações Unidas. Entre as principais abordagens destaca-se a ética incorporada ao design, a antecipação de princípios éticos no início do desenvolvimento, bem como a necessidade de sistemas autônomos serem transparentes e explicáveis. Além disso, ressalta-se a importância da equidade e justiça na aplicação dessas tecnologias, a proteção dos direitos humanos e privacidade dos indivíduos e a segurança e confiabilidade dos sistemas autônomos e inteligentes. Assim, ao adotar essas diretrizes, os projetos de IA visam respeitar e fomentar os valores humanitários essenciais, alinhando-se aos princípios fundamentais da Declaração Universal dos Direitos Humanos (IEEE Global Initiative, 2019).

A expansão do uso da IA na saúde traz consigo a responsabilidade de garantir que essas tecnologias sejam justas, inclusivas e equitativas. Isso requer um compromisso contínuo em examinar criticamente os sistemas de IA, identificando vieses e adotando medidas proativas para mitigá-los, a fim de proporcionar um atendimento médico de qualidade a todos, independentemente de sua raça ou origem.

Durante o desenvolvimento de soluções de IA, é esperado encontrar desafios durante o processo de *fine-tuning*, que ajusta o modelo para um conjunto de dados mais específico, não abordando ou reconhecendo vieses. Em vez disso, o modelo pode acabar perpetuando ou até mesmo ampliando vieses ao ser ajustado para o novo contexto.

Nesse sentido, no artigo “Detecção de discurso de ódio e mitigação de preconceito racial nas redes sociais com base no modelo BERT” (Mozafari, 2020), o autor demonstra como técnicas de *fine-tuning* podem ajudar a mitigar o racismo. Contextualizando, BERT é um modelo de inteligência artificial grande e pré-treinado do Google capaz de compreender o contexto das palavras em relação umas às outras em uma frase, permitindo também compreender estruturas de linguagem complexas, como a utilização de expressões idiomáticas. Neste artigo, foram separados dois conjuntos de dados do Twitter™, um com tweets escritos em inglês formal e outro escrito em “Black english” que é uma forma de inglês falado com influência africana. A partir disso, foi analisado como um modelo treinado previa rótulos em relação ao discurso de ódio nas redes sociais.

E aí os resultados mostram a existência de preconceito racial sistemático nesses classificadores treinados, uma vez que tendem a atribuir tweets escritos em “Black english” a classes negativas como racismo, sexismo, ódio e ofensivos com mais frequência do que tweets escritos em inglês formal.

No entanto, o preconceito racial nos classificadores usados no estudo reduziu significativamente depois que o mecanismo de alívio de preconceito foi incorporado. (Mozafari, 2020)

### **2.3 Racismo na área de saúde**

Profissionais na área de saúde em seu cotidiano, optam por atalhos baseados em raça durante a prática clínica. Este comportamento deriva de um racismo estrutural perpetuado desde a colonização europeia, fazendo com que a desigualdade racial aconteça também no atendimento de saúde. Como forma de mitigar a desigualdade racial, faz-se necessário que os profissionais de diversas áreas, e neste caso da área de saúde, tenham abordagens de atendimento clínico cada vez mais alinhados a práticas antirracistas.

Muitos estudantes de medicina iniciam seu treinamento com preconceitos raciais que são inconscientemente reforçados. A raça é frequentemente aprendida como um fator de risco independente para doenças, e não como um mediador de desigualdades estruturais resultantes de políticas racistas. As disparidades de saúde são apresentadas sem contexto, levando os alunos a desenvolver estereótipos nocivos com base na crença de que algumas populações são mais doentes do que outras (Cerdeña, et. al., 2020)

Ainda de acordo com Cerdeña et al. (2020), é proposto uma abordagem à medicina consciente da raça, onde se coloca em foco o racismo e não a raça dos pacientes. Desta forma, evita-se que vieses racistas perpetuados também durante a formação médica sejam perpetuados na prática clínica. Uma abordagem baseada em raça pode gerar um impacto desproporcional, como foi observado durante a Covid-19.

Segundo Souza et. al. (2021), a pandemia da Covid-19 destacou ainda mais essas desigualdades, com as pessoas negras sendo mais afetadas pela doença devido a fatores como a falta de acesso a informações precisas e a serviços de saúde adequados. A abordagem atual da medicina baseada em raça perpetua desigualdades na saúde, enquanto uma abordagem consciente da raça pode ajudar a mitigar essas desigualdades. Cerdeña et. al. (2020) também sugere várias recomendações para promover práticas antirracistas na medicina, incluindo a denúncia do uso da raça na formação de profissionais de saúde e nas diretrizes de tratamento, a inclusão de pessoas negras, indígenas e de outras etnias nas decisões de reforma das diretrizes de tratamento, e a consideração de barreiras estruturais à saúde que se sobrepõem à raça. Cerdeña et. al. (2020) então conclui que na área de saúde nos Estados Unidos da América (EUA), a colaboração entre setores é necessária para alcançar a equidade em saúde e que os profissionais de saúde devem usar seu capital cultural para advogar por políticas antirracistas.

Silva (2022) cita que foi realizado um estudo por pesquisadores estadunidenses sobre algoritmos comerciais usados na predição de necessidades por cuidados de saúde, com o intuito de detectar discriminação e viés contra grupos específicos. Revelou-se que algoritmos atribuíram injustamente a pacientes negros escores de risco elevados, afetando negativamente o acesso a tratamentos e recursos. Foi observado que, apesar de receberem escores similares, pacientes negros apresentavam condições de saúde mais graves em comparação aos brancos. A pesquisa indicou que tal discrepância tinha como origem preconceitos nos dados utilizados pelos sistemas, particularmente relacionados a despesas médicas passadas em relação à pacientes em mesma condição. Essa metodologia não levava em consideração as limitações econômicas enfrentadas por pacientes negros, decorrentes de desigualdades estruturais, e perpetuava uma distribuição desigual de cuidados médicos, ignorando o impacto de fatores socioeconômicos e raciais na saúde. O estudo é uma crítica a confiança em gastos anteriores como métrica de severidade de condições de saúde nos algoritmos, ressaltando que tal prática mantém a discriminação racial na alocação de serviços de saúde (Obermeyer, 2019).

Já no Brasil, segundo Leal et. al. (2017), foram identificadas disparidades raciais na atenção pré-natal e parto, evidenciando um gradiente de pior para melhor cuidado entre mulheres negras, pardas e brancas. Em comparação com as mulheres brancas, as mulheres negras foram as mais afetadas, com um período pré-natal inadequado, falta de vinculação à maternidade, ausência de acompanhante, grandes deslocamentos para o parto e menos anestesia local para a episiotomia. As mulheres negras e pardas também tiveram um risco maior de ter um período pré-natal inadequado e estar ausentes de um acompanhante durante o parto.

A conscientização dos profissionais de saúde, designers e desenvolvedores de IA sobre esses problemas é fundamental. A colaboração entre especialistas em saúde, ética, direitos humanos e IA pode contribuir para o desenvolvimento de diretrizes éticas e melhores práticas que minimizem o impacto negativo dos vieses na saúde.

## **2.4 Tecnologia para a Equidade: Design Antirracista em IA**

No cenário contemporâneo, a busca por equidade e justiça social tem sido fundamental em todas as esferas da sociedade, inclusive na tecnologia. O design antirracista em Inteligência Artificial (IA) emerge como uma abordagem indispensável para mitigar e combater vieses e discriminações presentes nos sistemas automatizados. Neste contexto, foram explorados diversos aspectos que sustentam a construção de produtos de IA orientados pela equidade e pela justiça racial. A utilização de ferramentas e frameworks especializados no design de produtos, especialmente em Inteligência Artificial (IA), representa uma abordagem essencial para a criação responsável e ética, visando a correção de preconceitos e a promoção de equidade e justiça racial. Schön (1988) nos lembra da importância do processo de enquadramento no design, destacando que para abordar efetivamente problemas complexos, como os encontrados no design antirracista em IA, os designers devem definir claramente os limites do problema, selecionar aspectos específicos para foco e impor uma coerência que guie as decisões e ações subsequentes.

Nesse contexto, o framework de Design de Equidade Centrado na Comunidade (ECCD), desenvolvido por Antionette Carroll (2014) e lançado pelo Creative Reaction

Lab, exemplifica perfeitamente como tais ferramentas podem ser utilizadas com eficácia para atender a essas necessidades. Ao integrar comunidades marginalizadas no processo criativo, o ECCD não apenas pode facilitar a análise e a mitigação de vieses raciais nos produtos de IA, mas também pode encorajar a concepção de soluções que genuinamente reflitam e respeitem a diversidade humana. Ao adotar frameworks especializados, que se mostram essenciais para o desenvolvimento de tecnologias antirracistas, transformamos práticas de design em ferramentas poderosas de mudança social. Assim, integram-se essas estruturas com práticas e ferramentas que promovem a equidade e a justiça racial, nos encaminhando para um futuro no qual a tecnologia é verdadeiramente inclusiva e empoderadora para todos.

### **3 METODOLOGIA**

Por ter necessidade de flexibilidade e por se tratar de uma pesquisa sobre tecnologia e fenômeno social, foi aplicado um estudo qualitativo e indutivo, utilizando como base a Teoria Fundamentada nos Dados (TFD) - construtivista, vertente criada por Charmaz (2009). A escolha por esta lente metodológica parte da natureza do problema, algo tão novo que ainda não possui teorias consolidadas.

Cada trabalho desenvolvido pede uma solução personalizada e que leve em consideração suas características peculiares. Portanto, o método também deve ser adaptado, a partir de uma estrutura básica, a fim de facilitar o desenvolvimento do projeto em questão (Panizza, 2004).

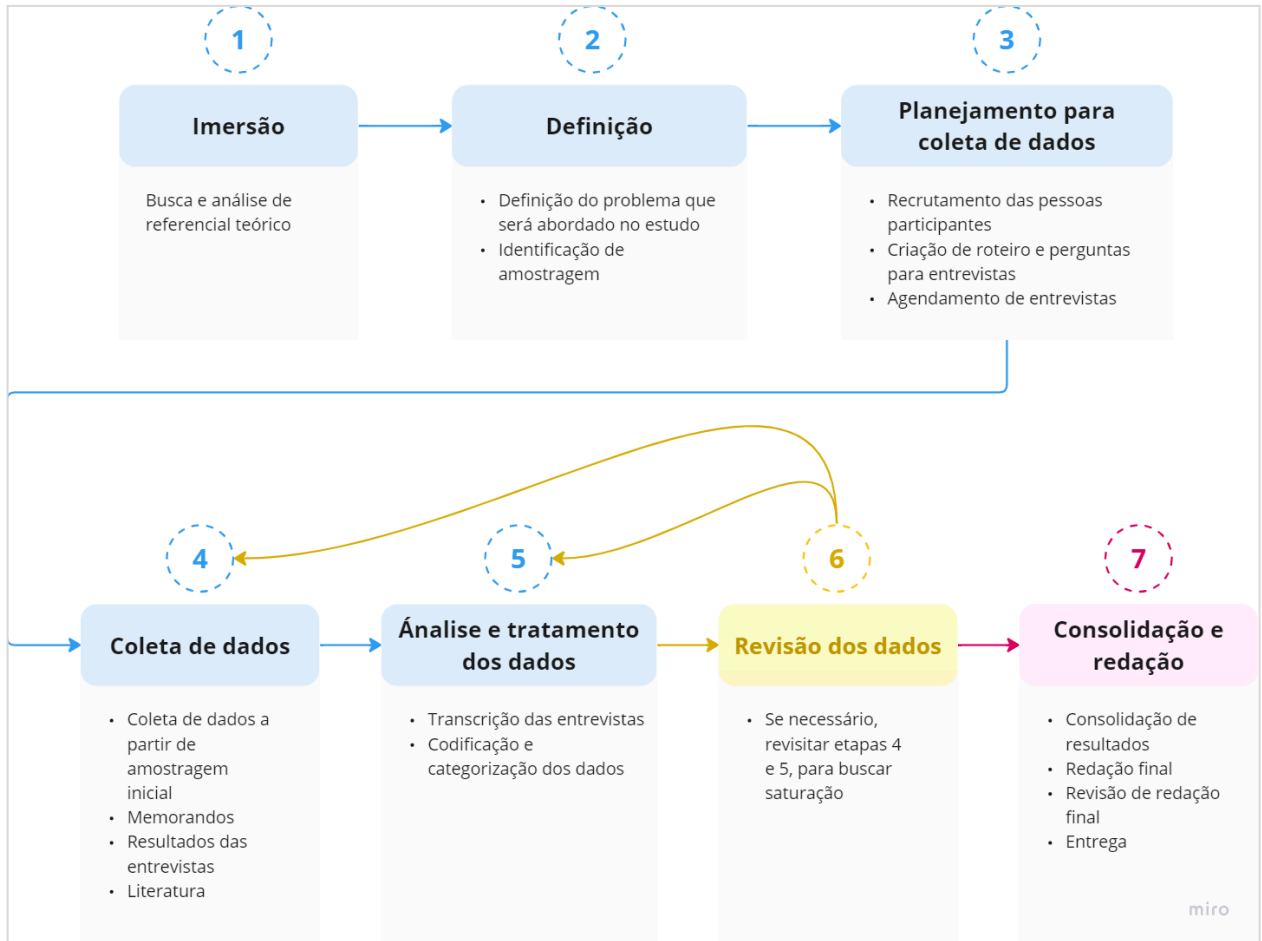
Charmaz (2009), afirma que a TFD serve como um modo de aprendizagem sobre o mundo que estudamos e como um método para elaboração de teorias para compreendê-los. Como somos parte do mundo que estudamos e também dos dados que coletamos, nós criamos nossas teorias fundamentadas a partir das interações com as pessoas, as perspectivas e as práticas de pesquisa. Nesta abordagem, Charmaz (2009) também admite que qualquer versão teórica apresenta uma perspectiva interpretativa do mundo estudado e não um retrato fiel dele.

A pesquisa será dividida em quatro etapas não necessariamente lineares. Algumas etapas podem ser revisitadas a qualquer momento que se fizer necessário, permitindo, desta forma, uma revisão bibliográfica antes, durante e após a coleta de dados. Na última etapa, analisada na ótica da TFD construtivista, foi possível coletar e analisar os dados ao mesmo tempo. Desta forma, permitiram-se comparações com as ideias identificadas nas entrevistas e como poderia aprofundar os dados.

### **3.1 Desenho da Pesquisa**

Neste estudo, foi adotada uma metodologia exploratória, indutiva e qualitativa, mapeada visualmente em um infográfico (Figura 2) que detalha o percurso percorrido. Inspirado nos princípios da teoria fundamentada nos dados de Charmaz (2009), o processo iterativo contemplou desde a imersão teórica inicial até a definição do problema e o meticuloso planejamento e realização da coleta de dados. A metodologia incluiu a coleta e análise aprofundada dos dados com revisões iterativas, culminando na consolidação e redação dos resultados. Cada fase foi projetada para fortalecer a integridade e a contribuição do estudo para o conhecimento no campo de pesquisa.

Figura 2: Desenho da Pesquisa.



Fonte: Elaborado pela autora (2023).

## **3.2 Etapas da Pesquisa**

### **3.2.1 Referencial teórico**

Essa etapa buscou gerar histórico sobre o tema. Para isto, foi realizado uma revisão abrangente da literatura, focada na compreensão de como o racismo estrutural se manifesta no setor de saúde. A investigação acadêmica envolveu a análise e síntese de uma diversidade de fontes, incluindo livros, artigos de periódicos, teses, dissertações e relatórios técnicos, tanto impressos quanto eletrônicos. Com base nestas fontes bibliográficas buscou-se entender a complexidade e as ramificações do racismo na saúde. Dessa forma, o referencial teórico propôs uma imersão nos pontos mais relevantes que conectam a área da saúde com as inovações e desafios apresentados pela Inteligência Artificial (IA)

### **3.2.2 Identificação de práticas e premissas**

Para identificar práticas e premissas, foi realizado uma revisão da literatura buscando por frameworks de design que incorporam práticas e premissas alinhadas com os desafios específicos ao tema dessa pesquisa. A análise focou na descoberta de estruturas que guiassem os designers na integração de considerações éticas e antirracistas no processo criativo. Foram selecionados seis frameworks significativos que ressaltam essas dimensões no desenvolvimento de produtos e soluções de saúde. Durante esse processo observou-se uma lacuna notável: a ausência de frameworks originários do Brasil, sugerindo uma oportunidade de desenvolvimento dentro do contexto nacional. A escolha por essa investigação reflete um esforço consciente para incorporar a ética e o compromisso antirracista como premissas fundamentais no design de novos artefatos, almejando que tais valores permeiem todas as fases da criação e se reflitam nos produtos finais na área da saúde.

Ao considerarmos o conceito de framework como um conjunto estruturado de conceitos e práticas e ferramentas que orientam a abordagem a problemas específicos, esta etapa buscou compreender, a partir da observação dos frameworks selecionados, como a ética e práticas antirracistas podem ser premissas indispensáveis, durante o processo de design de novos artefatos. A análise de Cross (2023) sobre o uso de frameworks no design corrobora diretamente com esta etapa da pesquisa, pois

segundo o autor, os designers se valem de estruturas conceituais e abordagens específicas para estruturar problemas de design e gerar soluções criativas.

Além disso, tal autor destaca a importância de uma visão sistêmica e ampla, onde os designers formulam perspectivas únicas para enfrentar os desafios, baseando-se em princípios fundamentais do design para orientar a concepção e desenvolvimento de soluções inovadoras. Essa abordagem, não apenas facilita a organização do pensamento criativo, mas também propicia uma maneira eficiente de lidar com problemas complexos e fomentar a inovação no processo de design. Assim, a integração de considerações éticas e antirracistas em *frameworks* de design pode ser vista como um passo essencial para garantir que novos artefatos reflitam valores de equidade e justiça social desde a sua concepção.

### **3.2.3 Levantamento do cenário atual**

Por meio de entrevistas, foi levantando o cenário atual quanto a perspectiva prática da ética e perspectiva antirracista das pessoas envolvidas no processo de criação de novas soluções que utilizam inteligência artificial na área de saúde.

Compreender o cenário atual por meio de entrevistas foi vital para captar a realidade vivenciada pelos profissionais que estão na linha de frente do design de soluções de saúde baseadas em inteligência artificial. Este levantamento se tornou um instrumento essencial, não só para mapear a aplicabilidade das normas éticas e das abordagens antirracistas descritas na literatura acadêmica, mas também para discernir como tais preceitos são transpostos para a prática cotidiana no mercado. Ao comparar as teorias e recomendações acadêmicas com as metodologias e estratégias realmente adotadas no desenvolvimento de produtos, tornou-se possível identificar tanto sinergias quanto discrepâncias, iluminando caminhos para a evolução de práticas que genuinamente tratam as complexidades éticas e combatem o racismo estrutural na saúde no contexto tecnológico.

### **3.2.4 Criação das teorias**

Sob a ótica da TFD e a partir das etapas anteriores, também foi analisado o uso da abordagem antirracista como um requisito essencial no design de produtos e serviços que façam uso de inteligência artificial (IA) em projetos destinados para a área da saúde.

## **3.3 Amostra de Participantes**

Para Charmaz (2009), ao se iniciar a amostra de participantes das entrevistas com diversidade de perfis e considerações raciais, o estudo emprega uma abordagem inicial ampla. À medida que a pesquisa avança, a amostragem teórica direcionará o aprofundamento nas categorias emergentes, guiando a coleta de dados subsequentes para refinar e elaborar as teorias. Portanto, alinhado a com os princípios da Teoria Fundamentada nos Dados de Charmaz (2009), a amostragem inicial de participantes (Tabela 1) foi composta por 6 pessoas de 3 perfis, sendo: dois profissionais da área de saúde, dois designers e duas pessoas desenvolvedoras de software, sendo que todos os perfis participam ou já participaram do processo de criação e desenvolvimento de produtos de IA. Quanto a quantidade de pessoas entrevistadas, Charmaz (2009), orienta que a amostragem deve ser direcionada para a construção da teoria, em vez de buscar representatividade populacional. Nesse sentido, a saturação teórica é alcançada quando as categorias emergentes se tornam completas e os dados adicionais não mais revelam novas propriedades. Portanto, a escolha do número de participantes é baseada na necessidade de atingir a saturação teórica, garantindo a profundidade e a qualidade da análise dos dados coletados. Devido à significativa disparidade de gênero e raça no setor tecnológico, foi considerado a diversidade racial e de gênero na seleção dos participantes. Essa abordagem visa enriquecer a pesquisa com uma variedade de narrativas e evitar a perpetuação das iniquidades existentes no contexto.

Tabela 1: Perfil dos participantes

<b>Perfil</b>	<b>Código de participantes</b>	<b>Pessoa Negra</b>	<b>Pessoa Branca</b>	<b>Participantes</b>
Pessoa Designer	D01, D02	1	1	2
Pessoa desenvolvedora	DEV01, DEV02	1	1	2
Pessoa que atua na área de saúde	S01, S02	1	1	2
<b>Total de participantes:</b>				6

Fonte: Elaborado pela autora (2023).

Antes da realização das entrevistas, foi elaborado um quadro detalhado, na qual os objetivos específicos foram desmembrados em três tópicos:

Quadro 1: Objetivos específicos da pesquisa em tópicos

<b>Tópico 1</b>	<b>Tópico 2</b>	<b>Tópico 3</b>
levantar o cenário atual quanto à perspectiva prática da ética	levantar o cenário atual quanto à diversidade de pessoas nas equipes de desenvolvimento de soluções com, e de IA	levantar o cenário atual quanto à perspectiva antirracista

Fonte: Elaborado pela autora (2024).

A partir dos dados coletados, foram desenvolvidas hipóteses que contribuiriam para orientar as discussões durante as entrevistas.

Essas hipóteses preliminares foram traduzidas em perguntas, destinadas a explorar as práticas, desafios e percepções dos entrevistados em relação à ética e à abordagem antirracista no contexto da inteligência artificial na área da saúde.

Segundo Charmaz (2009), ao elaborar entrevistas semiestruturadas, é recomendado que se planejem perguntas amplas e abertas, uma vez que tendem a incentivarem os participantes a compartilhar e refletir sobre suas experiências de maneira profunda. Além disso, é preciso enfatizar a necessidade de os entrevistadores serem empáticos

durante esse processo, facilitando a expressão dos participantes e buscando compreender suas interpretações pessoais sobre as experiências relatadas.

Com as entrevistas devem ser flexíveis, permitindo explorar temas emergentes, e que as perguntas não devem ser valorativas, para não limitar as respostas dos entrevistados. A intenção é manter o equilíbrio entre conduzir a entrevista de forma aberta e focar nas áreas significativas, ajustando-se à natureza semiestruturada da entrevista para capturar detalhes ricos e profundos sobre o tema em estudo, mantendo um diálogo que, apesar de seguir um formato menos casual que as conversas do dia a dia, permita uma exploração significativa dos tópicos relevantes.

Foi elaborada uma abordagem estruturada que possibilitou uma investigação minuciosa e abrangente do cenário atual, resultando em valiosas percepções. Assim, o quadro 2 foi organizado contendo a relação entre tópicos, hipóteses e perguntas.

Quadro 2: Perguntas da entrevista, de acordo com hipótese e tópicos

<b>Tópicos</b>	<b>Hipóteses</b>	<b>Perguntas</b>
<b>Tópico 1</b>	1 - Não avaliam impactos éticos antes ou depois da etapa de desenvolvimento	<b>1 - Nas soluções em que você teve envolvimento, existiu alguma análise ética do impacto da solução?</b> <b>Se sim:</b> <b>1.1</b> Como sua equipe fez essa análise? <b>1.2</b> Se estende para além da etapa de desenvolvimento? <b>Se não: Por quê?</b>
	2 - Focam em impactos éticos apenas no algoritmo	
<b>Tópico 2</b>	3 - Consideram essencial que a equipe para criar soluções com e de IA seja diversa racialmente para garantir a qualidade da solução	<b>3 - Na sua opinião, para a qualidade da solução, faz diferença ter diversidade na equipe de desenvolvimento de soluções com ou de IA?</b> <b>3.1</b> Por quê?
<b>Tópico 3</b>	4 - Acreditam que priorizar a ética de forma generalista, sem abordar diretamente o racismo, é o suficiente para soluções com equidade racial.	<b>4 - Você ou sua equipe consideram o racismo no processo de criação de soluções com IA?</b> <b>4.1 - Se sim, como</b> você e sua equipe consideram o racismo durante o processo de criação de soluções com e de IA? <b>4.2 - Se não, por que você e sua equipe não</b> consideram?

	5 - Equipes possuem poucas ou nenhuma prática antirracista durante o processo de criação de soluções com e de IA	<b>5</b> - Poderia compartilhar sobre as práticas antirracistas aplicadas pela sua equipe durante o desenvolvimento de soluções com ou de IA?
	6 - Não aplicam prática explicitamente antirracista à solução antes e durante o desenvolvimento e depois do lançamento da solução.	<b>6</b> - Como essas práticas são integradas após o lançamento da solução?
<b>Tópico 1 e 3</b>	7 - Equipes e lideranças possuem comunicação transparente sobre ética e práticas antirracistas	<b>7.</b> Em relação à transparência, como sua equipe promove a comunicação aberta sobre as decisões éticas tomadas durante o ciclo de vida da solução com IA? <b>8.</b> Há alguma iniciativa para envolver partes interessadas externas na discussão ética? <b>9.</b> Como a equipe busca garantir que as decisões éticas sejam compreendidas e consideradas pela comunidade afetada pela solução

Fonte: Elaborado pela autora (2024).

Para a condução das entrevistas foi elaborado uma ficha de planejamento (Figura 3) contendo o detalhamento sobre as ferramentas, participantes recrutados, roteiro, tópicos contemplados e hipóteses.

Figura 3: Ficha de planejamento das entrevistas

**Temática da pesquisa**

Abordagem antirracista como pilar fundamental no design de intervenções de inteligência artificial na área de saúde

**A entrevista**

Plataforma: Google Meet

**Ferramentas:**

- Gravação de entrevistas: [tdv - Gravador com tecnologia de IA para Google Meet](#)
- Agendamento de entrevistas: Google calendar

Tipo de entrevista: semi estruturada

**Objetivos da entrevista**

- **Objetivo 1:** levantar o cenário atual quanto a perspectiva prática da ética
- **Objetivo 2:** levantar o cenário atual quanto a diversidade de pessoas nas equipes de desenvolvimento de soluções com, e de IA
- **Objetivo 3:** levantar o cenário atual quanto a perspectiva antirracista

Roteiro		
<b>Apresentação</b>	<ul style="list-style-type: none"> <li>- Oi, meu nome é Fernanda e sou pesquisadora de Design e Inteligência Artificial. Muito obrigada por participar desta entrevista. Essa entrevista procura entender como acontece o processo de criação de soluções de IA atualmente no mercado.</li> <li>- Antes de começarmos, você autoriza que essa entrevista seja gravada, tanto em vídeo quanto em áudio? A gravação dessa chamada será utilizada para fins de pesquisa.</li> <li>- <b>Preparei alguns tópicos</b> para essa entrevista e gostaria de pedir que você responda da forma <b>mais honesta possível</b> e com base em suas próprias experiências pessoais.</li> <li>- Mesmo que sejam coisas positivas ou negativas, eu gostaria de ouvir tudo.</li> </ul>	
<b>Quebra-Gelo</b>	<ul style="list-style-type: none"> <li>- Para começar, <b>gostaria que você se apresentasse</b> e me falasse um pouco sobre você, em que área atua e como se relaciona com a criação de soluções com ou de IA.</li> </ul>	
<i>Equipes ou pessoas envolvidas na criação de soluções com, ou de IA...</i>		
Tópicos	Hipóteses	Perguntas
<b>Tópico 1</b>	1 - Não avaliam impactos éticos antes ou depois da etapa de desenvolvimento 2 - Focam em impactos éticos apenas no algoritmo	1 - Nas soluções em que você teve envolvimento, existiu alguma análise ética do impacto da solução? <b>Se sim:</b> 1.1 Como sua equipe fez essa análise? 1.2 Se estende para além da etapa de desenvolvimento? <b>Se não: Por quê?</b>
<b>Tópico 2</b>	3 - Consideram essencial que a equipe para criar soluções com e de IA seja diversa racialmente para garantir a qualidade da solução	3 - Na sua opinião, para a qualidade da solução, faz diferença ter diversidade na equipe de desenvolvimento de soluções com ou de IA? 3.1 Por quê?
<b>Tópico 3</b>	4 - Acreditam que priorizar a ética de forma generalista, sem abordar diretamente o racismo, é o suficiente para soluções com equidade racial. 5 - Equipes possuem poucas ou nenhuma práticas antirracistas durante o processo de criação de soluções com e de IA. 6 - Não aplicam prática explicitamente antirracista à solução antes e durante o desenvolvimento e depois do lançamento da solução.	4 - Você ou sua equipe consideram o racismo no processo de criação de soluções com IA? <b>4.1 - Se sim, como</b> você e sua equipe consideram o racismo durante o processo de criação de soluções com e de IA? <b>4.2 - Se não, por que</b> você e sua equipe não consideram? 5 - Poderia compartilhar sobre as práticas antirracistas aplicadas pela sua equipe durante o desenvolvimento de soluções com ou de IA? 6 - Como essas práticas são integradas após o lançamento da solução?
<b>Tópico 1 e 3</b>	7 - Equipes e lideranças possuem comunicação transparente sobre ética e práticas antirracistas	7. Em relação à transparência, como sua equipe promove a comunicação aberta sobre as decisões éticas tomadas durante o ciclo de vida da solução com IA? 8. Há alguma iniciativa para envolver partes interessadas externas na discussão ética? 9. Como a equipe busca garantir que as decisões éticas sejam compreendidas e consideradas pela comunidade afetada pela solução

**Participantes da entrevista**

Pessoas envolvidas no processo de criação de novas soluções que utilizam inteligência artificial na área de saúde.

**Recrutamento de participantes**

- 1 Participante entrevista piloto;  
P1 - Pessoa designer, pessoa desenvolvedora ou pessoa da área de saúde;
- 2 pessoas designers:  
Designer 1;  
Designer 2;
- 2 pessoas desenvolvedoras:  
Dev 1;  
Dev2;
- 2 pessoas da área de saúde  
Saúde 1;  
Saúde 2;

Fonte: Elaborado pela autora (2024).

Também como parte da preparação para as entrevistas foi realizada uma entrevista Piloto, com uma sétima pessoa, visando testar a eficácia das ferramentas, ter uma estimativa da duração das entrevistas. A entrevista desta sétima pessoa, não foi considerada como parte dos resultados desta pesquisa.

### **3.5 Recrutamento dos Participantes:**

As pessoas voluntárias para participação deste estudo, foram recrutadas formalmente via ferramentas digitais como *Whatsapp*<sup>™</sup> e *Linkedin*<sup>™</sup>, e-mail ou ligação telefônica.

### **3.6 Coleta de Dados**

A análise e interpretação dos dados começam com a codificação inicial, que permite ao pesquisador identificar conceitos e categorias emergentes, a partir dos segmentos de dados (Charmaz, 2009). Essa etapa inicial foi essencial para desenvolver uma compreensão básica do fenômeno em estudo e preparou o terreno para análises mais aprofundadas. Charmaz (2009) enfatiza a importância da codificação focalizada subsequentemente, selecionando códigos significativos para uma análise detalhada, o que facilita a integração dos dados e a identificação de padrões subjacentes.

A autora destaca a escrita de memorandos como uma ferramenta analítica essencial que auxilia na elaboração de ideias emergentes e na formulação de hipóteses sobre as relações entre as categorias (Charmaz, 2009).

Portanto, conforme recomendado por Charmaz, foram elaborados memorandos durante a fase de análise dos dados coletados, o que foi fundamental para o desenvolvimento das etapas seguintes, pois permitiu refletir sobre os dados, elaborar categorias, e estabelecer conexões teóricas entre eles. Os memorandos servem como um instrumento analítico que facilita a imersão no estudo, a compreensão profunda do fenômeno investigado e a construção progressiva da teoria fundamentada nos dados (Charmaz, 2009).

Conforme destacado por Charmaz (2009), durante a coleta e análise de dados pela Teoria Fundamentada nos Dados, é essencial a avaliação contínua da qualidade e a relevância dos dados, buscando alcançar a saturação teórica por meio de uma revisão constante que visa esclarecer dúvidas e aprimorar a compreensão do fenômeno estudado.

Seguindo as recomendações prévias, a coleta e análise dos dados foi conduzida da seguinte forma:

Os dados foram coletados por meio de pesquisas históricas e bibliográficas, entrevistas semi estruturadas e memorandos. As entrevistas foram conduzidas *on-line*, utilizando tanto o Google Meet quanto a ferramenta *on-line TL;DV*, que grava e transcreve reuniões virtuais utilizando inteligência artificial. A valorização de flexibilidade e adaptabilidade na coleta e análise de dados colocada por Charmaz (2009), permite aos pesquisadores a utilização de ferramentas que facilitem e acelerem o processo de pesquisa. Dessa forma, a escolha de uma ferramenta que utiliza Inteligência Artificial para a transcrição das entrevistas foi motivada pela necessidade de otimizar o tempo disponível na pesquisa, o que possibilitou a aceleração da transcrição, considerada tradicionalmente um processo demorado. Durante a coleta de dados via entrevistas, a ferramenta *TL;DV* transcreveu as entrevistas e ao fim ofereceu alguns *insights* gerados com inteligência artificial. Ademais, estes *insights* foram adicionados aos memorandos criados pela autora. Os insights gerados pela ferramenta *TL;DV* foram devidamente sinalizados nos memorandos e consultados como complemento dos *insights* elaborados pela autora.

### **3.7 Análise e interpretação dos dados**

Os dados foram analisados seguindo as etapas definidas por Charmaz (2009) na teoria fundamentada dos dados, estes dados são coletados, codificados, categorizados, analisados e, e quando necessário, revisitados, buscando uma maior compreensão ou aprofundamento sobre o tema de pesquisa e dessa forma atingindo a saturação.

Por meio de etapas interconectadas, a análise e interpretação dos dados se desdobrou da seguinte forma:

**Coleta de dados:** Foram registradas informações a partir de entrevistas, visando capturar a riqueza e a complexidade das experiências dos participantes.

**Codificação inicial:** Após a coleta, os dados passam pela codificação inicial, um processo onde os dados foram examinados linha por linha e também por perfis de pessoas entrevistadas para identificar conceitos iniciais que emergem diretamente das informações coletadas.

**Codificação focalizada:** foram selecionados os códigos ou conceitos mais significativos identificados na etapa anterior para uma análise mais profunda, permitindo uma organização dos dados de maneira mais coerente e focada. Alguns códigos que apresentavam significados próximos foram fundidos em um único código.

**Categorização dos dados:** nesta etapa, os códigos e conceitos foram organizados em categorias mais amplas, facilitando a estruturação e a compreensão dos dados, permitindo a identificação de padrões e relações entre os elementos analisados.

**Análise dos fenômenos emergentes na categorização:** durante a categorização, os fenômenos que emergiram foram analisados, explorando a profundidade e a complexidade das categorias e como elas se relacionam com as experiências dos participantes.

**Relação entre os fenômenos encontrados:** por fim, estabelece-se a relação entre os fenômenos identificados, construindo uma teoria que explique as conexões e interações entre as categorias e os conceitos, oferecendo uma compreensão holística do fenômeno estudado.

## 4 RESULTADOS

Este capítulo tem como finalidade apresentar os resultados alcançados conforme os objetivos específicos, e por fim, contemplando o objetivo geral da pesquisa.

## 4. 1 Identificação de práticas e premissas

Para atender ao objetivo de identificar práticas e premissas, foi conduzido um benchmarking com seis frameworks (Quadro 3) que podem ser utilizados no design de produtos de IA. Durante essa investigação, existiu uma certa dificuldade em encontrar não só um framework específico de design de produtos de IA, mas que também fosse focado na área da saúde. Nenhum framework brasileiro foi encontrado. Essa lacuna destaca a necessidade urgente de desenvolver e promover recursos mais inclusivos e abrangentes para orientar o desenvolvimento ético e antirracista no campo da inteligência artificial, principalmente quando aplicado à saúde.

Quadro 3: Frameworks selecionados para o benchmarking

<b>Número identificador</b>	<b>Framework</b>	<b>Autores</b>
1	A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle	Harini Suresh, John Guttag
2	Design of an Ethical Toolkit for the Development of AI Applications	Mario Sosa Hidalgo
3	Ethical framework for Artificial Intelligence and Digital technologies	Mona Ashok a, Rohit Madan a, Anton Joha a, Uthayasankar Sivarajah b
4	To Use or Not to Use Artificial Intelligence? A Framework for the Ideation and Evaluation of Problems to Be Solved with Artificial Intelligence	Timo Sturm, Mariska Fecho, Peter Buxmann
5	People + AI Guidebook	Google - People + AI Research (PAIR)
6	Project Muse - A Proposed Framework on Integrating Health Equity and Racial Justice into the Artificial Intelligence Development Lifecycle	Irene Dankwa-Mullan, Elisabeth Lee Scheufele, Michael E. Matheny, Yuri Quintana, Wendy W. Chapman, Gretchen Jackson, Brett R. South

Fonte: Elaborado pela autora (2024).

Nos frameworks analisados (Quadro 3), foram examinados diversos aspectos relacionados ao propósito de uso. Foi investigado se eles recomendavam a colaboração entre profissionais de diferentes áreas na criação de produtos de IA, se

incentivavam a diversidade na equipe, se continham diretrizes éticas específicas e quais eram essas diretrizes, se adotavam abordagens para mitigar vieses, quais eram essas abordagens, se mencionavam explicitamente o racismo, se incluíam medidas para lidar especificamente com vieses antirracistas e se tinham foco no setor de saúde.

Após a seleção dos frameworks (Quadro 3), os dados dos frameworks selecionados foram organizados em *post-it* na ferramenta Miro™, onde foi possível visualizar os dados graficamente.

Figura 4: Processo de análise dos frameworks selecionados  
(imagem meramente ilustrativa, sem intenção de ser legível)



Fonte: Elaborado pela autora (2024).

Esta visualização gráfica permitiu uma análise detalhada de cada um deles, abordando suas contribuições e limitações em relação aos temas centrais de ética, vieses, saúde e práticas antirracistas. A análise individual buscou explorar como cada *framework* se posiciona em relação à integração de considerações éticas e antirracistas ao longo do ciclo de desenvolvimento de produtos de IA, bem como sua aplicabilidade e relevância

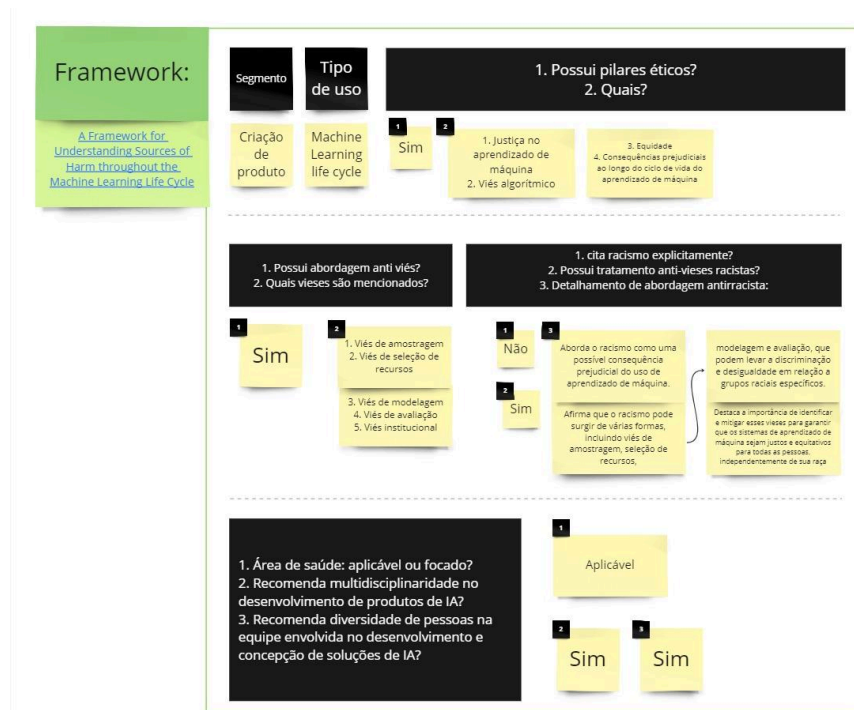
no contexto da saúde. Dada a natureza desta pesquisa, direcionou-se uma atenção especial às metodologias propostas para a mitigação de vieses, à promoção da equidade e à incorporação de diretrizes éticas e antirracista.

Além disso, foi investigado em que medida os *frameworks* recomendam a inclusão de vozes diversas e a colaboração interdisciplinar na concepção e desenvolvimento de soluções em IA. Esta análise de caráter exploratório e imersivo também focou na identificação de práticas sugeridas que possam servir como guias para a incorporação de uma consciência ética robusta e de uma perspectiva antirracista ativa. Tal atitude é fundamental para enfrentar os desafios éticos emergentes no desenvolvimento de tecnologias de IA na saúde. Além disso, a análise individualizada permitiu não apenas entender as nuances de cada proposta, mas também destacar como elas podem contribuir conjuntamente para um ecossistema de IA mais justo, ético e inclusivo na área da saúde, fornecendo *insights* valiosos para a elaboração de futuras diretrizes e *frameworks* brasileiros nesse campo.

## Análise individual dos frameworks selecionados:

- **1 - A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle (Suresh, 2021)**

Figura 5: Análise individual do framework 1.



Fonte: Elaborado pela autora (2024).

O framework 1, aborda a compreensão das fontes de danos ao longo do ciclo de vida do aprendizado de máquina. Ele propõe uma análise metódica dos riscos de um sistema, fornecendo terminologia específica para cada estágio do ciclo de vida do aprendizado de máquina. Essas terminologias incluem:

1. Viés histórico: refere-se a distorções nos dados de treinamento que refletem desigualdades históricas, resultando em modelos que perpetuam ou ampliam essas disparidades.
2. Viés de avaliação: diz respeito a distorções na avaliação de modelos de aprendizado de máquina, levando a decisões injustas ou prejudiciais para certos grupos.

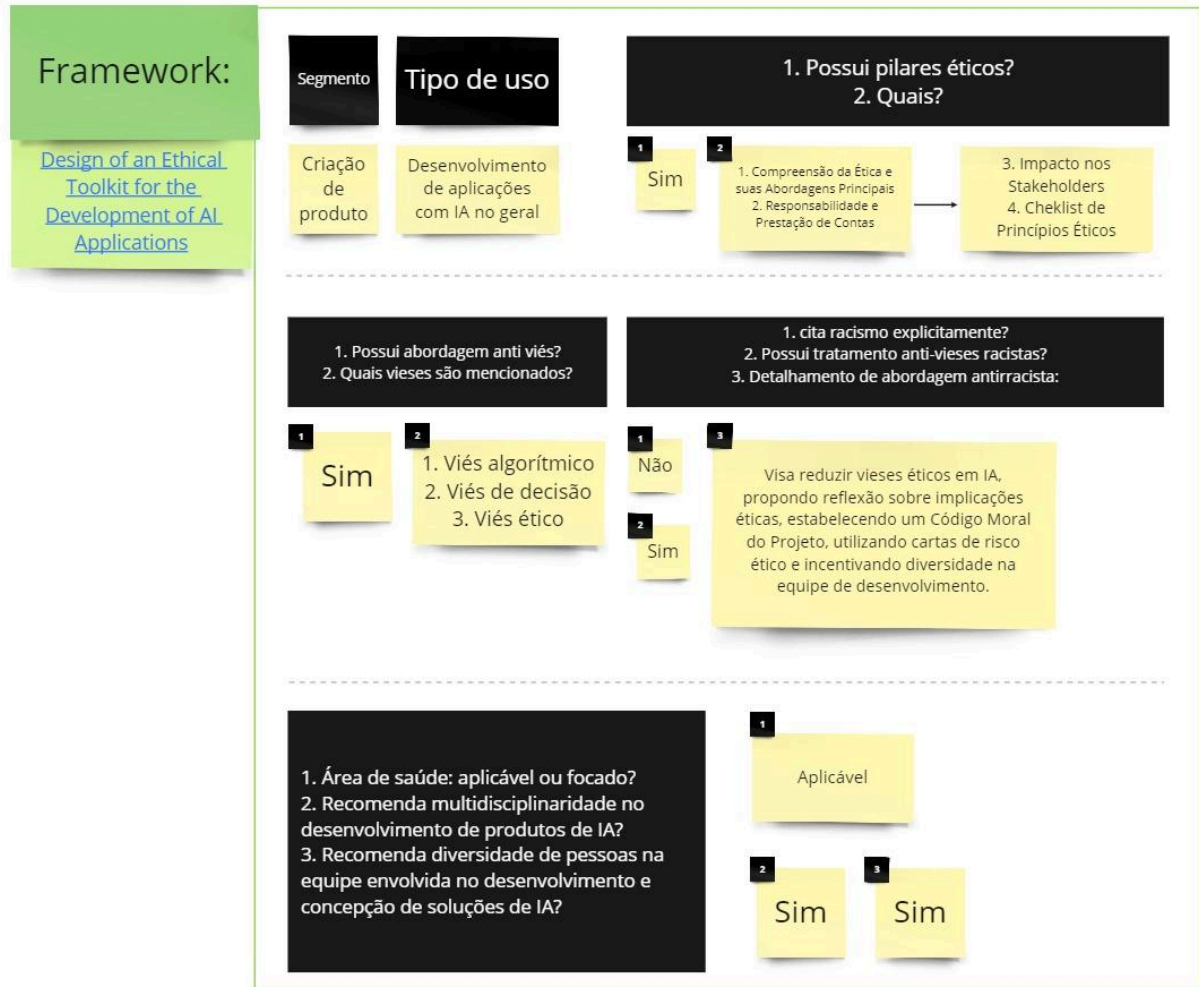
3. Dano alocativo: envolve a distribuição desigual de recursos, oportunidades ou serviços com base nas previsões ou decisões de modelos de aprendizado de máquina.
4. Dano representacional: refere-se à representação inadequada ou estereotipada de certos grupos na modelagem de aprendizado de máquina, levando a resultados discriminatórios ou prejudiciais.

Essas terminologias específicas ajudam a identificar e abordar de forma mais precisa as fontes de danos potenciais ao longo do ciclo de vida do aprendizado de máquina, facilitando uma análise mais aprofundada dos riscos e a implementação de medidas de mitigação adequadas.

Focado no ciclo de vida do aprendizado de máquina, aborda-se a criação de produtos de IA com ênfase nos pilares éticos, como justiça no aprendizado de máquina, vies algorítmico, equidade e as consequências prejudiciais ao longo do ciclo de vida do aprendizado. Esse enfoque ético é complementado pela exploração de como os vieses - incluindo o de amostragem, seleção de recursos, modelagem, avaliação, e institucional - podem ser identificados e mitigados. Embora o racismo não seja citado explicitamente como um dos vieses, o framework aborda de forma implícita a importância da abordagem antirracista ao considerar as possíveis consequências prejudiciais do uso de aprendizado de máquina, como o racismo. Ao destacar a necessidade de identificar e mitigar esses vieses, enfatiza-se a importância de garantir que os sistemas de aprendizado de máquina sejam justos e equitativos para todas as pessoas, independentemente de sua raça.

## **2 - Design of an Ethical Toolkit for the development of AI applications (Sosa hidalgo, 2019)**

Figura 6: Análise individual do framework 2.



Fonte: Elaborado pela autora (2024).

O Framework 2 é fruto de uma pesquisa de mestrado de Sosa hidalgo (2019) que aborda o desenvolvimento de um Toolkit Ético para Aplicações de Inteligência Artificial (IA) na empresa MOBGEN | Accenture Interactive. A MOBGEN é uma consultoria de serviços digitais que se concentra em estratégias móveis, criatividade e tecnologia para desenvolver soluções únicas que agregam valor às marcas. Em 2016, a MOBGEN foi adquirida pela consultoria multinacional Accenture, tornando-se parte do ramo digital europeu da empresa com o nome MOBGEN | Accenture Interactive.

Sosa hidalgo (2019) descreve que motivação para este projeto surgiu da crescente implementação da IA nas organizações, levantando questões éticas e consequências sociais. A pesquisa explora as visões éticas e preocupações futuras de todas as partes

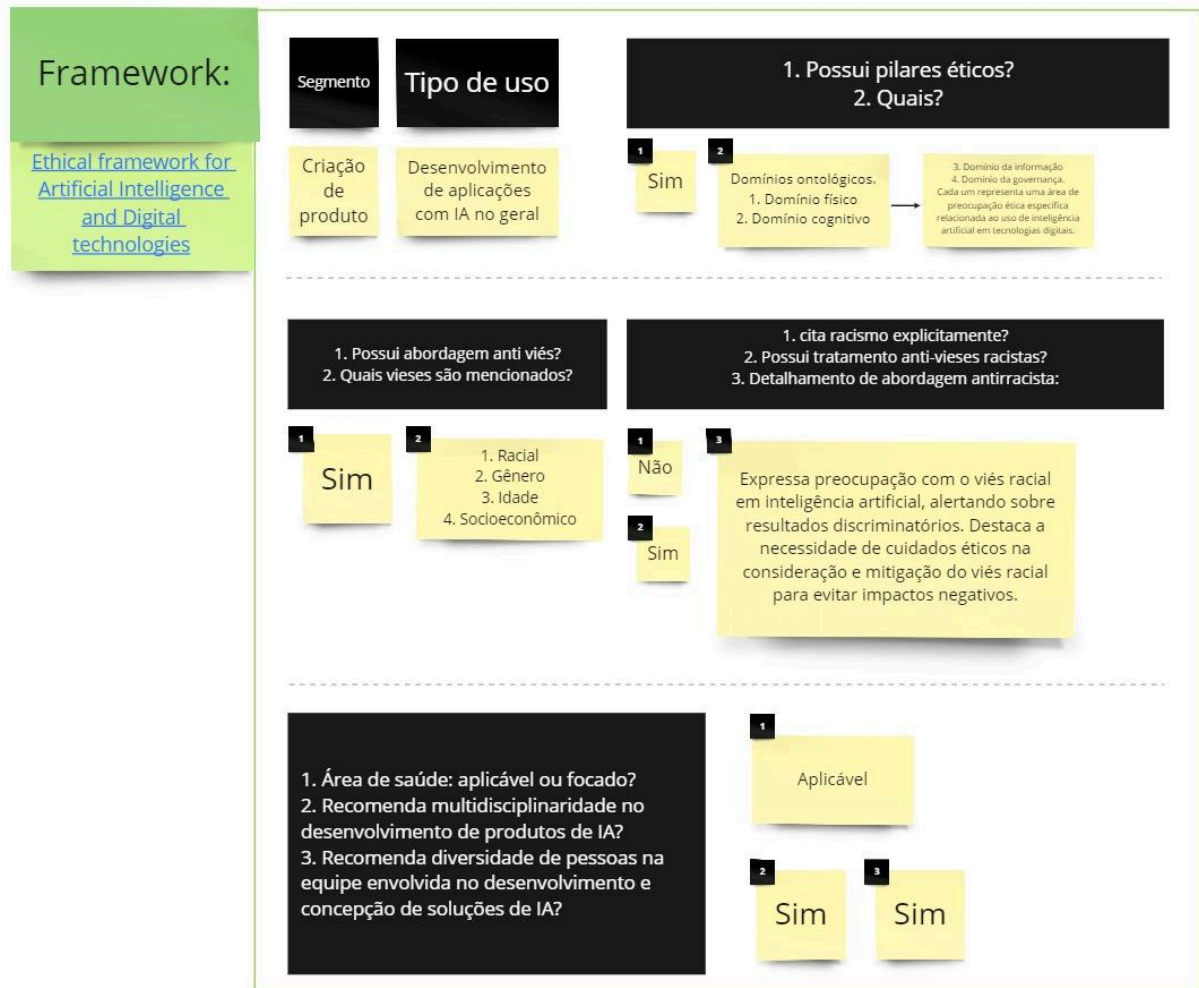
envolvidas no desenvolvimento de aplicações de IA utilizando o duplo diamante como abordagem metodológica de design e ciclos de Pesquisa-Ação para gerar conscientização sobre o tema. Foram realizadas revisões literárias, entrevistas aprofundadas e pesquisas empíricas para compreender o contexto ético da IA, resultando na criação de um framework teórico e um toolkit ético para auxiliar as equipes de desenvolvimento da MOBGEN | Accenture Interactive na avaliação ética de aplicações de IA. O toolkit foi validado em sessões com a empresa e outros *stakeholders*, fornecendo recomendações para o desenvolvimento futuro, como a criação de frameworks éticos, a medição do impacto ético de sistemas de IA e o aprimoramento da pesquisa sobre ética da IA sob uma perspectiva de design.

Por meio de uma abordagem proativa e abrangente na redução de vieses éticos em projetos de Inteligência Artificial (IA), tem-se o objetivo de promover uma reflexão cuidadosa sobre as implicações éticas inerentes ao desenvolvimento de aplicações de IA. As diretrizes e princípios centrais do framework incluem o estabelecimento de um Código Moral do Projeto criado em equipe, que define princípios éticos claros para orientar as ações da equipe de desenvolvimento, assegurando que todas as decisões reflitam valores éticos sólidos. Adicionalmente, a utilização de Cartas de Risco Ético permite a identificação e mitigação de potenciais problemas éticos via análises proativas de risco, ajudando a equipe a antecipar e evitar possíveis dilemas éticos antes que eles ocorram.

O *framework* também enfatiza o incentivo à diversidade na equipe de desenvolvimento, reconhecendo a importância de incorporar diferentes perspectivas e experiências no processo de criação. Essa diversidade é vista como fundamental para abordar de maneira eficaz e inclusiva as questões éticas, contribuindo para o desenvolvimento de soluções de IA mais éticas, inclusivas e justas. Ao promover um ambiente que valoriza a diversidade de ideias e experiências, o framework busca garantir que as tecnologias de IA sejam desenvolvidas com uma preocupação genuína com a equidade e a justiça, refletindo um comprometimento com a criação de produtos que beneficiem toda a sociedade, sem reforçar vieses ou discriminações preexistentes.

- **3 - Ethical framework for Artificial Intelligence and Digital technologies** (Ashok et al., 2022)

Figura 7: Análise individual do framework 3.



Fonte: Elaborado pela autora (2024).

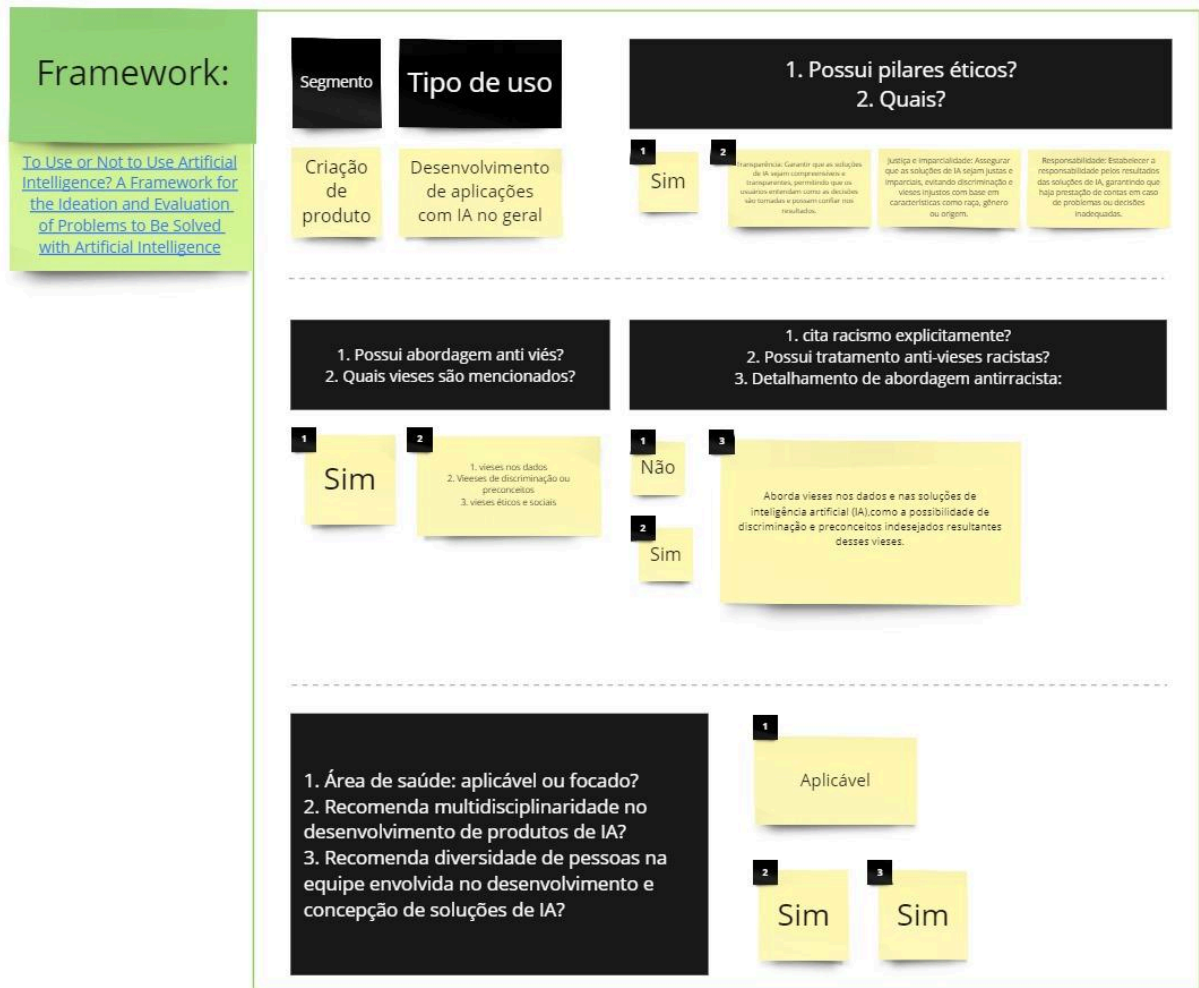
Este framework visa abordar às implicações éticas associadas ao uso da IA oferecendo uma estrutura conceitual para entender e orientar a implementação responsável dessas tecnologias. O estudo ressaltou a adoção de princípios de IA por governos e acadêmicos, mas apontou a falta de direcionamento prático para a implementação desses princípios. Mediante uma revisão sistemática da literatura de 59 artigos, identificou-se 14 implicações éticas do uso de IA em sete arquétipos de tecnologias digitais, utilizando um novo quadro ontológico que abrange as dimensões física,

cognitiva, informacional e de governança. Este é considerado o primeiro estudo a fazer tal identificação, apresentando um modelo conceitual com doze proposições que destacam o impacto das implicações éticas no impacto societal, moderado pelos arquétipos de tecnologias digitais e mediado pelo impacto organizacional.

O framework também expressa preocupação com o viés racial em inteligência artificial, alertando sobre os resultados discriminatórios que podem surgir de algoritmos e sistemas baseados em IA. Além disso, o artigo enfatiza a importância da diversidade nas equipes de desenvolvimento de IA para identificar e abordar questões éticas de forma abrangente. A inclusão de diversas perspectivas contribui para a criação de soluções mais éticas e inclusivas, refletindo uma genuína preocupação com a equidade e justiça no desenvolvimento de tecnologias de IA. Destaca-se que, enquanto a IA tem o potencial de transformar a sociedade, seu desenvolvimento e aplicação exigem uma consideração cuidadosa das implicações éticas para evitar resultados não intencionais e prejudiciais, reforçando a necessidade de ferramentas e frameworks que orientem a implementação responsável de IA em contextos organizacionais e sociais.

- **4 - To Use or Not to Use Artificial Intelligence? A Framework for the Ideation and Evaluation of Problems to Be Solved with Artificial Intelligence (Sturm et al., 2021)**

Figura 8: Análise individual do framework 4.



Fonte: Elaborado pela autora (2024).

Sturm et al. (2021) propõe um framework para auxiliar organizações a identificar e avaliar problemas que podem ser resolvidos por Inteligência Artificial (IA), especialmente por meio de aprendizado de máquina (ML). Baseando-se em entrevistas com 24 especialistas, o estudo reconheceu que, enquanto a solução de problemas tem sido tradicionalmente considerada uma atividade humana, a crescente capacidade das IA em superar humanos em determinadas tarefas exige uma reavaliação de como os

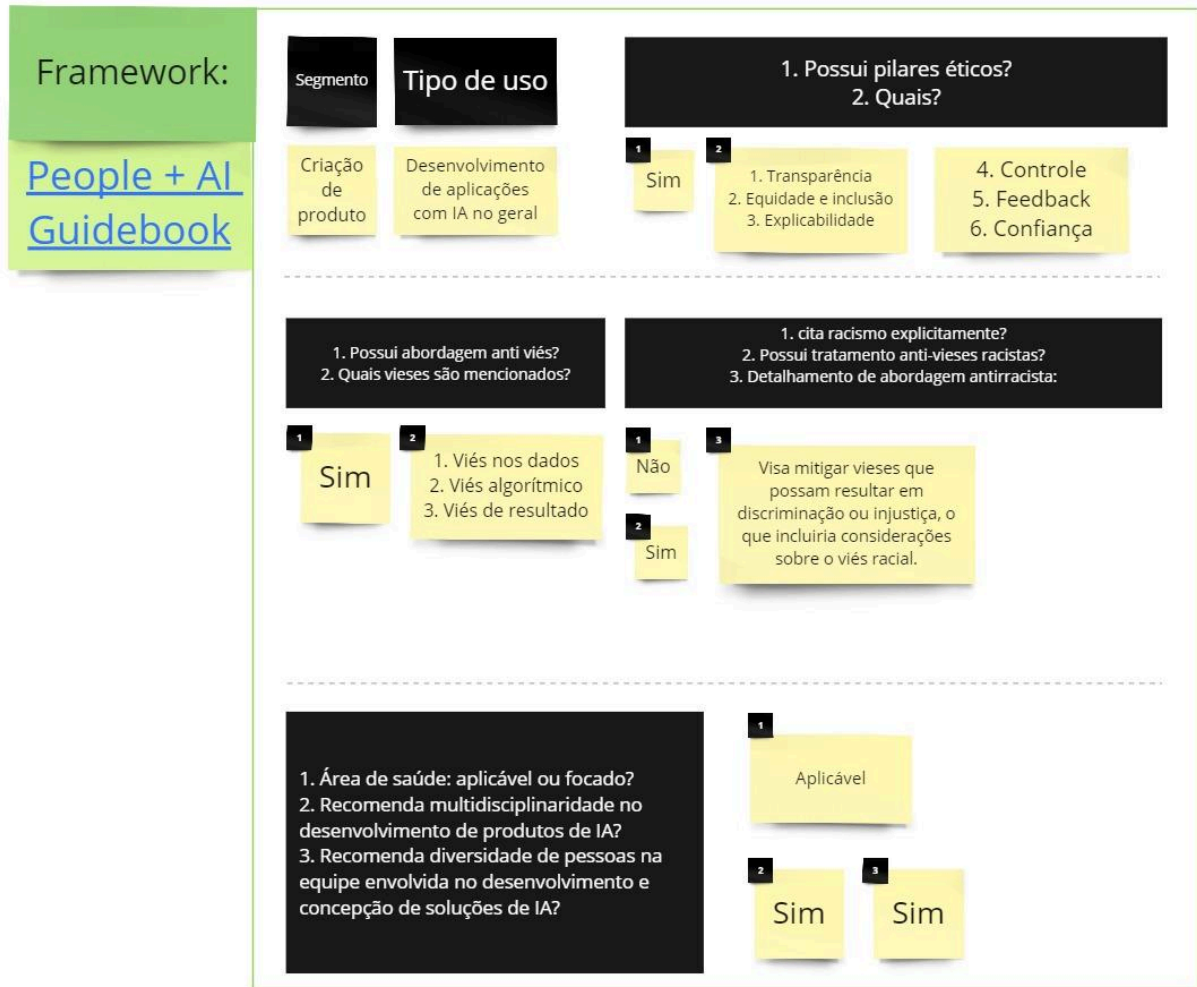
problemas são identificados e solucionados. O framework é construído sobre a premissa de que a identificação de problemas adequados para IA envolve compreender a clareza organizacional do propósito, a disponibilidade de dados relevantes e particularidades técnicas do aprendizado de máquina, seguindo fases de ideação e avaliação para garantir a relevância e viabilidade da solução proposta pela IA.

Durante a fase de ideação, os autores destacam a importância de explorar tanto o propósito organizacional quanto os dados disponíveis, sugerindo duas abordagens: foco no propósito ou nos dados. Isso envolve avaliar se um problema tem um propósito real e relevante para a organização e se os dados necessários para alimentar uma solução de IA estão disponíveis e são adequados. A fase de avaliação é dividida em avaliar a substância do problema para verificar se é intrinsecamente adequado para ser solucionado por IA. Ademais, deve-se avaliar as particularidades do problema, que dizem respeito a fatores como tolerância a erros, necessidade de transparência, justiça, adaptabilidade da solução e tempo de resposta. Essa abordagem sistemática visa não apenas a identificação de problemas adequados para soluções de IA, mas também a criação de soluções mais eficazes e eticamente responsáveis, alinhadas com os objetivos organizacionais e as capacidades técnicas disponíveis.

Apesar de o framework não abordar explicitamente a diversidade e práticas antirracistas, a inclusão de justiça como um dos critérios de avaliação sugere uma preocupação com a equidade nas soluções de IA. A atenção à justiça e à transparência pode contribuir para o desenvolvimento de tecnologias que evitem discriminações, inclusive raciais, ao assegurar que vieses sejam identificados e mitigados.

- 5 - People + AI Guidebook (People + AI Research Team, 2019)

Figura 9: Análise individual do framework 5.



Fonte: Elaborado pela autora (2024).

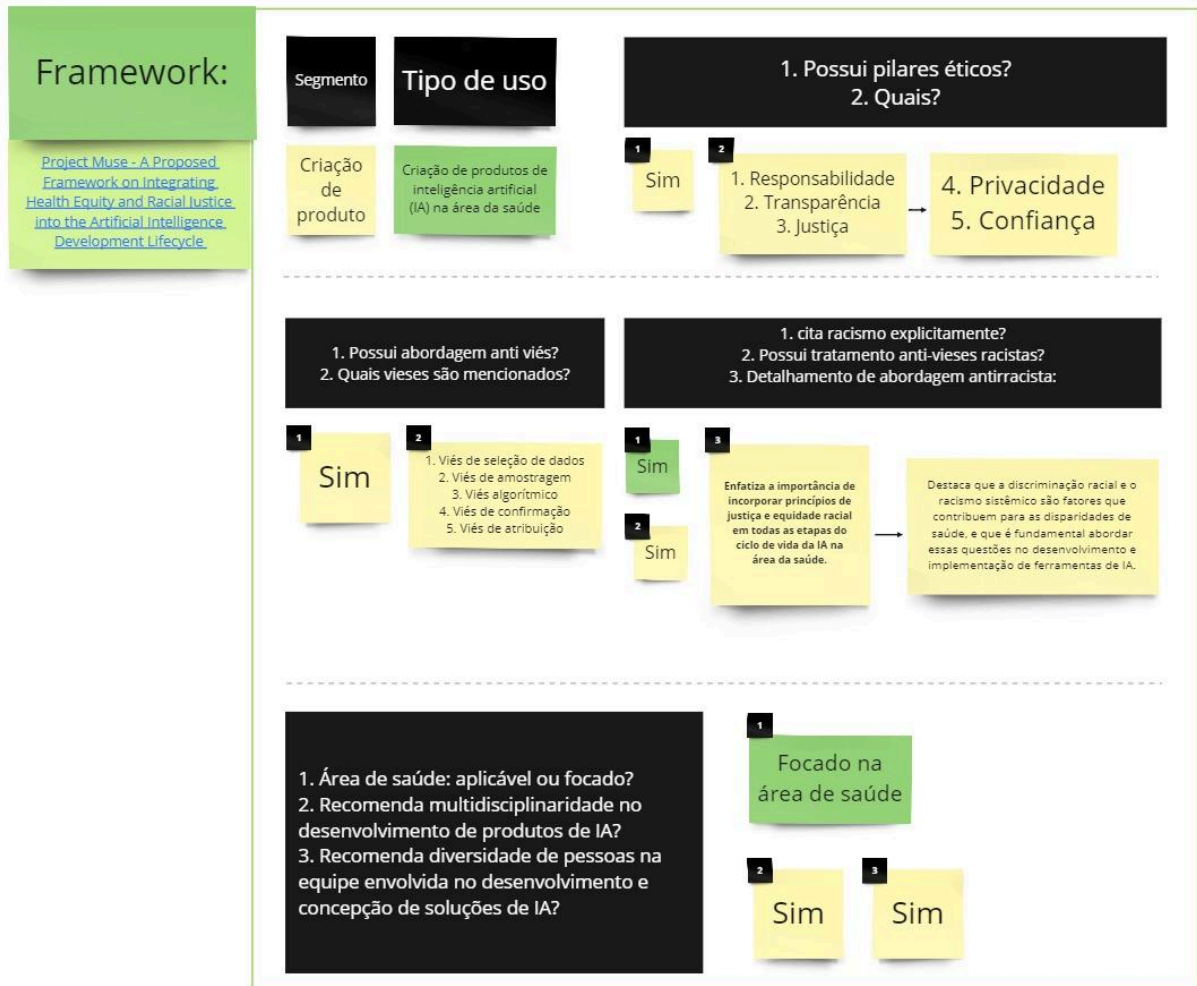
O "People + AI Guidebook" do Google PAIR (People + AI Research Team, 2019) é um framework abrangente projetado para orientar o desenvolvimento de produtos de IA de maneira ética e responsável, enfatizando a importância de atender às necessidades do usuário e definir o sucesso de forma clara. Ele sugere um processo de desenvolvimento centrado no ser humano, instigando a reflexão sobre problemas que a IA está singularmente posicionada para resolver, como a automação de tarefas desagradáveis ou a ampliação das capacidades humanas. Um dos aspectos-chave destacados é a

necessidade de uma função de recompensa bem projetada, que determine o que a IA considera como sucessos e falhas, influenciando diretamente a experiência do usuário. Embora o guia seja aplicável a uma variedade de setores, incluindo a saúde, ele serve como uma fundação geral que pode ser adaptada conforme as necessidades específicas de cada área.

O guia enfatiza pilares éticos como transparência, justiça e a prevenção de vieses, embora não especifique explicitamente "pilares éticos". Ressaltou a importância da diversidade nas equipes de desenvolvimento e na inclusão de diferentes perspectivas para criar soluções mais justas e inclusivas. A abordagem anti-viés é enfatizada, com a recomendação de considerar cuidadosamente a seleção de dados e o design do modelo para evitar a perpetuação de vieses existentes. Recomenda-se fortemente a multidisciplinaridade no desenvolvimento de produtos de IA e a diversidade dos envolvidos, visando a promoção de soluções equitativas que beneficiem amplamente a sociedade. A inclusão de diversas perspectivas e a recomendação de diversidade entre os envolvidos no processo de desenvolvimento sugerem um esforço para criar soluções que sejam justas e representativas da população na totalidade. Isso pode naturalmente incluir a consideração de questões raciais e a promoção de práticas antirracistas. No entanto, a abordagem antirracista requer uma intenção e ações específicas para identificar, entender e combater o racismo e suas manifestações dentro dos sistemas de IA, algo que pode ir além das recomendações gerais de diversidade e inclusão.

- **6 - Project Muse - A Proposed Framework on Integrating Health Equity and Racial Justice into the Artificial Intelligence Development Lifecycle (Dankwa-Mullan et al., 2021)**

Figura 10: Análise individual do framework 6.



Fonte: Elaborado pela autora (2024).

Dankwa-Mullan et al. (2021) apresenta um *framework* destinado a integrar a equidade de saúde e a justiça racial no desenvolvimento de tecnologias de Inteligência Artificial (IA), especialmente em contextos de saúde. Desenvolvido em resposta às disparidades reveladas pela pandemia de COVID-19, que afetou desproporcionalmente comunidades raciais e étnicas minoritárias e populações socialmente desfavorecidas, este framework

visa antecipar e mitigar desigualdades sociais e de saúde exacerbadas pela IA. Ele se apoia em pilares éticos como responsabilidade, justiça e inclusão, adotando uma postura explicitamente anti vieses para combater as discriminações algorítmicas e institucionais, além de enfatizar a importância das avaliações de impacto algorítmico. A abordagem antirracista é um ponto central, incentivando práticas proativas para assegurar tratamento e resultados equitativos entre todas as raças e etnias. O *framework* recomenda ainda a multidisciplinaridade e a diversidade na equipe de desenvolvimento, promovendo a inclusão de variadas perspectivas que enriquecem o processo criativo e garantem a relevância e eficácia das soluções de IA na saúde. Destinado especificamente ao setor de saúde, busca garantir que as inovações tecnológicas beneficiem equitativamente todos os segmentos da população, contribuindo para a construção de um futuro mais justo e equitativo no uso de IA em cuidados de saúde.

Por fim, a percepção geral da análise dos 6 frameworks revelou que todos os frameworks selecionados são direcionados para a criação de produtos de IA. Além disso, apenas o “Framework 1” abrange o ciclo de vida de machine learning e os demais o desenvolvimento de aplicações de IA de forma mais ampla. Entre os seis frameworks avaliados, apenas um é especificamente voltado para o setor de saúde (Framework 6), enquanto os demais têm um escopo mais geral. Mesmo que de forma não explícita, todos os frameworks incorporam pilares éticos em suas diretrizes, sendo que a equidade ou justiça é citada como um dos principais pilares em todos. Além disso, todos promovem abordagens práticas para mitigar vieses, visando garantir a imparcialidade e eficácia das soluções. Apenas o “Framework 6” menciona explicitamente o racismo e também promove práticas específicas para combater vieses racistas, destacando-se nesse aspecto.

Um ponto comum entre todos os frameworks é a recomendação enfática da multidisciplinaridade e diversidade das equipes envolvidas no desenvolvimento de soluções de IA. Eles reconhecem a importância de contar com profissionais de diferentes áreas e *backgrounds* para garantir uma abordagem holística e inclusiva no processo de criação e desenvolvimento. Essa ênfase na diversidade de perspectivas e

experiências reflete um compromisso com a construção de soluções mais robustas e éticas, capazes de atender às necessidades variadas dos usuários e da sociedade como um todo.

## 4.2 Principais resultados a partir das entrevistas

No cenário atual em relação à perspectiva prática da ética e da abordagem antirracista entre os profissionais envolvidos na criação de soluções que utilizam inteligência artificial, foram conduzidas entrevistas com 6 participantes conforme o Tabela 1.

A partir das transcrições e seguindo as etapas da TFD, os dados foram inicialmente organizados em códigos iniciais, posteriormente focalizados e, por fim, agrupados em categorias emergentes. Para realizar essa organização e desenvolver os códigos iniciais (Quadro 5) e focalizados (Quadro 6), foram realizadas análises detalhadas das respostas das pessoas entrevistadas para cada pergunta, complementadas por memorandos (Quadro 4) registrados pela pesquisadora durante as entrevistas e revisões das gravações.

Quadro 4: Resumo de memorandos criados durante as entrevistas

Trecho da transcrição	<b><i>Insights criados por IA com a ferramenta TL;DV</i></b>	<b>Memorando e <i>insights</i> criados pela autora</b>
<p><i>“olhando da perspectiva do usuário, ele não vai querer saber se a equipe que fez o produto são pessoas brancas, negras, azuis, amarelo, ele vai querer ter uma boa experiência naquilo que ele tá fazendo” (DEV01)</i></p> <p><i>“quando você quer, por exemplo, detectar o rosto, as aplicações que tem nos celulares tem, por exemplo, desbloqueio de tela com Reconhecimento facial, você não quer que uma pessoa asiática seja não tem o rosto reconhecido em relação a uma</i></p>	<p>Comenta sobre mostrar bons resultados inicialmente para depois corrigir problemas, como estratégia de startups. Reforça a ideia de mostrar resultados positivos inicialmente para depois melhorar aspectos ruins. (DEV01)</p> <p>Revela que nas soluções de IA onde trabalhou, não foram realizadas análises éticas de impacto, enfatizando uma maior atenção aos dados. Destaca a importância de equipes diversificadas e práticas</p>	<ul style="list-style-type: none"> <li>● Distanciamento de responsabilidade ética</li> <li>● Descrença em diversidade nas equipes</li> <li>● Fluxo de trabalho focado em dados</li> <li>● Desejo por equidade racial</li> <li>● Equidade racial apenas no contexto de visão computacional</li> </ul>

<p>“pessoa americana ou vice-versa” (DEV01)</p> <p>“Então primeiro tem que mostrar valor no sentido de mostrar o quanto a sua aplicação é boa em determinada coisa e depois acho que é muita da mentalidade “a gente faz isso aqui funcionar.”” (DEV01)</p> <p>“Atualmente no nosso fluxo de trabalho não ela é mais voltada para a parte dos dados o antes, né? E a gente não vê muito o depois assim.” (DEV02)</p>	<p>antirracistas específicas no reconhecimento de imagens, como o balanceamento de dados para melhor representar a diversidade. (DEV02)</p>	
<p>“A diversidade ela é fundamental. Não apenas para saúde como para outras áreas também, né? Trabalhamos com profissionais da engenharia na área de ciências da área de humanas então a diversidade ela tem papel fundamental na construção, inclusive desse banco de dados nós sabemos que aí a ela trabalha com o banco de dados que é pré-fornecido, né para que ela possa realmente através seja do aprendizado de máquina, seja de redes neurais.” (S01)</p> <p>“Não, não consideramos racismo não. Até porque também o escopo foi muito fechadinho e específico, mas é algo que a gente sempre tenta porque a gente sempre vê muito material aí que sai a questão do racismo.” (S02)</p>	<p>Afirma que a diversidade é fundamental. (S01)</p> <p>Responde ter considerado aspectos éticos, mas sem avaliação externa. Diz ter considerado aspectos éticos nos projetos seguindo recomendações da área, mas sem avaliação externa. (S02)</p> <p>Diz que diversidade é importante.(S02)</p> <p>seus projetos não consideraram racismo por serem modelos com escopo fechados e específicos. (S02)</p>	<ul style="list-style-type: none"> <li>● Diversidade como pilar fundamental</li> <li>● Conhecimento sobre importância de práticas antirracistas</li> <li>● Falta de integração de práticas antirracistas</li> </ul>
<p>Diversidade: “são questões que não passaram na minha cabeça a princípio e que pessoas da equipe falaram “tem uma situação tal que a gente tem que pensar nessa pessoa X”, aí foi bem legal, quando é uma equipe bem diversa.” (D01)</p>	<p>Confirma que discutiram impactos éticos, focando na diversidade no banco de dados. (D01)</p> <p>Discute a importância da diversidade na equipe e no banco de dados e práticas antirracistas adotadas. (D01)</p>	<ul style="list-style-type: none"> <li>● Importância de diversidade nos dados</li> <li>● Importância de diversidade nas equipes</li> <li>● Diversidade nas equipes como reflexo da diversidade de pessoas na empresa</li> </ul>

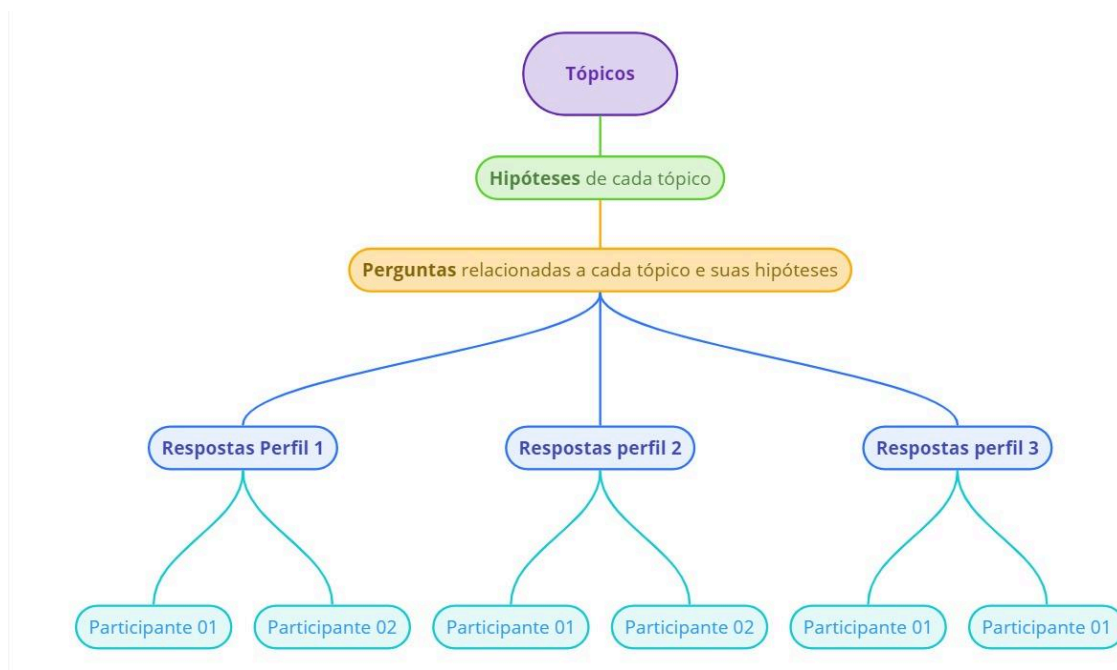
<p><i>“Minha impressão pessoal é de que acho que a fase do Design é que talvez mais pense na parte ética, e depois só na questão de negócio, sabe? Porque eles sabem que podem perder dinheiro, se não tiver legalmente correto. Mas não por uma preocupação genuína” (D02)</i></p>	<p>Não consegue lembrar de outras práticas antirracistas adotadas além da diversidade. (D01)</p> <p>Nenhuma ação ou tarefa específica antirracista foi acordada para o futuro. (D01)</p> <p><i>acredita que a diversidade na equipe melhora a qualidade da solução. (D02)</i></p>	<ul style="list-style-type: none"> <li>● Visão de práticas antirracistas apenas no desenvolvimento</li> <li>● Falta de preocupação genuína com a ética</li> <li>● Diversidade dá visibilidade para reflexões e narrativas diversas</li> </ul>
---	---	---

Fonte: Elaborado pela autora (2024).

Com a intenção de imergir na análise dos dados, foi criado o Quadro 6, que ilustra a aplicação da codificação inicial, uma etapa indispensável nesse processo. Conforme Charmaz (2009) articula, a análise inicia com uma imersão profunda nos dados coletados, onde a codificação inicial desempenha um papel fundamental ao identificar conceitos emergentes que refletem a riqueza e as múltiplas camadas presentes nas experiências vividas pelas pessoas entrevistadas. Esta fase é seguida pela codificação focalizada e a categorização dos dados, estruturando assim a fundação para uma compreensão mais aprofundada dos fenômenos em estudo.

Utilizou-se um mapa mental como ferramenta de design, elaborado na plataforma Miro™, que serviu como um auxílio visual essencial na organização dos dados obtidos durante as entrevistas. Esse recurso possibilitou uma visualização clara das relações entre os temas discutidos, as hipóteses formuladas e as perguntas relevantes, simplificando o acompanhamento das diversas respostas e percepções dos participantes entrevistados. A utilização dessa estratégia de mapeamento mental se revelou crucial para a estruturação eficiente dos dados de forma visual, apoiando significativamente a fase de codificação inicial e oferecendo flexibilidade para ajustes na direção da investigação à medida que novas compreensões surgiam a partir dos dados. Esse método de organização visual promoveu uma conexão profunda com os dados, facilitando a imersão nas nuances das perspectivas, reflexões e impressões dos entrevistados e evidenciando a relação dinâmica entre as experiências dos participantes e o arcabouço teórico adotado.

Figura 11: Mapa mental para organização visual dos dados coletados nas entrevistas.



Fonte: Elaborado pela autora (2024).

Com a estruturação dos dados coletados de forma visual (Figura 11), a codificação inicial foi realizada no quadro Miro™, onde trechos das transcrições foram examinados e associados a conceitos-chave. Este processo permitiu a construção da codificação inicial visualizada no quadro 5.

Quadro 5: Codificação inicial

Trechos das transcrições	Codificação inicial
<p>"Análise do <i>impacto ético é uma das questões que mais nos instiga</i> isso porque nós estamos trabalhando com algo denominamos <i>Inteligência Artificial</i>." (S01)</p> <p>"Será que realmente é possível tratarmos como <i>inteligência ou não? Isso porque nós temos um conjunto de algoritmo que vão trabalhar, né? Com 0 e 1.</i>" (S01)</p>	<p>1. Reflexão sobre a ética na Inteligência Artificial</p>
<p>"Conseguimos eliminar aí grande parte desse <i>viés racista, viés misógino ou até mesmo de comportamento indesejados para a sociedade, mas 100% ainda não.</i>" (S01)</p>	<p>2. Preocupação com viés e discriminação:</p>

<p>"A máquina ela tende a repetir esse comportamento. Então, em alguns momentos o próprio sistema se tornou racista, se tornou misógeno, se formou extremamente criminoso em outros aspectos." (S01)</p> <p>"A base de dados a gente tem que estar tentando até hoje montar ela de uma forma mais igualitária." (D01)</p>	
<p>"Embora a diversidade na equipe tenha sido considerada importante para a qualidade, os potenciais impactos do racismo não foram examinados profundamente durante o projeto." (S02)</p>	3. Importância da diversidade na equipe de desenvolvimento
<p>"É preciso que exista um acompanhamento após o lançamento também dessa ferramenta, né isso porque a partir do momento que eu tenho uma ferramenta que a proporcione o diagnóstico." (S01)</p> <p>"Então primeiro tem que mostrar valor no sentido de mostrar o quanto a sua aplicação é boa em determinada coisa e depois acho que é muita da mentalidade "a gente faz isso aqui funcionar." (DEV01)</p>	4. Necessidade de análise contínua pós-lançamento
<p>"Não discutimos a aplicação de quaisquer práticas antirracistas explícitas durante o desenvolvimento de soluções de IA." (S02)</p> <p>"Mais de saber que forma a gente abordaria as pessoas para participar das pesquisas (que iam ajudar a treinar o modelo) e que fosse ético mesmo, porque assim eu não posso chegar e dizer: me ajuda aqui de graça, mas no final vou ganhar dinheiro com isso, é complicado, sabe" (D01)</p>	5. Desafios na abordagem ética da IA
<p>"Esse processo iterativo de melhoria que podem incluir melhorias relacionadas a questões de preconceito, discriminação, etc." (DEV01)</p> <p>"Então você balancear os dados para que eles sejam representativos para a população." (DEV02)</p>	6. Processo iterativo de melhoria e correção de viés
<p>"Práticas antirracistas, acho que nunca. Nunca vi nem uma conversa sobre isso, sabe?" (D02)</p> <p>"Eu acho que pelo menos a empresa que eu tô não tem nenhuma iniciativa nessa área." (D01)</p> <p>"Não discutimos a aplicação de quaisquer práticas antirracistas explícitas durante o desenvolvimento de soluções de IA." (S02)</p> <p>"Eu não me lembro de nenhuma outra ação antirracista, além de buscar diversidade nos dados, que a gente tem usado no projeto." (D01)</p>	7. Falta de ações antirracistas explícitas
	8. Desafios na inclusão de diversidade na

<p>"A equipe teve essa <i>formação diversa por coincidência</i>, mas <i>não teve uma abordagem guiar vaga específica para pessoas "tal" não teve</i>, foi coincidência mesmo." (D01)</p>	<p>equipe de desenvolvimento</p>
<p>"Acredito que, por exemplo, <i>depende do contexto</i>, por exemplo, se eu tô no Brasil e <i>no Brasil a gente tem um contexto muito diverso</i>." (DEV01)</p> <p>"A máquina ela tende a repetir esse comportamento. <i>Então, realmente em alguns momentos o próprio sistema se tornou racista</i>, se tornou misógino, se formou extremamente criminoso em outros aspectos." (S01)</p>	<p>9. Conscientização sobre a importância da ética na IA</p>
<p>"<i>Não discuti a aplicação de quaisquer práticas antirracistas explícitas durante o desenvolvimento de soluções de IA</i>." (S02)</p> <p>"A máquina ela tende a repetir esse comportamento." (S01)</p> <p>"<i>Então nós conseguimos Estabelecer alguns produtos com a esse ditame ético</i>, mas ainda não conseguimos resolver a questão do viés." (S01)</p>	<p>10. Desafios na abordagem ética durante o desenvolvimento:</p>
<p>"Você tem pessoas pretas, brancas, azul, amarela misturado todo jeito, então <i>no Brasil no contexto do Brasil seria realmente super importante</i>." (DEV01)</p>	<p>11. Impacto da cultura e diversidade na IA:</p>
<p>"Toda a empresa que se preze de inteligência artificial, ela <i>tem que ter um processo de manutenção do modelo de IA</i>, isso é algo que é esperado, é o normal." (DEV01)</p> <p>"<i>Se ficar bom de primeira, o engenheiro vai achar que tá errado então</i>. É então o processo iterativo faz parte do processo de desenvolvimento artificial." (DEV01)</p>	<p>12. Desafios na manutenção de modelos de IA</p>
<p>"Quando você começa a provocar esse sistema com algumas informações que você começa a fornecer dados direcionados, <i>ele ainda possui ali limitações para entender</i>." (S01)</p> <p>"Que não é aceito na sociedade, tá? <i>Faz parte do processo de treinamento, né?</i>" (S01)</p>	<p>13. Limitações na compreensão da sociedade pela IA</p>
<p>"Eu acho que pelo menos a empresa que eu tô <i>não tem nenhuma iniciativa nessa área</i>." (D01)</p> <p>"A empresa também <i>não tinha nada e ninguém assim ninguém nem propôs, nem outras pessoas da equipe me propôs nenhuma outra ação</i>." (D01)</p>	<p>14. Desafios na implementação de práticas antirracistas</p>
<p>"A máquina ela tende a repetir esse comportamento. Então, realmente em alguns momentos o próprio sistema se tornou racista, se tornou misógino, se formou extremamente criminoso em outros aspectos." (S01)</p>	<p>15. Desafios na conscientização sobre viés e discriminação</p>

<p>"Eu acho que pelo menos a empresa que eu tô <i>não tem nenhuma iniciativa nessa área.</i>" (D01)</p> <p>"A empresa também <i>não tinha nada e ninguém assim ninguém nem propôs, nem outras pessoas da equipe me propôs nenhuma outra ação.</i>" (D01)</p>	16. Desafios na implementação de práticas éticas
<p>"A equipe teve essa <i>formação diversa por coincidência, mas não teve uma abordagem guiar vaga especifica para pessoas "tal" não teve, foi coincidência mesmo.</i>" (D01)</p> <p>"Para ser sincero. Assim, na prática, <i>talvez não</i> porque assim do jeito que a coisa funciona é o seguinte, no final das contas o que <i>eles querem saber tipo dar uma determinada é o quanto eu preciso é aquele modelo</i>" (DEV01)</p>	17. Desafios na conscientização sobre diversidade na equipe
<p>"Toda a empresa que se preze de <i>inteligência artificial, ela tem que ter um processo de manutenção do modelo de IA, isso é algo que é esperado, é o normal.</i>" (DEV01)</p> <p>"Se ficar bom de primeira, o engenheiro vai achar que tá errado então. É então o processo iterativo faz parte do processo de desenvolvimento artificial." (DEV01)</p>	18. Desafios na manutenção de ética na IA
<p>"Quando você começa a provocar esse sistema com algumas informações que você começa a fornecer dados direcionados, <i>ele ainda possui ali limitações para entender.</i>" (S01)</p> <p>"Que não é aceito na sociedade, tá? <i>Faz parte do processo de treinamento, né?</i>" (S01)</p>	19. Desafios na compreensão da sociedade pela IA
<p>"Práticas antirracistas acho que nunca. <i>Nunca vi nem uma conversa sobre isso, sabe?</i>" (D02)</p> <p>"Eu acho que pelo menos a empresa que eu tô <i>não tem nenhuma iniciativa nessa área.</i>" (D01)</p>	20. Desafios na implementação de práticas antirracistas explícitas

Fonte: Elaborado pela autora (2024).

Na sequência do processo de análise dos dados, avançamos para a fase de codificação focalizada. Esta etapa, fez-se essencial para aprofundar nossa compreensão dos dados coletados, envolveu uma seleção cuidadosa e uma análise mais detalhada dos códigos iniciais que se mostraram particularmente significativos. A codificação focalizada permitiu concentrar a atenção nas áreas que emergem como centrais para os fenômenos estudados, promovendo uma integração temática que guia a direção subsequente da pesquisa. O Quadro 6, ilustra esta etapa ao destacar códigos

focalizados como "Viés Racial em IA", "Práticas Antirracistas no Desenvolvimento de IA", "Ética e Impacto Social em IA", e "Melhoria Contínua de Modelos de IA". Cada código reflete uma síntese das discussões, práticas, e reflexões identificadas durante a análise dos dados, enfatizando a importância da ética, da equidade e da responsabilidade no desenvolvimento de tecnologias de IA. Este enfoque na codificação focalizada é fundamental para desvelar as complexidades dos temas em estudo e para construir uma base sólida para a elaboração de teorias substanciais que emergem da investigação.

Quadro 6: Codificação focalizada

<b>Código focalizado</b>	<b>Descrição</b>
1. Viés Racial em IA	Inclui discussões específicas sobre o viés racial em sistemas de IA e suas implicações.
2. Práticas Antirracistas no Desenvolvimento de IA	Refere-se às ações concretas e estratégias adotadas para promover a equidade e combater o racismo em soluções com IA.
3. Ética e Impacto Social em IA	Envolvem reflexões mais aprofundadas sobre a ética e o impacto social das soluções de IA considerando a diversidade na equipe como um fator crucial.
4. Melhoria Contínua de Modelos de IA	Destaca a importância do desenvolvimento iterativo e da diversidade nos dados de treinamento para aprimorar os modelos de IA.

Fonte: Elaborado pela autora (2024).

Após a criação dos códigos iniciais e focalizados, esses códigos deram origem as categorias (Quadro 7), com base também na incidência nas transcrições.

Com os dados categorizados, a análise revela fenômenos emergentes, permitindo uma exploração detalhada das interações e relações intrínsecas. Finalmente, essa abordagem metodológica conduz à identificação das conexões entre os fenômenos observados, culminando na elaboração de teorias substanciais que capturam a essência do estudo. O Quadro 7, exemplifica essa trajetória analítica, demonstrando como a partir dos dados transcritos das entrevistas, meticulosamente codificados, foi possível aprofundar-se em reflexões sobre a ética na inteligência artificial, preocupações com viés e discriminação, e outros temas pertinentes, seguindo os passos delineados por Charmaz (2009).

Quadro 7: Categorias finais

<b>Categoria</b>	<b>Descrição</b>
1. Viés e Discriminação em IA	Agrupa os códigos relacionados ao viés racial, discriminação e a necessidade de práticas antirracistas no desenvolvimento de IA.
2. Ética e Impacto Social em Desenvolvimento de IA	Engloba as discussões sobre ética, impacto social e a importância da diversidade na equipe para lidar com questões éticas e sociais.
3. Desenvolvimento Iterativo e Diversidade em Modelos de IA	Refere-se ao processo contínuo de melhoria dos modelos de IA destacando a diversidade nos dados de treinamento como um elemento fundamental.
4. Equidade e Inclusão em IA	Inclui conversas sobre a promoção da equidade e inclusão no desenvolvimento de soluções com IA considerando a diversidade como um fator essencial para resultados justos.
5. Transparência e Responsabilidade em IA	Refere-se à importância da transparência e responsabilidade nos processos de desenvolvimento de IA para garantir sistemas éticos, não discriminatórios e compreensíveis.

Fonte: Elaborado pela autora (2024).

Para a criação das codificações inicial, focalizada e também a criação de categorias foi utilizada a ferramenta Miro™ (Figura 12), para melhor visualização dos dados transcritos a partir das entrevistas.

Figura 12: Processo de codificação e categorização  
(imagem meramente ilustrativa, sem intenção de ser legível)



Fonte: Elaborado pela autora (2024).

### 4.3 Descrição de fenômenos

A partir da análise da relação entre as categorias encontradas, foram identificados seis fenômenos, estes representam aspectos essenciais das discussões e reflexões dos participantes sobre o tema em estudo. Cada um reflete uma dimensão distinta das questões abordadas, contribuindo para uma compreensão mais completa e aprofundada do assunto. Este capítulo visa fornecer uma visão abrangente e fundamentada sobre as complexidades inerentes ao tema, informando e enriquecendo o debate acadêmico e prático. Os fenômenos encontrados foram os seguintes:

→ Viés e Discriminação em Sistemas de IA:

“A inteligência artificial é reflexo da visão de mundo do homem, o algoritmo não conhece o argumento moral. [...] Se os dados são de baixa qualidade, o resultado também será ruim.” (Winetzki, 2020).

Os entrevistados demonstraram uma preocupação significativa em relação ao viés e discriminação em sistemas de Inteligência Artificial (IA). Destacaram a possibilidade de os algoritmos reproduzirem comportamentos humanos, incluindo viés e discriminação, ressaltando a importância de evitar tais padrões nos sistemas desenvolvidos. Além disso, a complexidade ética envolvida na IA foi enfatizada, com menções à dificuldade de modelar sistemas éticos diante da diversidade de critérios e variáveis presentes no comportamento humano, como cultura e religião. A necessidade de considerar o contexto cultural na aplicação da ética em IA foi reconhecida, evidenciando a importância de uma abordagem sensível e contextualizada para mitigar possíveis impactos negativos.

Porém, observou-se uma diversidade de pontos de vista em relação à consideração do racismo no desenvolvimento de soluções de IA. Enquanto alguns entrevistados ressaltaram a importância de práticas antirracistas e reflexões sobre possíveis impactos discriminatórios, outros reconheceram a falta de discussão explícita sobre essas questões durante o desenvolvimento de projetos de IA. A necessidade de aprimorar a consideração dos potenciais impactos do racismo nos projetos futuros foi destacada, indicando uma oportunidade de melhoria na abordagem ética e inclusiva no campo da Inteligência Artificial. As percepções dos entrevistados apontam para a complexidade ética e a importância de práticas antirracistas no desenvolvimento de soluções de IA ressaltando a necessidade de uma abordagem sensível e contextualizada para promover a equidade e a não discriminação nesse campo em constante evolução.

Também é uma preocupação a percepção entre algumas pessoas entrevistadas de que o viés racial tende a ser presente exclusivamente em produtos de inteligência artificial que envolvem a área de visão computacional, particularmente os produtos que tem funcionalidade voltadas para a identificação de tons de pele e o reconhecimento facial.

→ Práticas Antirracistas no Desenvolvimento de IA na área de saúde:

As pessoas entrevistadas revelaram uma ampla gama de opiniões em relação às práticas antirracistas no desenvolvimento de IA na área de saúde. Alguns mencionaram a importância de considerar o racismo durante o processo de criação de soluções de IA, porém não citaram práticas antirracistas específicas adotadas em seus projetos. Outros entrevistados admitiram não ter discutido explicitamente práticas antirracistas durante o desenvolvimento de soluções de IA na área de saúde, indicando uma possível falta de conhecimento ou implementação dessas práticas.

Almeida (2019) enfatiza que transformações sociais não ocorrem somente por meio de denúncias ou condenações morais ao racismo; elas exigem a implementação de atitudes e práticas antirracistas.

A análise presente no artigo “Rede Negras In Tech: apropriação de tecnologias por mulheres negras como estratégias de resistência” (Lima et al., 2020) revelam como a interação e o estabelecimento de comunidades online podem ampliar a visibilidade e a voz das mulheres negras no campo tecnológico. A rede demonstra um ecossistema de projetos que, ao se conectarem, criam um espaço de resistência, aprendizado e troca, essencial para a promoção de uma tecnologia inclusiva e antirracista (Lima et al., 2020). Esse fenômeno ilustra a necessidade de repensar os paradigmas no desenvolvimento de tecnologias de IA na saúde, incorporando práticas que reconheçam e valorizem as contribuições de grupos historicamente marginalizados.

As percepções sobre práticas antirracistas no desenvolvimento de IA na área de saúde revelam uma necessidade de maior conscientização e implementação de ações específicas para combater o racismo e promover a equidade nos sistemas de IA. A falta de menção a práticas antirracistas específicas sugere uma lacuna na consideração e implementação de medidas concretas para mitigar o viés e a discriminação racial nos produtos de IA na área de saúde. Essa constatação destaca a importância de promover uma abordagem mais ativa e consciente em relação às práticas antirracistas no desenvolvimento de soluções de IA visando garantir a equidade e a justiça no campo da saúde digital.

Ao incorporar considerações éticas e antirracistas desde a concepção até a implementação dos modelos de IA, é possível reduzir a possibilidade de reprodução de

preconceitos ou de valores de referência com base racista na medicina, além de garantir que as soluções sejam equitativas e justas para todos os grupos demográficos. Questões como privacidade dos dados, base de dados já racista, diagnósticos imprecisos devido a viés algorítmico e a necessidade de assegurar que as soluções sejam aplicadas de forma justa e igualitária em diferentes comunidades são fundamentais para promover a confiança e a eficácia dos sistemas de IA na área da saúde.

A conscientização sobre a importância das práticas antirracistas no desenvolvimento de soluções de IA na área de saúde também envolve a necessidade de avaliar constantemente os impactos sociais e éticos desses sistemas, não só durante o desenvolvimento, mas também durante o funcionamento útil deste produto. A busca por equidade e justiça na aplicação de tecnologias de IA na saúde requer uma abordagem proativa para identificar e mitigar os possíveis vieses e discriminações.

→ Ética e Impacto Social da IA:

Racismo e machismo são as maiores barreiras. No Brasil, a questão racial é uma das mais críticas, e isso se reflete também no setor tecnológico, ampliando a desigualdade quando esse campo, que é tão poderoso, se torna exclusivo a um grupo padrão: homens, jovens, brancos e de classe média (Bahia, 2019).

O fenômeno da ética e impacto social da Inteligência Artificial (IA) foi um dos destaques nos registros das entrevistas por meio das reflexões sobre as implicações éticas do desenvolvimento de soluções com IA. Um dos entrevistados (*D02*) ressalta a importância de considerar se as ferramentas de IA replicam o racismo ou qualquer outro tipo de preconceito, indicando a necessidade de abordar questões éticas de forma proativa. Essa preocupação ética é fundamental para garantir que os sistemas de IA não perpetuem vieses e discriminações, contribuindo para um impacto social positivo e equitativo.

Por exemplo, um dos entrevistados (*S01*) da área de saúde mencionou a necessidade de voltar às etapas iniciais do desenvolvimento do sistema para aprimorá-lo e evitar

comportamentos indesejados, como o viés racista. Destacou também a importância de considerar a ética e a aplicação de práticas antirracistas durante o processo de criação de soluções com IA evidenciando a preocupação com o impacto social dessas tecnologias. Essas reflexões demonstram a relevância de abordar questões éticas, como o racismo, de forma proativa no desenvolvimento de IA visando promover a equidade, a transparência e a responsabilidade social na aplicação dessas tecnologias.

→ **Diversidade e Inclusão na Equipe de Desenvolvimento:**

Conforme o relatório “O impacto e o futuro da Inteligência Artificial no Brasil” realizado pelo Google for Startups em parceria com a empresa Box1824 e com a Abstartups, falta diversidade nas empresas de I.A. Ademais, quanto maior a diversidade em uma empresa, se efetivamente colocada em prática, a organização tende a alcançar um leque mais amplo de soluções inovadoras.

É impossível falar de futuro sem considerar as populações historicamente marginalizadas e excluídas. Esse futuro está sendo construído agora e essas pessoas precisam estar fazendo esse presente também. A tecnologia pode ser muito transformadora no campo social, mas também pode acirrar as desigualdades se não houver inclusão (Bahia, 2019)

O reconhecimento da importância da diversidade na equipe de desenvolvimento de Inteligência Artificial (IA) também é evidenciado nas transcrições por meio das ponderações que surgiram sobre a relevância de se ter uma equipe diversificada para garantir uma abordagem mais ampla e inclusiva na criação de soluções tecnológicas. Um dos entrevistados (S01) destacou que a diversidade é fundamental não apenas para a área da saúde, mas também para outras áreas, ressaltando que profissionais de diferentes formações contribuem para a construção de soluções mais eficazes e alinhadas com as necessidades variadas dos usuários. Essa valorização da diversidade na equipe de desenvolvimento reflete a compreensão de que diferentes perspectivas e experiências enriquecem o processo de criação de tecnologias, possibilitando a consideração de uma variedade de pontos de vista e necessidades.

Por exemplo, ao discutir a influência da diversidade na equipe de desenvolvimento de soluções com IA, os entrevistados reconhecem que a diversidade é a coluna vertebral do desenvolvimento desses produtos. A consideração da diversidade não se limita apenas ao aspecto técnico, mas também abrange a ética e a inclusão, garantindo que as soluções sejam desenvolvidas para atender a uma ampla gama de usuários e a refletir a diversidade da sociedade.

→ Desenvolvimento Iterativo e Melhoria Contínua:

Este fenômeno foi evidenciado por meio das discussões propostas pelas pessoas entrevistadas sobre a importância de um processo iterativo para aprimorar os sistemas de IA. Uma das pessoas entrevistadas (*DEV01*) destaca que é esperado que o modelo inicial seja aprimorado ao longo do tempo, incluindo melhorias relacionadas a questões como preconceito e discriminação. Essa abordagem ressalta a necessidade de diversidade nos dados de treinamento para garantir a representatividade e a equidade dos modelos, bem como a constante revisão e aprimoramento dos sistemas para mitigar possíveis vies e garantir a eficácia das soluções tecnológicas.

Por exemplo, ao discutir o processo iterativo de melhoria dos modelos de IA, os entrevistados reconhecem que a evolução contínua dos sistemas é essencial para garantir a qualidade e a eficácia das soluções tecnológicas. A ênfase na revisão constante e no aprimoramento dos modelos destaca a importância de identificar e mitigar possíveis vies e discriminação, assegurando que os benefícios das tecnologias de IA sejam acessíveis e equitativos para todos os usuários. Essas reflexões ressaltam a necessidade de um processo iterativo e incremental no desenvolvimento de IA, considerando a evolução dos sistemas e a busca por soluções cada vez mais éticas, inclusivas e eficazes.

→ Equidade e Transparência em IA:

Foi ressaltado durante as entrevistas a importância de considerar o vies presente nos dados de treinamento dos algoritmos de IA além de reconhecer que essas informações podem refletir preconceitos existentes na sociedade. Ademais, enfatizam a necessidade de incluir uma variedade de exemplos e perspectivas nos conjuntos de dados para

evitar a exclusão ou marginalização de grupos minoritários. Essas discussões são fundamentais para garantir que os sistemas de IA sejam desenvolvidos de forma ética e não discriminatória, promovendo a equidade e a inclusão em suas aplicações.

Outro aspecto relevante é a importância da transparência e responsabilidade no desenvolvimento de tecnologias de IA. Os entrevistados destacam a necessidade de adotar medidas proativas para mitigar o viés nos algoritmos, como a inclusão de marcações explícitas para indicar quando uma imagem ou texto foi gerado por inteligência artificial. Essas práticas éticas visam promover a transparência e conscientização sobre a origem dos dados e resultados produzidos pelos sistemas de IA, contribuindo para uma maior responsabilidade na utilização dessas tecnologias. Portanto, a equidade e a transparência emergem como pilares fundamentais para o desenvolvimento de sistemas de IA éticos e não discriminatórios, capazes de promover um ambiente mais justo e inclusivo.

## **5 DISCUSSÃO**

As relações entre os fenômenos encontrados revelam um panorama complexo das interações entre poder, responsabilidade, viés racial e práticas antirracistas no processo de desenvolvimento e criação de soluções de IA com participação de profissionais de diversas áreas do conhecimento. A percepção de poder e alcance global por parte das pessoas entrevistadas, cria uma dinâmica em que a responsabilidade ética é diluída, sendo que muitas vezes os profissionais envolvidos responsabilizam a empresa financiadora dos projetos, levando a um afastamento ético em relação aos produtos criados. Esta desconexão entre a percepção individual de responsabilidade e o impacto global das tecnologias desenvolvidas destaca a necessidade de uma abordagem mais coletiva e consciente no processo de criação de IA.

A negação de responsabilidade ocorre na tentativa de enquadramento da questão como externa a sua ferramenta, como apenas uma reprodução de estereótipos existentes (Silva, 2022, p. 84).

Mesmo com a importância de práticas antirracistas durante a criação e desenvolvimento de produtos de IA no discurso de todas as pessoas entrevistadas, foi

identificado na análise dos dados a ausência de práticas antirracistas integradas. Desta forma, salienta-se que existe uma lacuna na percepção ética e prática de designers, profissionais da área de saúde e pessoas desenvolvedoras que contribuem nesse processo de criação de produtos de IA. Almeida (2019) ressalta que o silêncio diante do racismo torna a pessoa ética e politicamente responsável pela manutenção do racismo, enfatizando a importância da tomada de posturas e adoção de práticas antirracista.

Além disso, a falta de conhecimento e aplicação de práticas antirracistas por parte também das pessoas designers ressalta a necessidade urgente de educação e conscientização sobre a importância da inclusão de perspectivas antirracistas no desenvolvimento de tecnologias além de práticas anti-vieses adotadas no processo de programação e treinamento de máquina.

Isso significa que quando o conhecimento sobre práticas antirracistas não é compartilhado de forma ampla e integrada com a diversidade dos dados e técnicas de treinamento de máquina, acaba-se revelando uma grande falha na compreensão dos efeitos do preconceito racial nos sistemas de IA. Em outras palavras, é crucial não apenas estar ciente das práticas antirracistas, mas também aplicá-las de forma eficaz nos dados e métodos de treinamento dos modelos de inteligência artificial para evitar a replicação ou ampliação de viés racial nos sistemas. Isto alerta para outro ponto que é a falta de consideração sobre a possibilidade desde os dados já estarem contaminados por informações racistas, principalmente por meio de padrões adotados historicamente na medicina em relação a diagnósticos, o que indica a necessidade de uma análise mais abrangente e crítica dos fundamentos dos modelos de IA. A reflexão sobre esta lacuna pode orientar a ação para promover uma IA mais equitativa e inclusiva.

A análise dos fenômenos identificados destacou também a importância de teorizar as conexões entre as esferas de vidas dos desenvolvedores, profissionais da área de saúde, designers e as estruturas sociais mais amplas que influenciam o desenvolvimento de soluções de IA. Esta interconexão ressalta a necessidade de considerar as implicações éticas e sociais em todas as etapas do processo de criação tecnológica para cada produto.

Com base nos dados e análise interpretativa apresentada até aqui, é possível identificar uma série de lacunas e desafios significativos no desenvolvimento ético e inclusivo de

soluções de inteligência artificial (IA). Analisando também as hipóteses em paralelo às transcrições, é adicionado uma camada de compreensão sobre onde essas lacunas podem estar e como podem ser abordadas de forma mais eficaz.

Uma das principais lacunas identificadas é a falta de avaliações específicas de impacto ético antes e depois da etapa de desenvolvimento. Embora haja uma preocupação geral com a ética, a ausência de uma prática formalizada ou sistemática de análise ética sugere a necessidade de implementação de processos mais robustos nessa área. Isso inclui a consideração de impactos éticos não apenas no algoritmo, mas em todas as áreas relacionadas ao desenvolvimento e implementação de soluções de IA. Priorizar a ética de forma generalista pode não ser suficiente para abordar as questões específicas relacionadas ao racismo e à inclusão racial.

Além disso, a falta de diversidade racial nas equipes de desenvolvimento é uma questão crítica que impacta diretamente a qualidade e abordagem das soluções. A ausência de práticas antirracistas e discussões sobre diversidade evidencia uma lacuna significativa que precisa ser abordada urgentemente. É essencial reconhecer que a diversidade de perspectivas é fundamental para mitigar o viés racial e promover a equidade nas soluções de IA.

Da Hora (2023), em entrevista à revista GALILEU, abordou a questão dos estereótipos predominantes no campo da computação, pois é um espaço majoritariamente ocupado por homens brancos, que “não falam com ninguém e ficam no computador, destacando a necessidade de diversificar o perfil dos profissionais na tecnologia”. Além disso, Da Hora (2023) enfatizou a importância crítica de incluir pessoas negras em posições de liderança nas decisões sobre o desenvolvimento e a criação de tecnologias, argumentando que a ausência dessa diversidade contribui para a perpetuação de um abismo significativo no setor.

Um caminho para contribuir, para estimular a presença das pessoas negras nesse mercado, é formar essas pessoas, ajudar essas pessoas a criarem conexões. Também falar para essas empresas que não vai dar para ser como sempre foi. Todo mundo quer ter o selo da diversidade, mas existe uma estrutura para receber essas pessoas? Existe uma

política antirracista e antimachista na empresa? Como a gente fomenta essa cultura também? (Bahia, 2019).

A falta de aplicação explícita de práticas antirracistas ao longo do processo de desenvolvimento e após o lançamento das soluções de IA é uma lacuna crítica que precisa ser abordada. Integrar considerações antirracistas em todas as fases do desenvolvimento tecnológico é essencial para garantir que as soluções sejam verdadeiramente inclusivas e equitativas.

Em resumo, os resultados destacam uma série de desafios e oportunidades no desenvolvimento ético e inclusivo de soluções de IA. Abordar esses desafios e oportunidades requer uma abordagem holística que inclua processos formais de avaliação ética, diversidade racial nas equipes de desenvolvimento, abordagem direta do racismo e aplicação explícita de práticas antirracistas em todas as fases da concepção, desenvolvimento e implementação das soluções.

Argumentou que a verdadeira inovação deve transcender aspectos técnicos e econômicos, aspirando a ser socialmente revolucionária e desafiadora do *status quo*. Uma inovação deve ser avaliada não apenas por seu resultado imediato, mas principalmente pelo impacto social que promove, contribuindo para a construção de uma sociedade que valorize a inclusão, a equidade, a justiça social e a solidariedade. Enfatizou ainda que as inovações do século XXI devem ser projetadas visando promover esses valores fundamentais. (Oliveira, 2020)

## **6 CONSIDERAÇÕES FINAIS**

### **6.1 Conclusão**

Este trabalho explorou a criação e desenvolvimento ético e inclusivo de soluções de inteligência artificial (IA), com foco na identificação de práticas e premissas que norteiam a criação de produtos de IA, especialmente na área da saúde. Por meio de um *benchmarking* de *frameworks* e entrevistas com profissionais envolvidos no

desenvolvimento de IA, emergiram desafios significativos e a necessidade urgente de abordagens mais inclusivas e explicitamente antirracistas.

Foi constatado que, apesar da existência de frameworks que promovem a ética e a equidade como pilares fundamentais no desenvolvimento de IA, há uma lacuna evidente no que diz respeito à abordagem específica de questões raciais e à inclusão de práticas antirracistas. A análise realizada nessa pesquisa revelou que apenas um dos frameworks selecionados é direcionado especificamente para o setor da saúde, e somente um aborda explicitamente o racismo, indicando uma oportunidade para o desenvolvimento de recursos mais focados nesse contexto.

As entrevistas com profissionais da área destacaram a importância da diversidade de equipes e da inclusão de múltiplas perspectivas no processo de criação de soluções de IA. Contudo, foi identificado que práticas antirracistas explicitamente integradas são raramente aplicadas, apontando para uma desconexão entre o reconhecimento da importância dessas práticas e sua efetiva implementação.

A falta de avaliações de impacto ético específicas, tanto antes quanto depois da etapa de desenvolvimento, emergiu como um desafio crítico. Isso sugere a necessidade de estabelecer processos formais e sistemáticos de análise ética que contemplem não apenas o algoritmo, mas todas as áreas envolvidas no desenvolvimento e na implementação de soluções de IA.

A ausência de diversidade racial nas equipes de desenvolvimento foi identificada como uma lacuna significativa, impactando a qualidade e a abordagem das soluções desenvolvidas. A implementação de práticas antirracistas e a promoção da diversidade emergem como aspectos fundamentais para mitigar o viés racial e garantir a equidade nas soluções de IA.

A pesquisa ressaltou a importância da aplicação explícita de práticas antirracistas em todas as fases do desenvolvimento de produtos de IA, desde a concepção até a implementação e pós-lançamento da solução ou produto. É essencial que estas considerações sejam integradas de forma holística, envolvendo processos formais de avaliação ética, diversidade racial nas equipes de desenvolvimento, e uma abordagem direta ao racismo e às práticas antirracistas. Por fim, mesmo não sendo uma aplicação direta neste trabalho, a perspectiva filosófica do *Transition Design* foi inspiração e motor

crucial para a abordagem desta pesquisa. Conforme discutido por Tonkinwise (2015), esta abordagem enfatiza a necessidade de abordar as crises sociais contemporâneas por meio de práticas de design que promovam mudanças sistêmicas e sustentáveis. Portanto, este conceito proporcionou uma lente crítica para examinar e questionar as práticas e estruturas existentes.

Este estudo contribui para o debate acadêmico e prático sobre ética na IA, destacando desafios, oportunidades e a necessidade urgente de uma abordagem mais ética, inclusiva e responsável no desenvolvimento de soluções de IA. Através da conscientização, educação e ação coletiva, podemos promover uma IA mais equitativa e justa para todos.

## **6.2 Trabalhos futuros e direções estratégicas**

A partir dos resultados e discussões apresentados, este capítulo delinea direções estratégicas para pesquisas futuras e sugere abordagens práticas visando superar os desafios identificados e fomentar o avanço de soluções de IA que sejam não apenas inovadoras, mas também éticas, inclusivas e explicitamente antirracistas.

Primeiramente, pesquisas futuras devem ter foco no desenvolvimento e validação de *frameworks* específicos para o design de produtos de IA na saúde que integrem explicitamente considerações éticas e antirracistas. Isso inclui a criação de diretrizes detalhadas para avaliação de impacto ético que abordem especificidades do contexto da saúde, garantindo que as soluções desenvolvidas promovam equidade e não perpetuem vieses ou discriminações.

Outro campo fértil para pesquisa é a investigação de metodologias e ferramentas que facilitem a incorporação da diversidade nas equipes de desenvolvimento de IA. Os estudos podem explorar estratégias eficazes para recrutamento, retenção e capacitação de talentos de grupos sub-representados, bem como examinar o impacto da diversidade de equipe nas fases de concepção, desenvolvimento e implementação de soluções de IA.

Além disso, é crucial avançar no desenvolvimento de técnicas e algoritmos que minimizem vieses raciais e promovam a justiça algorítmica. As pesquisas podem se concentrar em abordagens inovadoras para o treinamento de modelos de IA.

Uma área relacionada que merece atenção é a avaliação contínua de soluções de IA após o lançamento, com foco em monitorar e mitigar impactos éticos negativos e vieses que possam surgir com o uso contínuo. Isso implica o desenvolvimento de marcos regulatórios e mecanismos de governança que assegurem transparência, responsabilidade e revisão ética contínua.

Na prática, organizações, designers, desenvolvedores e todos os envolvidos na criação de produtos de IA, devem adotar uma abordagem proativa para a inclusão de práticas antirracistas e éticas em todas as etapas do desenvolvimento de produtos de IA. Isso inclui a implementação de treinamentos específicos sobre ética em IA e antirracismo para todas as equipes envolvidas, bem como a criação de espaços de diálogo e reflexão crítica sobre as implicações sociais das tecnologias desenvolvidas.

Para facilitar a adoção dessas práticas, recomenda-se o estabelecimento de parcerias entre acadêmicos, indústria, organizações governamentais e não governamentais, visando o compartilhamento de conhecimentos, recursos e melhores práticas. Essas colaborações podem impulsionar a inovação responsável e garantir que os desenvolvimentos em IA atendam às necessidades e respeitem os direitos de todas as partes interessadas.

Como consideração final, faz-se necessário que as futuras iniciativas de pesquisa e implementação de IA estejam alinhadas com um comprometimento robusto em relação à ética, inclusão e justiça social, servindo também como um meio para alcançar objetivos de Governança Ambiental, Social e Corporativa (ESG) nas empresas. Isso implica uma visão integrada que entende a IA não meramente como uma inovação tecnológica, mas como uma força capaz de moldar e melhorar os aspectos sociais, econômicos e culturais, alinhando-se às metas de sustentabilidade e responsabilidade social corporativa. Ao seguir estas diretrizes estratégicas, as organizações podem demonstrar como a tecnologia de IA, desenvolvida e aplicada conscientemente, é fundamental para promover um futuro mais sustentável, equitativo e ético, contribuindo significativamente para os pilares ESG dentro do contexto empresarial e mercadológico.

## REFERÊNCIAS

ALMEIDA, S. **Racismo estrutural**. São Paulo: Jandaíra, 2019.

ASHOK, M; MADAN, R; JOHA, A; SIVARAJAH, U. Ethical framework for artificial intelligence and digital technologies. **International Journal of Information Management**, v. 62, 102433, 2022. Disponível em: <https://doi.org/10.1016/j.ijinfomgt.2021.102433>. Acesso em: 05 dez. 2023

BAHIA, S. **[Entrevista] – “O mercado de tecnologia pode ser um lugar para as mulheres negras”**. 9 de outubro de 2019. Disponível em: <http://inciti.org/2019/10/09/entrevista-o-mercado-de-tecnologia-pode-ser-um-lugar-para-as-mulheres-negras/>. Acesso em: 01 fev. 2024.

BENTO, C. **O pacto da branquitude**. São Paulo: Companhia das Letras; 2022.

BUGHIN, J. et al. **Notes from the AI frontier: Modeling the impact of AI on the world economy**. McKinsey Global Institute, 04 set. 2018. Disponível em: <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy>. Acesso em: 04 mar. 2023.

CARROLL, A. D. **Equity-Centered Community Design Field Guide**. 2018. Disponível em: <https://static1.squarespace.com/static/5e3b20447d777f2b32c1bc1c/t/5e667103feb2830f1b1b68d4/1583771908636/ECCD+FIELD+GUIDE+-+DOWNLOAD.pdf>. Acesso em: 02 fev. 2024.

CERDEÑA, J. P.; PLAISIME, M. V.; TSAI, J. **From race-based to race-conscious medicine: how anti-racist uprisings call us to act**. *The Lancet*, v. 396, n. 10257, p. 1125–1128, 10 out. 2020a.

CHARMAZ, K. **A construção da teoria fundamentada: guia prático para análise qualitativa**. Porto Alegre: Artmed Editora S.A., 2009. 272 p. Tradução: Joice Elias Costa.

CHEN, L; CHEN, P; LIN, Z. **Artificial Intelligence in Education: a Review**. *IEEE Access*, v. 8, n. 2169-3536, p. 75264–75278, 17 abr. 2020.

CROSS, N. **Design Thinking: Understanding How Designers Think and Work**. 2. ed. Bloomsbury Visual Arts, 2023.

DANKWA-MULLAN, I. et al. **A Proposed Framework on Integrating Health Equity and Racial Justice into the Artificial Intelligence Development Lifecycle**. *Journal of Health Care for the Poor and Underserved*, v. 32, n. 2, Supplement, p. 300-317, maio 2021. Johns Hopkins University Press. Disponível em: <https://doi.org/10.1353/hpu.2021.0065>. Acesso em: 8 ago. 2023.

DA HORA, N. **Nina da Hora: "As big techs já ultrapassaram os limites da influência delas"**. Revista Galileu, 02 abr. 2023. Disponível em: <https://revistagalileu.globo.com/tecnologia/noticia/2023/04/nina-da-hora-as-big-techs-ja-ultrapassaram-os-limites-da-influencia-delas.ghtml>. Acesso em: 02 fev. 2024.

DISTRITO. **Inteligência Artificial Report**. São Paulo: Distrito, Janeiro 2021.

GERMANI, A. C. et al. **Interdisciplinaridade, Interprofissionalidade E Diversidade Racial Na Formação Antirracista Do Profissional De saúde: Vozes E Aprendizados**. Saúde Em Debate, v. 46, n. spe6, p. 175–184, 1 jan. 2022.

GOOGLE FOR STARTUPS; BOX1824; ABSTARTUPS. **O Impacto E O Futuro Da Inteligência Artificial No Brasil**. Disponível em: <https://startup.google.com/intl/pt-BR/inteligencia-artificial/>. Acesso em: 5 mar. 2023.

IEEE GLOBAL INITIATIVE. **Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, Version 2**. 2019. Disponível em: [https://standards.ieee.org/wp-content/uploads/import/documents/other/ead\\_v2.pdf](https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf). Acesso em: 01 dez. 2023.

KRIPPENDORFF, Klaus. **Design centrado no usuário: uma necessidade cultural**. Estudos em design, Rio de Janeiro, v. 8, n. 3, 2000.

LEAL, M. C. et al. **A cor da dor: iniquidades raciais na atenção pré-natal e ao parto no Brasil**. Cadernos de Saúde Pública, v. 33, p. e00078816, 24 jul. 2017.

LIMA, D. C. et al. **Negras in tech: apropriação de tecnologias por mulheres negras como estratégias de resistência**. **Cadernos Pagu**, [s.l.], n. 59, 2020. Disponível em: <http://dx.doi.org/10.1590/18094449202000590006>. Acesso em: 12 dez. 2023.

MOZAFARI, M.; FARAHBAKHS, R.; CRESPI, N. **Hate speech detection and racial bias mitigation in social media based on BERT model**. PLoS ONE, v. 15, n. 8, e0237861, 27 ago. 2020. Disponível em: <https://doi.org/10.1371/journal.pone.0237861>. Acesso em: 10 ago. 2023.

WINETZKI, A.. **Treine bem seus algoritmos para evitar decisões tendenciosas**. 2020. Disponível em: <https://febrabantech.febraban.org.br/temas/inteligencia-artificial/treine-bem-seus-algoritmos-para-evitar-decisoes-tendenciosas>. Acesso em: 10 jan. 2024.

OBERMEYER, Z. et al. **Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations**. Science, v. 366, n. 6464, 2019.

O'BRIEN, J.; CLARE, C. A. **Race-based Versus Race-conscious Medicine in Obstetrics and Gynecology**. Clinical Obstetrics and Gynecology, v. 66, n. 1, p. 95, mar. 2023.

OLIVEIRA, A. de A. de. **ESPECIAL INOVADORES NEGROS: 10 profissionais que estão fazendo história nas grandes empresas de tecnologia.** Forbes Brasil, 24 jul. 2020. Disponível em: <https://forbes.com.br/forbes-tech/especial-inovadores-negros/2020/07/especial-inovadores-negros-10-profissionais-que-estao-fazendo-historia-nas-grandes-empresas-de-tecnologia/>. Acesso em: 15 jan. 2024.

PANIZZA, J. F. et al. **Metodologia E Processo Criativo Em Projetos De Comunicação Visual.** [s.l.: s.n.]. Disponível em: <<https://www.teses.usp.br/teses/disponiveis/27/27148/tde-04082006-120606/publico/me-todo-criatividade.pdf>>. Acesso em: 14 jun. 2023.

PEOPLE + AI RESEARCH TEAM. **People + AI Guidebook.** [S.l.]: Google PAIR, 8 maio 2019. Disponível em: <https://pair.withgoogle.com/guidebook>. Acesso em: 06 ago. 2023

PEREZ, I. U. et al. **Design de transições para a sustentabilidade: expansão e decolonização no design contemporâneo.** In: 14º Congresso Brasileiro de Pesquisa e Desenvolvimento em Design, 2022, vol. 10, n. 5. Disponível em: <https://www.proceedings.blucher.com.br/article-details/design-de-transies-para-a-sustentabilidade-expansao-e-decolonizacao-no-design-contemporneo-38340>. Acesso em: 15 ago. 2023.

RODA VIVA. **Entrevista com Cida Bento.** Disponível em: <https://www.youtube.com/live/pA7bZnpRWnY?si=OIPMlID5a45Ve7uk>. Acesso em: 4 mar. 2024.

ROSSETTI, R.; ANGELUCI, A. **Ética Algorítmica: questões e desafios éticos do avanço tecnológico da sociedade da informação.** Galáxia (São Paulo, online), N° 46, pp. 1-18, 2021. Disponível em: <https://www.scielo.br/j/gal/a/R9F45HyqFZMpQp9BGTfZnyr/?lang=pt> . Acesso em: 20 fev. 2024.

SCHÖN, D. A. **Designing: Rules, types and worlds.** *Design Studies*, v. 9, n. 3, p. 181-190, 1988. Disponível em: [https://doi.org/10.1016/0142-694X\(88\)90047-6](https://doi.org/10.1016/0142-694X(88)90047-6). Acesso em: 20 ago. 2023.

SILVA, T. **Visão Computacional e Racismo Algorítmico: Branquitude e Opacidade no Aprendizado de Máquina.** Revista ABPN, v. 12, p. 428-448, 2020. Disponível em: <https://abpnrevista.org.br/site/article/view/744/774>. Acesso em: 10 nov 2023.

\_\_\_\_\_, T. **Racismo algorítmico: inteligência artificial e discriminação nas redes digitais.** 1. ed. São Paulo: Edições Sesc SP, 2022. 223 p. Disponível em: <https://racismo-algoritmico.pubpub.org/> . Acesso em: 5 dez. 2023.

SOUZA, M. C. et al. **Racismo estrutural como barreira de acesso à saúde: O que a pandemia da COVID-19 nos revela?**. J. of Multiprofessional Health Research, v. 2, e02.102-e02.106, 02 fev. 2021.

SOSA HIDALGO, M. **Design of an Ethical Toolkit for the Development of AI Applications**. 2019. Dissertação (Mestrado em Strategic Product Design) - Delft University of Technology, Delft, 2019. Disponível em: <http://resolver.tudelft.nl/uuid:b5679758-343d-4437-b202-86b3c5cef6aa>. Acesso em: 8 ago. 2023

SHANKAR, V; COOK, A. **Human-Centered Design for AI**. Kaggle, [s.d.]. Disponível em: <https://kaggle.com/code/var0101/human-centered-design-for-ai>. Acesso em: 5 mar. 2023.

STURM, T; FECHO, M; BUXMANN, P. **To Use or Not to Use Artificial Intelligence? A Framework for the Ideation and Evaluation of Problems to Be Solved with Artificial Intelligence**. Technical University of Darmstadt. Disponível em: <https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/73b2d588-203d-45aa-9aed-b34e95e57d9a/content>. Acesso em: 5 ago. 2023.

STANFORD UNIVERSITY. **AI Index Report 2021**. Disponível em: <https://aiindex.stanford.edu/ai-index-report-2021/>. Acesso em: 5 mar. 2023.

SURESH, H.; GUTTAG, J. **A Framework for Understanding Unintended Consequences of Machine Learning**. 2021. Disponível em: <https://arxiv.org/pdf/1901.10002.pdf>. Acesso em: 8 ago. 2023.

TONKINWISE, C. **Design for Transitions – from and to what?** Design Philosophy Papers, 2 jan. 2015.

WILES, J. **Beyond ChatGPT: The Future of Generative AI for Enterprises**. Gartner, 26 jan. 2023. Disponível em: <https://www.gartner.com/en/articles/beyond-chatgpt-the-future-of-generative-ai-for-enterprises>. Acesso em: 20 maio 2023.