



**Elas na inteligência  
artificial – Questões de gênero**

*Renata Wassermann*

## resumo

Há muitos anos se discute a baixa representatividade de alguns grupos nas áreas tecnológicas, principalmente mulheres e pessoas negras. Neste artigo, examinamos o caso da inteligência artificial, em que o panorama não é diferente. Discutimos os efeitos da falta de diversidade nos grupos que desenvolvem a IA e iniciativas ao redor do mundo para tentar mitigá-los. Mostramos um caso específico – sistemas de tradução automática, uma subárea da IA – em que os estereótipos de gênero aparecem.

---

**Palavras-chave:** IA e gênero; diversidade; viés de gênero; viés de raça; tradução e viés.

## abstract

*For many years, the low representation of some groups in technological areas, especially women and black people, has been discussed. In this article, we examine the case of artificial intelligence, where the panorama is not different. We discuss the effects of a lack of diversity in groups developing AI and initiatives worldwide to mitigate them. We show a specific case – machine translation systems, a sub-area of AI – in which gender stereotypes appear.*

---

**Keywords:** AI and gender; diversity; gender bias; race bias; translation and bias.

**N**os anos 80 e início dos anos 90, o termo “inteligência artificial” era quase sinônimo de sistemas especialistas, baseados em regras que codificavam de forma simbólica, através de alguma linguagem formal, o conhecimento de especialistas, como médicos ou químicos. No início do século 21, com o avanço das abordagens não simbólicas, baseadas em aprendizado de máquina, passamos a ver referências a “uma inteligência artificial”, ou à “IA da empresa X”, tratando cada sistema como um agente dotado de inteligência. Em 2022, com a popularização dos grandes modelos de linguagem como o GPT e o Bard, a IA passou a ser vista pelo público leigo como sendo apenas esta subárea de pesquisa. Os modelos de linguagem

são baseados em redes neurais profundas e aprendizado autossupervisionado, técnicas que já existem há décadas. No entanto, a escala e o escopo dos modelos aumentaram enormemente. Por exemplo, o GPT-3 tem 175 bilhões de parâmetros e pode ser adaptado por meio de instruções em linguagem natural para fazer uma ampla gama de tarefas, apesar de não ter sido treinado explicitamente para realizar muitas delas (Bommasani et al., 2022).

A primeira falta de diversidade que surge com a popularização dos grandes modelos é a de linhas de pesquisa. Rapidamente, a atenção e o financiamento se voltaram à construção de modelos de linguagem cada

---

**RENATA WASSERMANN** é professora associada do Departamento de Computação do Instituto de Matemática e Estatística (IME) da USP e pesquisadora do C4AI e do Lawgorithm, também da USP.

vez maiores, que para funcionar necessitam cada vez mais de recursos computacionais e fontes de dados para treinamento. Com isso, cada vez menos instituições possuem os recursos necessários, fazendo com que um número bastante pequeno de *big techs* domine o mercado. Países e continentes inteiros ficam excluídos do desenvolvimento desse tipo de inteligência artificial.

Outro grande problema é a amplificação de preconceitos e vieses presentes nas bases de dados utilizadas para treinamento desses modelos. Bommasani et al. (2022) descrevem diferentes tipos de vieses encontrados nos grandes modelos de linguagem. Apesar de serem problemas que aparecem em outros sistemas de inteligência artificial e ciências em geral, como os modelos são mal compreendidos e utilizados em diferentes aplicações, os danos são potencialmente maiores, como já ficou claro desde o lançamento do livro de Cathy O’Neil (2017).

Um caso bastante estudado atualmente é o de sistemas com viés de raça. Danos podem ir desde inconvenientes, como o dispenser de sabão automático que viralizou na internet por apenas reconhecer mãos de pessoas brancas, até carros que não reconhecem pedestres negros<sup>1</sup> e sistemas que caracterizam negros como criminosos ou mais propensos a reincidir em crimes. Buolamwini (2017) verificou sistematicamente ferramentas de reconhecimento facial e mostrou como as bases de dados usadas pelas ferramentas eram pouco diversas e funcionavam melhor para homens brancos. Muitas vezes, as pessoas nem sabem o que está sendo levado em

conta na sua avaliação, como no caso recente das universidades americanas<sup>2</sup>. É importante lembrar que nos Estados Unidos pessoas de origem latina não são consideradas brancas e também sofrem discriminação.

De um modo geral, os vieses nos dados refletem preconceitos presentes na sociedade em que os dados foram criados. Como disse Benjamin (2019, p. 60)<sup>3</sup>, não é necessário ter má intenção, basta não se preocupar em como o passado molda o presente. Se tradicionalmente cozinhar era uma ocupação feminina, existem mais imagens de mulheres do que homens e, assim, sistemas passam a reconhecer pessoas na cozinha como mulheres<sup>4</sup>.

Parece bastante claro que algo precisa ser feito para mitigar os danos causados por vieses (Boinodiris, 2024), mas mesmo assim propostas de ações afirmativas ainda encontram bastante resistência, com alegações de que são discriminatórias e desnecessárias<sup>5</sup> (West; Whittaker; Crawford, 2019). Podemos pensar em três aspectos de diversidade:

- diversidade dentro dos grupos que fazem a IA – se as pessoas que trabalham no desenvolvimento de sistemas de IA são brancas, do gênero masculino, podem passar despercebidas falhas como uso de estereótipos e linguagem inadequada;

1 Ver: <https://www.businessinsider.com/self-driving-cars-worse-at-detecting-dark-skin-study-says-2019-3>.

2 Ver: <https://themarkup.org/machine-learning/2021/03/02/major-universities-are-using-race-as-a-high-impact-predictor-of-student-success>.

3 “No malice needed, no N-word required, just lack of concern for how the past shapes the present.”

4 Ver: [https://brasil.elpais.com/brasil/2017/09/19/ciencia/1505818015\\_847097.html](https://brasil.elpais.com/brasil/2017/09/19/ciencia/1505818015_847097.html).

5 “AI researchers greeted the announcement of the Black in AI workshop at NeurIPS, a leading machine learning conference, by questioning whether the event was necessary and arguing that it would be discriminatory.”

- diversidade nas fontes de dados utilizadas – os algoritmos usados em IA em geral são conhecidos. O que diferencia os sistemas normalmente são as bases usadas para treinamento. Se as bases são enviesadas, o sistema aprende e repete o padrão;
- diversidade entre as pessoas para quem os sistemas são desenvolvidos – o sistema reconhece rostos/vozes/linguagens diversas?

Ainda que distintos, os três aspectos estão diretamente relacionados e iniciativas para criar equipes de desenvolvimento diversas podem resultar em dados menos enviesados e em sistemas adaptados para usuários diversos. Na próxima seção, comentamos algumas iniciativas existentes para diversificar quem faz a IA.

## INICIATIVAS PARA AMPLIAR A DIVERSIDADE EM IA

Diversas iniciativas surgiram para tentar incentivar a presença de grupos sub-representados na inteligência artificial. Na grande maioria, os grupos estão mais presentes na América do Norte.

- **Black in AI**<sup>6</sup>: começou em 2017 como um grupo no Facebook e hoje tem mais de 5 mil membros. Organiza workshops na NeurIPS todo ano. Tem iniciativas de mentoria e ajuda financeira para participar de eventos da área. O grupo também oferece ajuda para lidar com vistos e outras barreiras burocráticas;
- **Queer in AI**<sup>7</sup> – organiza workshops regularmente desde 2018. Oferece bolsas de estudo para estudantes LGBTQIA+ em todos os níveis. Também oferece ajuda para inscrições em programas de pós-graduação;
- **Latinx in AI**<sup>8</sup> – surgiu em 2018 e organiza diversos workshops anualmente e grandes conferências na área de IA. Oferece mentorias e financiamento para participação nas conferências principalmente para estudantes provenientes da América Latina;
- **Women in Machine Learning**<sup>9</sup> – organiza workshops anuais desde 2006 junto à NeurIPS. Oferece mentoria para mulheres na área de aprendizado de máquina e mantém uma lista de mulheres que trabalham na área;
- **Women in AI**<sup>10</sup> – começou em 2016 também como um grupo no Facebook e hoje conta com mais de 8 mil membros. Dentre as organizações listadas, é a mais voltada para a indústria. Oferece programas de aceleração de startups lideradas por mulheres, *hackathons* e programas de atualização em IA.

Alguns grupos mais recentes incluem {Dis}Ability in AI<sup>11</sup> e Indigenous in AI<sup>12</sup>.

Nos últimos três anos, grandes conferências internacionais, como IJCAI, AAAI, AAMAS e NeurIPS, passaram a

---

6 Ver: <https://blackinai.github.io/>.

---

7 Ver: <https://www.queerinaai.com/>.

8 Ver: <https://www.latinxinaai.org/>.

9 Ver: <https://wimlworkshop.org/>.

10 Ver: <https://www.womeninaai.co/>.

11 Ver [https://elesa.github.io/ability\\_in\\_AI/](https://elesa.github.io/ability_in_AI/).

12 Ver: <https://www.indigenous-ai.net/>.

ter chamadas públicas para organização de eventos relacionados à diversidade. Além de workshops e eventos sociais devotados a grupos pouco representados, as conferências contam também com políticas de inclusão e diversidade na sua organização, levando em conta a diversidade na composição dos comitês avaliadores e nos convites a palestrantes.

Em um desses eventos, na AAAI de 2021, foi apresentado um sistema em desenvolvimento para medir o índice de diversidade de conferências (Hupont et al., 2022), levando em conta palestrantes convidados, comitês de programa e organização e autoria de artigos. A diversidade é avaliada em três dimensões: gênero (por enquanto, binário e baseado unicamente em reconhecimento de nomes), localização (país) e área de atuação (academia, indústria ou centro de pesquisa).

No Brasil ainda há poucas iniciativas específicas para o aumento da diversidade em IA. A Sociedade Brasileira de Computação (SBC) apoia desde 2007 o WIT – Women in Information Technology<sup>13</sup>, workshop que acontece junto ao seu congresso principal para discutir os assuntos relacionados a questões de gênero e à tecnologia da informação (TI) no Brasil. Um dos focos é debater políticas de educação e treinamento de mulheres em TI, assim como seu recrutamento em empresas e grupos de pesquisa. O Programa Meninas Digitais<sup>14</sup> da SBC possui dezenas de iniciativas afiliadas em todas as regiões do país, sendo voltado a motivar alunas do ensino fundamental e médio a seguir carreira

na área de computação. O grupo Mulheres em IA<sup>15</sup> foca encontros para troca de experiências entre mulheres trabalhando em IA no Brasil. A Lawgorithm<sup>16</sup>, organização voltada à pesquisa em inteligência artificial e direito, possui entre seus núcleos um dedicado a IA e gênero e outro dedicado a IA e raça. No Centro de Inteligência Artificial (C4AI)<sup>17</sup>, temos desde 2021 um comitê de diversidade e inclusão que, além de ações de conscientização, financia bolsas de estudo para estudantes de graduação da Universidade de São Paulo. Por entendermos que a diversidade é necessária em muitos aspectos diferentes, procuramos selecionar estudantes que venham de áreas de conhecimento diversas, com variabilidade de gênero e raça. Dentro das ciências exatas, por exemplo, privilegiamos a escolha de bolsistas mulheres, dada a sua pouca representatividade.

## GÊNERO EM IA

A preocupação com a falta de inclusão de mulheres em IA não é nem um pouco nova nem restrita às técnicas mais modernas. Já nos anos 80, pesquisadoras se depararam com estereótipos de gênero em sistemas especialistas e também com a falta de mulheres estudando IA. Bruce e Adam (1989) já se propunham a estudar até que ponto os sistemas reproduzem questões de gênero e quais as oportunidades para que mulheres participem do desenvolvimento dessas tecnologias.

---

13 Ver: <https://csbc.sbc.org.br/2023/wit/>.

14 Ver: <https://meninas.sbc.org.br/>.

15 Ver: <https://linktr.ee/mulheres.em.ia>.

16 Ver: <https://www.lawgorithm.org.br/>.

17 Ver: <https://c4ai.inova.usp.br/>.

Hall e Lovegrove (1988) discutem a preocupação com a falta de representatividade das mulheres nas equipes que desenvolvem a IA desde a época da formação. Como grande parte dos profissionais vem da computação e já era perceptível na época a diminuição de estudantes mulheres na área, como evitar que a IA fosse feita apenas por homens?

Recentemente, o Observatório de Inteligência Artificial da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) publicou um artigo (Caira; Russo; Aranda, 2023) defendendo que os países-membros atuem de forma a facilitar o acesso das mulheres à educação necessária para entender como os sistemas de IA estão sendo desenvolvidos, como uma maneira de evitar que a desigualdade no mercado de trabalho aumente. O artigo também chama a atenção para o fato de que a grande maioria dos CEOs das grandes empresas de tecnologia são homens, assim como a maior parte dos pesquisadores em IA.

Em um artigo recente sobre as pessoas mais importantes por trás da IA moderna, 12 homens foram mencionados<sup>18</sup>. No Brasil, de acordo com o LinkedIn, 25% das pessoas trabalhando em IA eram mulheres em 2022<sup>19</sup>.

Com o avanço da IA generativa, a internet foi inundada de *deep fakes*, conteúdo falso em formato de fotos ou vídeos, muitas vezes contendo pornografia ou imagens altamente sexualizadas de mulheres reais

ou criadas por IA. Lamensch (2023) conta que um relatório publicado em 2019 por DeepTrace indicou que, de 15 mil vídeos classificados como *deep fake*, 96% continham pornografia com rostos de mulheres colocados artificialmente nas imagens. Ainda no mesmo artigo, Lamensch descreve um experimento feito com um aplicativo de criação de avatares a partir de fotos pela repórter Melissa Heikkilä. Quando ela tentou criar um avatar de si mesma, o aplicativo propôs avatares seminus e sexualizados. Por outro lado, os avatares dos colegas homens de Heikkilä estavam totalmente vestidos e pareciam astronautas, exploradores e inventores. Segundo Lamensch, a maior parte dos algoritmos de geração de imagem a partir de texto é treinada com a base de dados LAION-5B<sup>20</sup>, com mais de 5 bilhões de pares de imagem/texto extraídos da internet. Como a internet está repleta de conteúdo misógino e racista, com falta de representatividade de diversos grupos, isso é propagado para as ferramentas que utilizam a base.

Não é apenas a IA generativa que trata mulheres de forma diferente. Um estudo mostrou que os algoritmos de classificação de imagens usados nas redes sociais para evitar imagens ofensivas classificam como inadequadas imagens de mulheres em que aparecem pequenas porções de pele, enquanto imagens de homens de torso nu não são censuradas (Mauro; Schellmann, 2023). Barrigas de grávidas são classificadas como tendo alta probabilidade de conter conteúdo sexual. O mesmo acontece com campanhas para

---

18 Ver: <https://www.nytimes.com/2023/12/03/technology/ai-key-figures.html>.

19 Ver: [https://www.linkedin.com/posts/linkedin-noticias\\_iwd2023-activity-7039627442877083648-Pl-F?utm\\_source=share&utm\\_medium=member\\_desktop](https://www.linkedin.com/posts/linkedin-noticias_iwd2023-activity-7039627442877083648-Pl-F?utm_source=share&utm_medium=member_desktop).

---

20 Ver: <https://laion.ai/blog/laion-5b/>.

prevenção de câncer de mama. O impacto dessas classificações nas redes sociais pode ser muito maior do que imaginamos. Muitas vezes, a imagem tida como possivelmente inadequada não é bloqueada, mas sofre um processo conhecido em inglês como *shadowbanning*, em que o algoritmo restringe o alcance de uma postagem. Enquanto o bloqueio de conteúdo considerado apropriado é notificado ao usuário, no caso de *shadowbanning* isso normalmente não acontece. O artigo cita dois casos de mulheres cujo trabalho depende da divulgação em redes sociais e que sofrem constantemente com o *shadowbanning*. Assim, suas postagens não têm o mesmo alcance que as postagens de concorrentes do sexo masculino, com consequente aumento na disparidade econômica.

Outro fenômeno que atraiu atenção no final da década passada foi o surgimento de assistentes digitais baseadas em IA, como Siri e Alexa, entre outras. Não é coincidência o fato de praticamente todas as assistentes terem nomes e vozes femininas. De acordo com West, Kraut e Ei (2019), para justificar a decisão de tornar os assistentes de voz femininos, empresas como a Amazon e a Apple citaram trabalhos acadêmicos que demonstram que as pessoas preferem uma voz feminina a uma voz masculina. Por outro lado, outros estudos sugerem que a maioria das pessoas percebe as vozes femininas como cooperativas, além de úteis, enquanto as vozes masculinas são consideradas autoritárias. Ou seja: as assistentes digitais devem ter vozes femininas para os consumidores se sentirem à vontade para dar ordens a elas! Pesquisadores observaram que o uso de assistentes virtuais ensina (inclusive a crianças), que pessoas do gênero feminino devem ser subservientes e responder a comandos. Ou

seja, a tecnologia pode servir para amplificar as desigualdades de gênero (West; Kraut; Ei, 2019). Aqui no Brasil temos a BIA<sup>21</sup>, a IA do Bradesco, que foi apresentada desde o início (em 2018) como feminina. Após diversas interações com usuários ofensivas e agressivas, o próprio Bradesco teve de lançar um programa contra o assédio que sua assistente sofria<sup>22</sup>.

O último ponto que eu gostaria de mencionar nesta seção é a disparidade de gênero no uso de IA, notada em um artigo recente (Costa, 2023). Uma das razões levantadas é a de que, como os estudos em áreas como a computação e a engenharia são dominados por homens, as mulheres se sentem menos à vontade para adotar novas tecnologias. Outro ponto que tem aparecido na mídia é que as mulheres têm mais medo de serem substituídas por IA em seus empregos do que os homens e, portanto, veem a tecnologia com desconfiança. Esse medo não é infundado, como mostram estudos (Unesco/OECD/IDB, 2022).

## O CASO DO PROCESSAMENTO DE LINGUAGEM NATURAL

Processamento de linguagem natural (PLN) é a área da inteligência artificial que lida com interpretação e geração de linguagem “humana”. O termo “natural” aparece em contraste com a ideia de linguagem formal, como a lógica matemática ou uma linguagem de programação,

---

21 Ver: <https://banco.bradesco/bia/>.

22 Ver: <https://banco.bradesco/aliadosbia/>.

desenhada de acordo com regras rígidas de sintaxe e semântica e com propósitos bem definidos. As linguagens naturais, como o português e o inglês, por outro lado, surgiram de forma espontânea e foram evoluindo ao longo do tempo, de acordo com mudanças na sociedade e necessidades de comunicação entre seus falantes. PLN sempre esteve em uma posição de destaque dentro da IA, já que para a popularização do uso de ferramentas de IA elas devem ser acessadas sem a necessidade de uma linguagem formal intermediária. Um exemplo de uso são as já citadas assistentes virtuais, que recebem perguntas ou comandos em linguagem natural e também formulam a resposta na mesma linguagem.

Assim como outras áreas da IA, as técnicas de PLN atuais são quase sempre baseadas em aprendizado de máquina. Como já discutido, o uso de grandes bases de dados para treinamento dessas ferramentas tem o potencial de propagar os vieses de gênero dos textos utilizados. Uma das técnicas para predição de textos, conhecida como *word embedding*, utiliza uma representação espacial de termos e textos para fazer projeções de vetores. Assim, pode-se projetar a relação homem/rei para o espaço onde está a palavra mulher e encontrar a relação mulher/rainha. Também a partir da relação França/Paris, podemos encontrar Japão/Tóquio. No entanto, nem sempre as relações encontradas são as desejadas, como alertado por Bolukbasi et al. (2016). No artigo, os autores discutem como alguns vieses estão presentes na própria língua inglesa, mas também como sistemas treinados em bases supostamente idôneas também aprendem estereótipos de gênero.

Esse foi talvez o primeiro artigo a estudar o fenômeno em profundidade e a propor

técnicas de “desenviesamento” (*debiasing*) de algoritmos. A técnica proposta, no entanto, baseia-se no fato de que, na língua inglesa, a maior parte das palavras é neutra em relação a gênero, o que não ocorre em português. Se olharmos para palavras referentes a profissões, por exemplo, temos *teacher*, que pode significar professor ou professora, *doctor*, que pode significar médico ou médica, e assim por diante. Claramente, essa diferença entre as línguas traz uma dificuldade extra para traduções, sejam elas automáticas ou realizadas por seres humanos. Como podemos traduzir “*the teacher has arrived*”? A tradução para o português deve optar por um gênero? Ou tentar utilizar termos que mantenham a indefinição da sentença em inglês? O Google Tradutor, quando consultado sobre a tradução de “*the teacher has arrived*” para português, devolve como resposta (em janeiro de 2024): “A professora chegou”. No entanto, ao mudarmos a profissão da consulta para *doctor*, a resposta passa a ser “o médico chegou”. Aqui podemos ver o resultado do aprendizado estatístico produzindo um estereótipo de gênero – mulheres como professoras, homens como médicos. A situação fica ainda pior quando traduzimos entre duas linguagens com marcação de gênero, como português e espanhol. Traduzindo “a médica e o enfermeiro chegaram” para o espanhol, retorna como resultado “*llegaram el médico y la enfermera*”, porque as traduções entre as duas línguas são feitas com uma passagem intermediária pelo inglês (português → inglês → espanhol) e, assim, perdemos os gêneros marcados na frase original. Esse fenômeno foi estudado por Prates et al. (2020) em relação a línguas como o húngaro, que têm ainda menos marcações de gênero do que o inglês.

No caso das traduções do inglês para o português, já comentamos que elas apresentam um desafio mesmo para tradutores humanos. Existe todo um movimento social preconizando a adoção de uma linguagem inclusiva, muitas vezes chamada de linguagem neutra, que possa abarcar situações em que o gênero seja desconhecido ou não binário. Em inglês, já é bastante consensual o uso dos pronomes *they/them* para essas situações. Já em português, a criação de tal linguagem inclusiva vem despertando fortes reações contrárias, e até proibições de uso<sup>23</sup>. Isso se deve ao fato de a bandeira ter sido levantada inicialmente por populações marginalizadas, como as pessoas transgêneros ou que se identificam com alguma das muitas possibilidades de gênero existentes para além das opções binárias (masculino/feminino) e que são excluídas pelo uso da norma binária. Mesmo sem entrar na questão sociológica, vemos que a linguagem inclusiva não é apenas uma questão de discutir gêneros não conformantes, mas também de transmitir informações quando o gênero das pessoas é desconhecido. Nesse caso, os modelos de linguagem modernos baseados em aprendizado de máquina têm pouco a dizer, pois (ainda?) não existe uma linguagem inclusiva que seja consenso e muito menos grandes conjuntos de textos bilíngues em linguagem inclusiva. Em Lopes et al. (2023), fizemos um estudo preliminar utilizando a linguagem proposta por Cassiano (2019) e testamos um protótipo de tradutor de português “correto” para um português “inclusivo”.

---

23 Ver: <https://g1.globo.com/educacao/noticia/2023/02/13/linguagem-neutra-barrada-pelo-stf-lei-que-proibe-o-uso-de-linguagem-neutra-existe-em-3-estados-e-2-capitais.ghtml>.

Lopes et al. (2023) realizaram também experimentos com o ChatGPT, que apresentou um bom desempenho para o Prompt “Escreva as frases seguintes usando linguagem neutra: <[frases]>”, mas utilizou alguns artifícios considerados inadequados. Exemplos positivos desse uso são as seguintes reescritas: “Aqueles que terminaram podem sair → Quem terminou pode sair” e “Nós somos alunos de teatro → Nós somos estudantes de teatro”. Exemplos negativos são: “As cantoras denunciaram os assédios → As artistas denunciaram os assédios”, no qual o gênero feminino não foi neutralizado, e “Você é biólogo? → Você é biólogx?”, no qual o uso do caractere “x” como neutralizador foi aplicado, o que não é recomendado atualmente.

## DISCUSSÃO

A maior parte das questões ligadas a gênero é inerente à sociedade e não exclusividade da área de inteligência artificial. No entanto, como vimos, as novas tecnologias têm o potencial de amplificar e fossilizar estas questões. Amplificar, pois o alcance da IA aumentou e as pessoas têm a tendência a imaginar a tecnologia como sendo imune a preconceitos tipicamente humanos e, portanto, não questionar. Além disso, conforme um modelo de linguagem enviesado serve de base para a implementação de outras ferramentas, os vieses vão sendo transmitidos em cadeia. Fossilizar, porque, como os modelos de linguagem são treinados com grandes volumes de dados já existentes, mesmo que a sociedade evolua, existe o risco de os preconceitos do passado serem usados para moldar o futuro.

Para mitigar os danos, podemos pensar em propostas de curto, médio e longo prazos. No curto prazo, seria muito importante exigir das empresas transparência em relação aos dados usados no treinamento das ferramentas e formular os requisitos de representatividade que tais dados devem cumprir. A médio prazo, políticas afirmati-

vas de contratação podem gerar equipes de desenvolvimento mais diversas e, portanto, mais sensíveis ao sexismo. A longo prazo, políticas públicas de educação igualitária e de atração de meninas para as áreas de tecnologia têm o potencial de fazer com que as próximas gerações não precisem se preocupar com questões de gênero.

## REFERÊNCIAS

- BENJAMIN, R. *Race after technology: abolitionist tools for the New Jim Code*. Cambridge, Polity, 2019.
- BOINODIRIS, P. *The importance of diversity in AI isn't opinion, it's math*. 2024. Disponível em: <https://www.ibm.com/blog/why-we-need-diverse-multidisciplinary-coes-for-model-risk/>.
- BOLUKBASI, T. et al. "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings", in *Proceedings of the 30th International Conference on Neural Information Processing Systems (NIPS'16)*.
- BOMMASANI, R. et al. *On the opportunities and risks of foundation models*. 2022. Disponível em: <https://arxiv.org/abs/2108.07258>.
- BRUCE, M.; ADAM, A. "Expert systems and women's lives: a technology assessment". *Futures*, v. 21, issue 5, 1989, pp. 480-97.
- BUOLAMWINI, J. *Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers*. Dissertação de mestrado. Massachusetts, MIT, 2017. Disponível em: <https://www.media.mit.edu/publications/full-gender-shades-thesis-17/>.
- CAIRA, C.; RUSSO, L.; ARANDA, L. *Artificially inequitable? AI and closing the gender gap*. 2023. Disponível em: <https://oecd.ai/en/wonk/closing-the-gender-gap>.

- CASSIANO, O. *Guia para "linguagem neutra" (PT-BR)*. 2019. Disponível em: <https://medium.com/guia-para-linguagem-neutra-pt-br/guia-para-linguagem-neutra-pt-br-f6d88311f92b>.
- COSTA, M. "Why are fewer women using AI than men?". *BBC*, 2023. Disponível em: <https://www.bbc.com/news/business-67217915>.
- HALL, W.; LOVEGROVE, G. "Women and AI". *AI & Society*. v. 2, 1988, pp. 270-1.
- HUPONT, I. et al. "Monitoring diversity of AI conferences: lessons learnt and future challenges in the DivinAI project". 2022. Disponível em: <https://arxiv.org/abs/2203.01657>.
- LAMENSCH, M. *Generative AI tools are perpetuating harmful gender stereotypes*. 2023. Disponível em: <https://www.cigionline.org/articles/generative-ai-tools-are-perpetuating-harmful-gender-stereotypes/>.
- LOPES, W. R. et al. "Eu, tu, ele, ela, elu, nós, vós, eles, elas, elus por um modelo de linguagem neutra". *Anais do Encontro Nacional de Inteligência Artificial e Computacional (Eniac)*, 20, 2023. Disponível em: <https://sol.sbc.org.br/index.php/eniac/article/view/2575>.
- MAURO, G.; SCHELLMANN, H. "'There is no standard': investigation finds AI algorithms objectify women's bodies". 2023. Disponível em: <https://www.theguardian.com/technology/2023/feb/08/biased-ai-algorithms-racy-women-bodies>.
- O'NEIL, C. *Weapons of math destruction*. Londres, Penguin Books, 2017.
- PRATES, M. O. R.; AVELAR, P. H.; LAMB, L. C. "Assessing gender bias in machine translation: a case study with Google Translate". *Neural Computing & Applications*, v. 32, 2020, pp. 6.363-81.
- UNESCO – Organização das Nações Unidas para a Educação, a Ciência e a Cultura; OECD – Organização para a Cooperação e Desenvolvimento Econômico; IDB – Banco Interamericano de Desenvolvimento. *The effects of AI on the working lives of women*. Paris, Unesco, 2022.
- WEST, M.; KRAUT, R.; CHEW, H. E. "The rise of gendered AI and its troubling repercussions". *I'd blush if I could: closing gender divides in digital skills through education*. Paris, Unesco, 2019.
- WEST, S. M.; WHITTAKER, M.; CRAWFORD, K. "Discriminating systems: gender, race and power in AI". *AI Now Institute*, 2019. Disponível em: <https://ainowinstitute.org/publication/discriminating-systems-gender-race-and-power-in-ai->