FUNDAÇÃO GETULIO VARGAS
ESCOLA DE DIREITO FGV DO RIO DE JANEIRO
GRADUAÇÃO EM DIREITO

BRUNA DINIZ FRANQUEIRA

COMO A INTELIGÊNCIA ARTIFICIAL REFORÇA A DISCRIMINAÇÃO DE GÊNERO NO AMBIENTE DE TRABALHO

Rio de Janeiro, Dezembro/2019

FUNDAÇÃO GETULIO VARGAS
ESCOLA DE DIREITO FGV DO RIO DE JANEIRO
GRADUAÇÃO EM DIREITO

BRUNA DINIZ FRANQUEIRA

COMO A INTELIGÊNCIA ARTIFICIAL REFORÇA A DISCRIMINAÇÃO DE GÊNERO NO AMBIENTE DE TRABALHO

Trabalho de Conclusão de Curso, sob orientação da professora **Ligia Fabris,** e do professor **Ivar Hartmann**, apresentado à FGV DIREITO RIO como requisito parcial para obtenção do grau de bacharel em Direito.

Rio de Janeiro, dezembro de 2019

FUNDAÇÃO GETULIO VARGAS

ESCOLA DE DIREITO FGV DO RIO DE JANEIRO
GRADUAÇÃO EM DIREITO

COMO A INTELIGÊNCIA ARTIFICIAL REFORÇA A DISCRIMINAÇÃO DE GÊNERO NO AMBIENTE DE TRABALHO

Elaborado por BRUNA DINIZ FRANQUEIRA

Trabalho de Conclusão de Curso apresentado à FGV DIREITO RIO como requisito parcial para obtenção do grau de bacharel em Direito.

Comissão Examinadora:
Nome da orientadora: Ligia Fabris
Nome do co-orientador: Ivar Hartmann
Nome do Examinador 1: Caitlin Mulholland
Nome da Examinadora 2: Luca Belli

Assinaturas:

_____
Ligia Fabris

_____
Ivar A. Hartmann

_____
Caitlin Mulholland

_____
Luca Belli

Nota Final: _____

Rio de Janeiro, _____ de _____ de 2019

“A ciência da abelha, da aranha e a minha muita gente desconhece”

- João do vale

Às mulheres

AGRADECIMENTOS

Aos meus avós, a quem devo minha educação. Todos os privilégios que tive na vida foram proporcionados por vocês, que me ensinaram a reconhecê-los e aproveitá-los. Espero um dia poder retribuir.

À minha mãe, por me mostrar a magnitude da força de uma mulher. Cada dia com você era uma aula de luta, generosidade e amor. Mesmo de longe, o amor incondicional que você tem por nós se reflete na sua busca incessante em garantir a segurança com a qual você não pôde contar. Eu espero que você seja feliz sempre. Meu futuro é nosso.

À minha irmã, por não deixar de acreditar em mim e não sair do meu lado nem mesmo por um segundo. Sua confiança em mim é maior do que a minha: da forma mais paradoxal, você foi fonte de calma, paciência e esperança. Estaremos juntas para sempre.

Ao Igor, por todo companheirismo e amor. O seu apoio vem na forma de compreensão das minhas necessidades, de incentivo às minhas teorias conspiratórias - que nos rendem os debates mais calorosos e cômicos-, de risadas e de cuidado. Eu não poderia pedir além. Ainda temos muita vida para cantar.

A todas e todos que aguentaram meu gênio matinal ao longo de cinco anos. Vocês tornaram esse processo mais leve e proveitoso. Obrigada pelas risadas e colocações brilhantes. Vocês são impressionantes. Beijos infinitos e intermináveis a cada um.

À professora Ligia, que plantou a primeira semente em mim, a qual cultivarei ao longo da minha vida. Em toda minha luta, sempre terá um pouco de você.

Ao professor Ivar, sem o qual o presente trabalho não seria possível. Me faltam palavras para agradecer. Como professor, orientador e acadêmico, despertou em mim interesses que vão além da faculdade de direito e que hoje orientam meus caminhos profissionais.

Ao Leandro, professor, amigo e pai ao longo de toda a graduação. Eu não me formaria o seu apoio emocional e orientação – que sempre foi muito além da pesquisa. Espero você para entregar meu diploma.

Ao Colégio Pedro II e às amigas e amigos que fiz durante minha trajetória, com os quais desenvolvi a vivacidade e curiosidade que me estimulam. O pouco que consigo enxergar além do meu universo começou com vocês. Zum zum zum paratimbum.

RESUMO

Estamos entrando em uma nova fase de representações estereotipadas, e ainda não sabemos como a sociedade deve encará-las. Tecnologias de inteligência artificial são utilizadas para tomadas de decisão acerca de diversos aspectos de nossas vidas. Hoje, algoritmos automatizados definem até mesmo as relações de trabalho e a empregabilidade de alguém. No entanto, esses sistemas de tomada de decisão incorporam vieses de diferentes formas, fazendo com que discriminações sociais sejam reproduzidas em escala por máquinas, sem que tenhamos conhecimento acerca dos seus objetivos, funcionamentos e *inputs*. Para mulheres, essas novas formas de discriminação podem reforçar a divisão sexual do trabalho, de forma que a relevância social dos papeis atribuídos às mulheres (relacionados com a satisfação de homens) seja retomada a partir do uso de inteligência artificial em processos seletivos. Mesmo diante da incerteza dos efeitos, a automatização da contratação de pessoal é uma tendência atual, e empresas como a HireVue oferecem até mesmo serviços de entrevistas preditivas em vídeos para facilitar e reduzir o tempo de duração de processos de contratação. No presente trabalho, apresento as formas pelas quais o emprego dessas tecnologias pode reforçar a discriminação contra mulheres no mercado de trabalho, utilizando como principal caso de estudo a tecnologia da HireVue. Por ser um debate ainda tímido no Brasil, o objetivo principal é apenas fornecer uma compreensão geral de como ocorre a reprodução de vieses de gênero a partir do emprego de tecnologias de inteligência artificial.

PALAVRAS-CHAVE: mulheres; gênero; inteligência artificial; vieses; mercado de trabalho; divisão sexual do trabalho; discriminação; HireVue; entrevistas preditivas em vídeo, processo seletivo

ABSTRACT

We are now beginning a new phase of stereotyped representations, and we do not know how society should approach them yet. Artificial Intelligence is being used for decision-making processes that effect different domains of our lives. Nowadays, even labor relations and someone's degree of employability are defined by automated algorithms. These decision-making systems, though, embed bias by different means, reproducing social discrimination on a larger scale, without us knowing about their objectives, functioning and inputs. For women, these new means of discrimination may reinforce the sexual division of labor. The social role assigned for women (related to the satisfaction of men's needs) is then reinstated through the application of Artificial Intelligence in selection processes. Notwithstanding the uncertainty of its effects, the automating hiring procedures is a current trend. Corporations, such as HireVue, offer predictive video interviewing services, in order to facilitate and reduce the length of others companies recruiting procedures. In this work, I present some of the means through which these technologies may reinforce discrimination against women in the workplace, by showing how HireVue might be doing so. The main objective of this essay is only to offer an overview about how the application of Artificial Intelligence reproduce gender bias.

KEY-WORDS: women; gender; Artificial Intelligence; bias; workplace; sexual division of labour; discrimination; HireVue; predictive video interviewing; hiring processes.

ÍNDICE

Introdução

Imagine viver em uma sociedade na qual são atribuídas pontuações a todas as interações sociais. Nesse mundo hipotético, a média da avaliação de cada pessoa influencia em sua situação socioeconômica, e pode servir como limitador de acesso à moradia, bens de consumo, empregos. A popularidade em redes sociais e grau de interação que o restante das pessoas têm com suas publicações são alguns dos parâmetros utilizados para orientar o modelo matemático que gera sua média. Serviços médicos poderiam ser negados a pessoas que fogem dos padrões comportamentais largamente aceitos pela sociedade. Um tanto quanto assustador. A maior parte das pessoas acharia esse cenário surreal e aquelas que tenham assistido ao episódio *Nosedive*, da série britânica do Netflix, responderiam a essa teoria com um simples comentário: *Isso é tão Black Mirror*[1].

O episódio conta a trajetória de uma mulher, obcecada com sua pontuação – já que até mesmo para se mudar para um apartamento melhor, ela precisa manter uma média elevada. Ao longo do episódio, sua média, antes admirável, começa a cair e, como consequência, a protagonista começa a ser impedida de comprar passagens de avião, alugar carros novos, e frequentar determinados locais. Em sua trajetória, conhece uma motorista de caminhão, que ficou viúva de seu marido acometido por um câncer, que não pôde realizar tratamentos porque sua nota não era suficiente para acessar aos cuidados médicos devidos. De fato, esse mundo parece alguma teoria da conspiração fora do universo da ficção – comparável ao controle das máquinas sobre humano, em um futuro próximo.

A verdade, no entanto, é que vivemos em uma sociedade em que a avaliação das pessoas é determinada por máquinas. Nossas chances de empregabilidade, os limites dos nossos cartões de crédito[2], e o grau de periculosidade que representamos para a sociedade (AGWIN, et al, 2016; HAO e STRAY, 2019) são determinados por tecnologias de inteligência artificial. Há algum tempo, pessoas têm seus créditos negados de acordo com sua orientação sexual ou por conta da forma pela qual elas cuidam do quintal de suas casas (SOSNICK, 2016, p. 22). O valor do preço dos planos de saúde pode variar de acordo com nossos padrões de consumo alimentar ou pesquisas no Google (SOSNICK, 2016, p. 23-24). Modelos matemáticos, que

[1] IMDB. Nosdive, Black Mirror. 21/10/2016, Disponível em <https://www.imdb.com/title/tt5497778/>

[2] GLOBO. Algoritmos reproduzem machismo e racismo por se basearem em práticas discirminatórias humanas. Novembro, 2019 disponível em: <https://oglobo.globo.com/economia/algoritmos-reproduzem-machismo-racismo-por-se-basearem-em-praticas-discriminatorias-dos-humanos-24085081> acesso em: 17/11/2019

pouco sabem sobre nós – e simplificam diversos comportamentos para fazer inferências acerca de nossas personalidades (O' NEIL, 2016, p. 105-112) – são instruídos a fazer correlações, muitas vezes injustas, sob o pretexto de otimizar e agilizar tomadas de decisões.

O cenário – inicialmente, absurdo - do episódio de *Black Mirror* só não é mais real porque, diferente do mundo ficcional, somente os detentores de poder é que têm acesso às avaliações e aos parâmetros utilizados para negar resultados positivos a algumas pessoas. Como será que ficam os grupos já marginalizados nesse mundo em que as decisões são tomadas por máquinas? Será que os valores discriminatórios são minimizados, a ponto de a inteligência artificial ser desejável para promoção de diversidade? Há a crença de que a inteligência artificial pode minimizar decisões injustiças em face de mulheres, frequentemente penalizadas por decisões humanas enviesadas, tomadas com base em estereótipos que justifiquem a sua segregação.

Em um mundo marcado por noções sexistas, no entanto, a inteligência artificial pode ser utilizada para reforçar representações de gênero estereotipadas. Assim como tudo o que nos cerca, algoritmos automatizados são desenvolvidos por humanos, orientados, em grande parte, por uma ótica machista, que atribuí às mulheres apenas funções complementares aos homens. Esses valores, claramente discriminatórios, são incorporados por sistemas automatizados de tomada de decisão, e a discriminação é reforçada, ao invés de minimizada. Um dos grandes problemas é que, mesmo quando não há intenção de discriminar (ou quando há a intenção de efetivamente produzir efeitos mais justos), padrões sociais discriminatórios podem ser aprendidos ou reproduzidos por sistemas automatizados, tornando ainda mais essencial entender como isso acontece e como é possível combater essa nova forma de reforço de preconceitos de gênero.

Nesse sentido, com o objetivo de alertar ao estabelecimento de uma sociedade em que a grande parte das nossas decisões são tomadas por máquinas, o presente trabalho irá explorar as possíveis formas de discriminação automatizada contra mulheres. Diante da magnitude de seus efeitos, irei focar, principalmente, na discriminação perpetrada contra mulheres no ambiente de trabalho por meio do uso de inteligência artificial. Como essa forma de discriminação começa pelas barreiras que mulheres encontram para acessar o espaço público, o foco será igualmente no começo do processo de busca por trabalho: o processo seletivo.

Para isso, no primeiro capítulo irei esclarecer (i) como algoritmos podem se tornar discriminatórios, a partir da incorporação de vieses de diversas fontes, (ii) e quais as

características inerentes às tecnologias de inteligência artificial que as tornam perigosos meios de reprodução de estereótipos.

No segundo capítulo, passo a apresentar o cenário de discriminação de gênero no ambiente laboral, segundo a divisão sexual do trabalho, de acordo com o trabalho Sexual Harassment of Working Women, de Catharine MacKinnon. Em seguida, apresento as formas pelas quais essa discriminação de gênero ocorre de forma automatizada no mercado do trabalho.

No último capítulo, utilizo o exemplo da HireVue, companhia que oferece serviços de entrevistas preditivas em vídeo, para ilustrar como a discriminação de gênero pode ser reproduzida pela inteligência artificial. A tecnologia da empresa foi escolhida seu funcionamento possibilita introdução de vieses de todas as categorias apresentadas no trabalho, além de apresentar as características perniciosas de um algoritmo.

1. Como os algoritmos discriminam

Sistemas informatizados são desenvolvidos por indivíduos com preferências e ideais moldados por suas experiências sociais, que acabam fazendo escolhas que levam à reprodução de seus valores nos modelos adotados por seus sistemas. Ao adotar uma visão mais restritiva, é possível argumentar em favor de uma inevitabilidade de algoritmos serem marcados por preferências, valores e vieses dos seus desenvolvedores, justamente por se tratarem de criações humanas.

Para aqueles que defendem adventos da revolução cibernética como as novas promessas de progresso econômico, político e social, por outro lado, soluções tecnológicas voltadas para tomada de decisão são vistas como remédios para a presença de incentivos intuitivos. Não por outro motivo, diversas companhias e atores estatais desenvolvem sistemas de *machine learning* movidos pelo desígnio (ao menos em seus discursos) de reduzir a relevância de impressões pessoais em processos de tomada de decisão, já que decisões pautadas por preferências e valores pessoais podem gerar uma desconsideração de aspectos técnicos ou ameaçar o bem-comum.

Esse é o caso do COMPAS, algoritmo utilizado pelo sistema criminal dos Estados Unidos da América (EUA) que fornece uma "pontuação do risco" para cada réu ainda em fase de julgamento, com relação às chances de esse voltar a cometer novos crimes (HAO e STRAY, 2019). O objetivo do sistema é oferecer parâmetros, não vinculativos, para que juízes tomem decisões menos baseadas em suas intuições ou preferências quando deliberando acerca da necessidade de prisão preventiva.[3]

No entanto, até mesmo quando estamos lidando com as mais nobres intenções, é possível que vieses sejam inseridos em algoritmos, de forma que os resultados (*outcomes*) da sua operação reproduzam discriminações de grupos previamente marginalizados: após uma investigação realizada pela ProPublica (ANGWIN, et al, 2016), confirmada em artigo publicado na revista digital do MIT, a Technology Review (HAO e STRAY, 2019),

[3] Importante destacar que, não obstante a presunção de que o objetivo declarado de determinados algoritmos é o objetivo verdadeiro dos mesmos, o presente trabalho reconhece a sua limitação em determinar se há ou não interesses escusos por trás da implementação do algoritmo, os quais poderiam se resumir a apenas um desígnio de fornecer justificativas "numéricas" - as quais, normalmente, são dotadas de maior confiança - para perpetuar o encarceramento seletivo. Essa inferência seria possível quando consideramos que o algoritmo é um produto oferecido por atores privados, com interesses meramente lucrativos. "*The algorithm used to create the Florida risk scores is a product of a for-profit company, Northpointe. The company disputes our analysis*" (AGWIN, et al, 2019)

verificou-se que a probabilidade de o COMPAS atribuir elevadas pontuações de risco de reincidência é duas vezes mais alta para negros do que para brancos, até mesmo quando o sistema considera apenas réus que nunca haviam sido presos anteriormente.

Não obstante os prejuízos sociais da expansão dessa forma de tomada de decisão que se pretende neutra serem gritantes, a preocupação social com relação a episódios como esse ainda é latente. Em grande parte, esse cenário de despreocupação se origina diante da falta de compreensão com relação ao meio pelo qual vieses são introduzidos em sistemas informatizados (e, consequentemente, pelo desconhecimento da existência dessa - não tão - nova forma de discriminação).

Para trazer luz ao problema e para que seja possível começar a esboçar os possíveis caminhos de respostas jurídicas em vista a mitigar os potenciais negativos da discriminação incorporada em algoritmos, o presente capítulo irá apresentar uma categorização de vieses com bases nas suas fontes de origem, para, em seguida, identificar os atributos de modelos matemáticos lesivos que fazem com que suas ameaças tenham alcances maiores e sejam mais difíceis de combater.

Nem todo algoritmo, por mais repleto de vieses que seja, irá impor, necessariamente, uma discriminação em face a determinado grupo de indivíduos: outros aspectos - como quais são os efeitos de sua implementação, quais dados utilizados, ou quais os objetivos almejados pelo uso do software – precisam ser levados em conta conjuntamente com a existência de vieses antes de denunciar o caráter discriminatório de um algoritmo.

A necessidade de estabelecer, preliminarmente, qual o conceito de viés que será utilizado ao longo do trabalho é forçosa. Desde já, aponto que será utilizada a definição oferecida pelas autoras Batya Friedman e Helen Nissenbaum, em seu trabalho "*Bias in computer systems*", que determina que são enviesados os sistemas que "sistematicamente e injustamente discriminam certos indivíduos ou grupos de indivíduos em favor de outros"[4], desde que tais discriminações originem resultados não desejados, que sejam inapropriados ou razoáveis.

[4] No original: "(...) [computer systems that] systematically and unfairly discriminate against certain individuals or groups of individuals in favor of others" (tradução minha).

1.1 Categoria de vieses

A categorização utilizada no presente trabalho é uma reinterpretação da classificação de sistemas enviesados oferecida também por Friedman e Nissenbaum. A releitura da classificação das autoras, que leva em conta apenas vieses incorporados nos próprios modelos matemáticos, foi necessária já que algumas subcategorias não são necessárias para o escopo do presente trabalho, como se verá adiante. Além disso, foi acrescida uma quarta categoria, que leva em conta vieses introduzidos nos dados usados para operar as instruções de um algoritmo (e não nos modelos matemáticos). Em 1996, enquanto as autoras publicavam o trabalho oferecendo uma classificação, o fenômeno que hoje faz essencial a adição dessa classe de vieses estava ainda se consolidando. Foi apenas no ano seguinte, em 1997, que foi rotulado o Big Data, advento tecnológico que propiciou a evolução de métodos de inteligência artificial, como o *machine learning*[5].

Optei por adicionar tal categoria tendo em mente a relevância da tecnologia de Big Data nos dias de hoje, essencial para viabilizar a coleta, o armazenamento e a manipulação da quantidade colossal de dados que a inteligência artificial requer. Essa categoria é importante porque, quando tratamos de algoritmos utilizados para viabilizar técnicas de *machine learning*, a escolha, quantidade, e padrões dos dados vai direcionar o resultado buscado pela implementação da tecnologia de inteligência artificial.

A seguir, explicarei cada uma das diferentes naturezas de vieses apresentadas por Nissenbaum e Friedman: vieses pré-existentes, vieses técnicos, vieses emergentes, que podem ser verificados concomitantemente em um só algoritmo, juntamente com os vieses incorporado pelos dados.

1.1.1. Vieses pré-existentes

Vieses pré-existentes são aqueles que antecedem o desenvolvimento do algoritmo, i.e., **encontram seus fundamentos na cultura, nas instituições, em valores, preferências, ideologias.** Vieses pré-existentes podem ser originados a partir de inserções conscientes e voluntárias de valores e preferências ideológicas individuais do desenvolvedor ou do cliente,

[5] "Visualization provides an interesting challenge for computer systems: data sets are generally quite large, taxing the capacities of main memory, local disk, and even remote disk. We call this the problem of big data. When data sets do not fit in main memory (in core), or when they do not fit even on local disk, the most common solution is to acquire more resources" (ELLSWORTH e COX, 1997, p.1)

que requer que o funcionamento do programa proporcione resultados alinhados aos seus valores.

Nesse caso estaríamos tratando uma discriminação explícita (SOSNICK, 2016, p. 5), pela qual bens ou resultados positivos são negados a um determinado grupo por conta de atributos que a ele sejam inerentes. É o caso de um modelo de coleta e análise de dado para concessão de crédito que levavam em consideração gestos afeminados como um dos critérios para decisão (PASQUALE, 2015)[6] . A opção de restringir o benefício de crédito a grupos de pessoas com esta característica era determinada pela própria instituição financeira. As elites controladores das instituições financeiras, antes mesmo de contratar o responsável por desenvolver o algoritmo, já determina limitações à grupos específicos de forma intencional, movido por preconceitos, preferências discriminatórias amplamente difundidas na sociedade. Grupos já marginalizados, por sua orientação sexual, ou identificação de gênero, são discriminados por parâmetros voluntariamente fornecidos aos algoritmos[7].

A discriminação pode ser produzida também através de uma reprodução involuntária de valores. Na categorização de Friedman e Nissenbaum, vieses pré-existentes poderiam ainda ser fruto de um movimento inconsciente do desenvolvedor: neste caso, não é perquirido o objetivo de promover resultados discriminatórios. A inserção das preferências pessoais seria consequência de institucionalização de valores, ou sua ampla dispersão na cultura da qual o desenvolvedor faz parte.

O resultado de um experimento realizado por Joel Cooper e Charles Huff, já em 1987, nos fornece um retrato dessa forma de discriminação, para além de evidenciar a que essa diferenciação é desimportante, apenas para os fins do objetivo proposto no presente trabalho. Na hipótese, ao selecionar 43 educadores[8] para desenvolver um software que ensinasse às crianças o uso adequado de vírgulas, os autores constataram que, quando era pressuposto que os usuários do programa seriam meninos, os sistemas eram mais atrativos do que aqueles desenvolvidos para meninas (HUFF, COOPER, 1987).

---

[6] "*From Credit History to Score: The Original Black Box. Credit Bureaus pioneered black box techniques, making critical judgments about people, but hiding their methods of data collection and analysis. In the 1960s, innuendo percolated into reports filed by untrained investigators". They included attributes like messiness, poorly kept yards and "effeminate gestures*"(PASQUALE, 2015, p. 22)

[7] É necessário considerar que algoritmos são instruções/receitas necessárias para alcançar uma resposta, não necessariamente de forma automatizada.

[8] Foram selecionados profissionais com uma média de 2,2 linguagens programação cada, e uma média de 15,7 anos de experiência em educação,

Os programadores haviam sido separados em três grupos distintos, nos quais os educadores deveriam desenvolver softwares: (i) exclusivamente para o uso de meninos; (ii) exclusivamente para meninas e (iii) para meninos e meninas. A hipótese dos autores de que estereótipos de gênero eram levados em consideração na hora do desenvolvimento foi confirmada, e eles puderam aferir que era os desenvolvedores presumiam que meninos teriam maior pretensão a gostar de jogos (que foram, no caso, os softwares posteriormente classificados como mais atrativos para crianças). Os autores notaram ainda que, no caso dos programas idealizados para meninas e meninos, o funcionamento do sistema era mais similar àqueles desenvolvidos para meninos, que eram, essencialmente, jogos. Ou seja, pressupunha-se que a maior parte de usuários de computadores seriam meninos.

Em um só experimento, foi possível extrair dois diferentes casos em que vieses pré-existentes, pautados em preconceitos de gênero arraigados na sociedade[9], foram introduzidos em programas, afastando meninas do uso do software, já que aqueles desenvolvidos para elas não era tão estimulante quanto àqueles desenvolvidos pensando em usuários masculinos. E, apesar do viés não ter sido inserido conscientemente, com o desejo de afastar de fato as meninas do aprendizado da vírgula, sua inserção foi realizada por meio de uma decisão ativa de programadores. Por isso, não se faz necessária, para o trabalho, a presente distinção, entre vieses pré-existentes individuais ou sociais. Ambos são frutos de um descaso com a perpetuação de vieses e valores (seja intencional ou não) e, através de atitudes ativas do programador, promovem os mesmos resultados discriminatórios.

1.1.2. Vieses técnicos

Vieses técnicos, por sua vez, não encontram suas origens em escolhas (conscientes ou não) humanas, mas sim em **limitações técnicas variadas**. Alguns exemplos de vieses técnicos são oferecidos por Friedman e Nissenbaum, como falhas no desenvolvimento do programa, ou a falha em gerar números aleatórios, de modo que um grupo de números seja sempre favorecido, criando uma falsa sensação de aleatoriedade. Tamanho do dispositivo, limitações próprias de hardwares, também são outros exemplos de limitações técnicas próprias da máquina (computador).

No entanto, para o presente trabalho, a hipótese mais relevante trazida por Friedman

[9] Talvez aqui inserir algum dado numérico que apresente o estereótipo de gênero de mulheres serem menos tendentes/capazes de atuar em áreas exatas e lidar com sistemas computacionais.

e Nissenbaum diz respeito à dificuldade (ou, em alguns casos, até mesmo impossibilidade) de quantificar informações qualitativas da vida humana. Trata-se da falha em "formalizar construções humanas" e na tentativa de torná-las legíveis por computadores, em "discretizar" o contínuo (FRIEDMAN, NISSENBAUM, 1996).

Vetores de palavras (*word-embeddings*), tecnologia utilizada para o processamento de linguagem natural (NPL, na sigla em inglês), abrem margem para uma presença expressiva de vieses técnicos dessa natureza. Trata-se da representação de palavras como uma sequência de números, em que sequências próximas umas às outras revelam a proximidade do significado das palavras correspondentes. Essas sequências numéricas são organizadas em redes, nas quais a representação de cada palavra é expressa por um vetor ou partes de vetores. Para isso, contexto da palavra também é levado em consideração no processamento de palavras. É uma espécie de dicionário (na linguagem de computação, chave-valor) de classificações para programas de computadores. (BOLUKBASI, et al. 2016). Essa classificação, muitas vezes, é feita a partir até mesmo das proximidades geométricas entre palavras, considerando o emprego usual dela nos dados utilizados para ensinar à máquina como fazer a classificação das palavras. "Por exemplo, Inteligência Artificial tem a habilidade de objetivamente preencher a palavra "rainha" na sentença 'Homem está para rei, assim como mulher está para X' (FEAST, 2019)[10].

Para viabilizar essa tecnologia, é preciso que as palavras passem por uma classificação humana. E para quantificar o qualitativo, é preciso fazer escolhas e aproximações, sempre deixando algo de fora. Essas escolhas, normalmente, são feitas por homens[11], que podem inserir, propositalmente ou não, vieses pré-existentes, pautados na divisão sexual do trabalho. Segundo artigo "*4 ways to address gender bias in ai*", publicado na Harvard Business Review, "a grande maioria de sistemas de inteligência artificial comerciais utilizam *machine learning* supervisionada" (FEAST, 2019)[12]. No *Machine learning* supervisionado, os dados do grupo teste, que são usados para informar como a máquina deve se comportar, são rotulados, para que sejam legíveis para o computador. Assim, pessoas dotadas de vieses introduzem (de forma consciente ou não) padrões

[10] No original: *For example, AI has the ability to objectively fill in the word 'queen' in the sentence 'Man is to king, as woman is to X* (tradução da autora)
[11]Veremos mais adiante as diferenças de distribuição de força de trabalho entre mulheres e homens na área de inteligência artificial.
[12]No original: "*The vast majority of commercial AI systems use supervised machine learning"* (tradução da autora)

discriminatórios em modelos de *machine learning* (FEAST, 2019). Ao quantificar (em sequência de números) palavras – essencialmente, informações qualitativas-, humanos imprimem vieses e fazem correlações até mesmo para relacionar o significado de palavras (e identificar o contexto em que é utilizada).

1.1.3. Vieses decorrentes

Vieses decorrentes emergem por conta de alterações dos contextos **regionais/sociais** em que o sistema é executado. Tal categoria de vieses tem estreita relação com o fato de que desenvolvedores possuem expectativas com relação aos seus potenciais usuários. Esse tipo de viés é identificado somente após o início da operacionalização do algoritmo, ou do surgimento de novos conhecimentos que não sejam incorporados aos algoritmos posteriormente. Os resultados desinformados podem promover prejuízos a grupos em detrimento de outros, sem bases razoáveis, por desconsiderar parâmetros essenciais. É o caso de um software de automatização de decisões legais, que deixa de incorporar em suas instruções um novo enunciado jurídico, ou uma decisão erga omnes que deve passar a ser levada em consideração.

A diferença de expectativa de contextos pode se dar ainda em face do distanciamento entre a presunção dos programadores em relação aos valores e nível de conhecimento de seus potenciais usuários, e quais são os seus verdadeiros valores. O mesmo experimento de Huff e Cooper pode ser utilizado também para exemplificar um processo desse gênero de incorporação de vieses. Em seu teste, os autores buscavam comprovar a tese de que "as expectativas que os desenvolvedores possuem em relação aos usuários do software o qual eles estão desenvolvendo são centrais para determinar a forma pela qual software irá interagir com o usuário" (HUFF, COOPER, 1987)[13].E ao analisar os padrões discriminatórios identificados, revelaram que suas origens decorrem de expectativas dos programados pautadas em estereótipos de gênero. Os valores que os designers pressupunham que meninas possuíam era, em realidade, distinto das reais preferências das alunas.[14]

O exemplo nos mune com uma forte demonstração de que um mesmo sistema pode ter vieses de diferentes categorias. Essa informação é de extrema relevância para evidenciar

[13]Do original: "The expectations software designers hold about the users of the software they design are central in determining the way the software they design interacts with the user." (tradução da autora).
[14] "Differences in programs occurs as function of the designer's expectations of the characteristics of potential users of the software and result in sex stereotyped software." (HUFF, COOPER, 1987)

como diferentes etapas do desenvolvimento - e também da implementação - de um algoritmo podem dar margem para inserção de padrões discriminatórios, que fazem com que sejam negados resultados positivos a determinados grupos. Assim, a atenção para identificar possíveis vieses deve estar presente a todo momento, inclusive no processo de seleção de dados.

1.1.4. Vieses de dados.

Como mencionamos anteriormente, mesmo quando há a pretensão de neutralizar enviesamentos, ao programar sistemas com o objetivo de tornar processos de decisão mais céleres, desenvolvedores alimentam modelos informatizados com dados produzidos pela própria sociedade[15]. "*Ai-enabled machines are only as smart as the knowledge they have been fed*" (ANGWIN, et. al 2019). No lugar da neutralidade, dá-se espaço para o reforço de preconceitos imbricados nas informações que produzimos e, consequentemente, para a reprodução, em escala, de desigualdades. Essa é a principal origem de *vieses de dados.*

Assim como os vieses pré-existentes, vieses de dados podem ser inseridos em um algoritmo por conta da **seleção de dados que refletem discriminações históricas,** práticas culturais institucionalizadas (SOSNICK, 2017). Ora, alimentar um algoritmo com dados produzidos ao longo de um período histórico em processos de tomadas de decisões por humanos nada mais é do que ensinar à máquina os valores, instituições e vieses que orientaram tais decisões.

O caso do algoritmo de seleção automatizada de currículos, desenvolvida pela Amazon, é um grande exemplo dessa categoria de viés. A grande companhia de e-commerce desenvolveu um software próprio para ranquear os currículos enviados como requisito para as vagas de trabalho na companhia, com o intuito de automatizar as tomadas de decisão para contratação. Para alimentar os algoritmos responsáveis por fazer essa seleção, foram utilizados dados coletados ao longo de 10 anos anteriores, a fim de que o programa pudesse reconhecer os padrões que determinavam a exclusão de determinados candidatos em processos seletivos anteriores, realizados por intermédio de indivíduos.

Em 2015, após realização de diligência prévia, a companhia optou por não utilizar o

[15]"*É que a criação dos modelos se dá com base nos dados disponibilizados em etapas anteriores ao momento da tomada de decisão. Se os dados utilizados -para o treinamento forem dotados de vieses, a máquina aprenderá e replicará.*". (MATOS, 2019, p. 569)

algoritmo: não obstante os esforços para neutralizar as decisões realizadas pelo programa, o algoritmo reproduzia vieses de gênero, preterindo candidatas mulheres, e favorecendo homens[16]. Isso acontecia pois, com base nos dados fornecidos como parâmetro para o alcance um resultado ótimo de contratação (i.e., selecionar candidatos que se assemelham aos profissionais que se destacam na companhia), aprendidos pelo algoritmo por meio da aplicação do método de *machine learning*, a exclusão de candidatos era feita reproduzindo os mesmos vieses percebidos em decisões de contratação realizada por humanos nos anos anteriores.

Ou seja, o algoritmo desenvolvido pela companhia não estava alcançando os seus objetivos precípuos, por conta do treinamento que recebeu para seu funcionamento, o qual foi feito com base em dados classificados por humanos, em um momento anterior. Trata-se da reprodução da discriminação de gênero de forma automatizada no ambiente de trabalho.

São diversas as formas pelas quais vieses podem ser impressos em dados organizados em grandes bancos, inclusive através da própria limpeza de tais dados e da manipulação dos mesmos para que sejam ordenados. É possível até mesmo que vieses de dados se originem diante da união de diferentes fontes de dados sobre diferentes grupos, que se inserem em contextos sociais, políticos e econômicos distintos, os quais, consequentemente, levam a condições de sucesso (com bens ou resultados positivos) distintas. (SOSNICK, 2017), como seria o caso de um algoritmo elaborado para gerar a nota do Exame Nacional do Ensino Médio, que não considera condições socioeconômicas (e até mesmo geográfica) de cada estudante.

Dada a amplitude de espectros da vida que passam a ser determinados por algoritmos de inteligência artificial, desde a empregabilidade, permeando até mesmo a liberdade de uma pessoa - a exemplo do COMPAS-, a identificação da origem de vieses que corrompem decisões é essencial para que se possa combater (ou minimizar ao máximo) a disseminação de padrões discriminatórios.

[16]Reuters. *Amazon scraps secret AI recruiting tool that showed bias against women.* disponível em: ttps://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G , acesso em 03 de junho de 2019.

1.2 . O perigo da discriminação automatizada

Para atingir os objetivos almejados (sejam eles identificar o caráter do consumidor, para, então, propor um anúncio em detrimento de outro; ou ranquear/classificar resultados em plataformas de pesquisa), algoritmos fazem uso de modelos matemáticos que, por meio de uma representação abstrata de processos sociais, simplificam a realidade para prever resultados em determinados contextos, pressupondo ações e fatos que poderão ser originados a partir de certos padrões de comportamento. Por ser, por definição, uma simplificação matemática da vida, escolhas precisam ser feitas acerca do que deve ou não ser incluído nos modelos: "*Nenhum modelo pode incluir toda a complexidade do mundo real ou as nuances da comunicação humana"*[17] (O'NEIL, 2016).

Tais escolhas, conforme apresentado repetidamente, irão refletir valores e ideias dos indivíduos que estão desenvolvendo modelos[18]. É claro que a coleta de dados (que deve almejar ser mais ampla o possível) é uma tarefa árdua - a qual depende da acessibilidade dos mesmos - e, na tentativa de automatizar decisões por meio de algoritmos, alguns grupos de dados sempre serão desconsiderados. No entanto, essa ausência de determinadas informações, pode levar a correlações discriminatórias, uma vez que dados essenciais são substituídos por outros, e aproximações genéricas (e irreais) são feitas com base em construções sociais preconceituosas amplamente difundidas. Assim, dados que sejam falsos, incompletos ou generalizados tendem a elevar as chances de que os algoritmos reproduzam tais representações, as quais buscam seus fundamentos em estereótipos socialmente consolidados. Mas o que torna essa discriminação automatizada não pior, mas tão ruim quanto a discriminação realizada por humanos?

1.2.1. Feedback Loop Pernicioso.

Quando não são testados constantemente (por meio de protótipos, grupos de testes, ou outros meios), as premissas/suposições perversas presentes nos sistemas simplificados passam a ser utilizadas como a própria justificativa dos resultados – discriminatórios - da

[17] Do original: "*No model can include all of the real world's complexity or the nuance of human communication*" (tradução da autora)
[18] "*Models are opinions embedded in mathematics".* (O'NEIL, 2016, p.21)

aplicação dos algoritmos. Sem revisões que reparem eventuais resultados discriminatórios não previstos após uma primeira operação do modelo as projeções realizadas por ele continuarão sendo falsas, genéricas, inconsistentes, tal qual os dados selecionados para informá-los, já que os próprios resultados discriminatórios servirão como informação para próximas vezes em que o algoritmo for utilizado para o mesmo fim. Tal fenômeno é o que a autora Cathy O'Neil descreve como "***pernicious feedback loop***" (O'NEIL, 2016).

O feedback loop é uma característica inata de ciências da computação. No entanto, ao tratar de algoritmos discriminatórios, o feedback loop torna-se, como pontua a autora, pernicioso: os algoritmos se alimentam de uma auto validação da discriminação promovida por eles próprios. Conforme esclarece Sosnick, "algoritmos podem tomar decisões para um grupo de pessoas que aumentem (ou diminuam) as chances de mesmo grupo de receber resultados negativos (ou positivos) no próximo ciclo" (SOSNICK, 2017)[19]. Esse caráter ainda aumenta no caso compartilhamento de dados em cascata, casos em que resultado de um algoritmo já enviesado seja usado como parâmetro para treinar decisões de um outro algoritmo. Uma discriminação pretérita serve como justificativa para uma restrição de resultados positivos em outras esferas.

1.2.2. Escala.

As reproduções de padrões discriminatórios promovidas por algoritmos, que limitam resultados positivos a determinados grupos da sociedade, são feitas ainda em **escala**. Dado o baixo custo de disseminação dos sistemas informatizados (quando comparados a sistemas manuais), basta que o algoritmo seja desenvolvido para seus impactos sejam amplamente propagados (FRIEDMAN, NISSENBAUM, 1996). Nesse caso, por exemplo, um mesmo sistema de tomada de decisão é facilmente replicado por diferentes companhias em diferentes lugares do mundo. Se antes decisões eram pautadas em impressões que poderiam ao menos variar de avaliador para avaliador, a aplicação em escala homogeneíza os valores que limitam resultados positivos a alguns grupos (O'NEIL, 2016). A restrição da pluralidade de valores e de grupos que são contemplados com *outcomes* ótimos é consequência lógica do fenômeno. E um grupo cada vez mais seleto delimita quais os valores devem prosperar e quem irá ou não se beneficiar deles.

[19] No original: "algorithms can make decisions for a group of people that increases (or decreases) their chances of receiving a negative (or positive) outcome in the next cycle" (tradução da autora)

Outra dimensão de escala se refere aos dados. Não só a tecnologia é aplicada em escala, mas os mesmos dados idênticos são utilizados para informar decisões de algoritmos desenvolvidos por companhias de diferentes áreas. A transferência de dados ocorre em cascata entre companhias de forma tão rápida, quanto a própria coleta de dados novos (por meio de raspagem de dados online). E todos esses dados são combinados de formas distintas para fazer correlações sobre indivíduos que podem melhorar ou piorar suas chances de obter resultados positivos. Da mesma forma, as correlações são feitas em cascata, já que as respostas anteriores são utilizadas por novas empresas para fazer novas correlações. E enquanto a privacidade de seres humanos é contraída (pelo uso em cascata de correlações de dados), o sigilo das empresas aumentam (PASQUALE, 2015)[20].

1.2.3. Falta de transparência.

A disseminação em escala de padrões de discriminação que reproduzem valores dominantes é também obscura, inteligível. A **falta de transparência** – chamada de "caixa preta" – dos modelos matemáticos dificulta a verificação dos dados utilizados para informar um algoritmo, das instruções fornecidas em escala, e até mesmo da amplitude do impacto discriminatório. Sem transparência, impossibilita-se até mesmo verificar se os objetivos declarados de algoritmos estão em concordância com o seu funcionamento, i.e., se o algoritmo de fato opera de forma neutra, e não possui *inputs* ou instruções que claramente levam a resultados discriminatórios.

A falta de transparência coloca o usuário em uma posição de elevada insegurança e hipossuficiência: o usuário do algoritmo não sabe nem mesmo as razões pelas quais lhes foram negados resultados positivos e, em uma instância institucional, como combater e quem responsabilizar caso os resultados negativos sejam frutos de um processo discriminatório. Isso ocorre em um cenário no qual o mesmo algoritmo de decisão é adotado em escala.

Vamos analisar o exemplo de um candidato que, participando de mais de um processo seletivo através da plataforma Vagas.com, é rejeitado para uma vaga de emprego ofertada por uma companhia Y. As companhias que realizam seus processos de contratação pelo Vagas.com contam com as tecnologias da plataforma para fases iniciais, como testes de personalidade ou lógica, e softwares de triagem e ordenação inteligente de currículos.[21] Os

[20] "Corporate secrecy expands as the privacy of human beings contracts."(PASQUALLE, 2015, p.)
[21] https://forbusiness.vagas.com.br/plataforma-recrutamento-selecao/

candidatos, por sua vez, podem utilizar o currículo elaborado dentro da própria plataforma para diversos processos.

Voltando à narrativa do indivíduo rejeitado, ao se candidatar para uma vaga da companhia Z, esse fracassa no processo realizado através da plataforma novamente. O currículo utilizado para avaliação de ambas as companhias foi o mesmo, gerado através do preenchimento de um formulário disponibilizado pela própria plataforma (utilizado para informar, por sua vez, o mesmo algoritmo de inteligência artificial que determina quais parâmetros serão levados em consideração para ambas as empresas). Da mesma forma, o mesmo teste de personalidade ou lógica é aplicado pelas duas companhias, e as respostas de um mesmo candidato tendem a ser as mesmas. Imersos em uma situação de completa desinformação (um breu acerca dos parâmetros), os candidatos não sabem quais foram os padrões que os colocaram no lado de fora da peneira, e não poderão corrigi-los, perpetuando sua situação de marginalização de oportunidades.

Além disso, quanto mais complexos forem os sistemas, mais ocultos serão seus vieses. Considerando, no caso de algoritmos utilizados em *machine learning*, que o treinamento, inerente ao funcionamento de tal modelo de inteligência artificial, depende ainda da classificação prévia das informações contidas no conjunto de dados que será utilizado para alimentar os algoritmos, o desconhecimento dos critérios utilizados para tal também oculta eventuais vieses camuflados por algoritmos ditos neutros.

Por isso, faz-se essencial entender as dimensões da falta de transparência - que se relaciona com medidas adotadas por companhias ou por atores estatais para dificultar ou impedir a compreensão de seus algoritmos, e as diferentes concepções do conceito de transparência no âmbito de cada área de conhecimento - conceitos que devem ser complementares.

Frank Pasquale, em sua obra Black Box Society, estabelece três estratégias utilizadas por companhias para manterem seus algoritmos obscuros: sigilo real, sigilo legal e a ofuscação. A verificação conjunta dos três leva à opacidade tecnológica. O sigilo real dispensa uma explicação mais aprofundada, já que nada mais é que criar mecanismos que impeçam o acesso a determinado conteúdo, como a necessidade do uso de senha para acessar um e-mail. Da mesma forma o sigilo legal ocorre quando o conteúdo não pode ser revelado por determinações normativas, que podem ser estipuladas por lei, regulações de autoridades da administração pública direta ou indireta, mas também por meio de estatutos de empresas

privadas.

A ofuscação é um mecanismo utilizado por companhias com o objetivo dificultar o acesso de suas informações por parte da sociedade, empregado quando o sigilo real ou legal é rompido por alguma razão. Foi o que ocorreu após a crise de *subprimes* de 2008, em razão da determinação legal de que os bancos norte-americanos deveriam divulgar, anualmente, seus históricos de créditos em uma plataforma oficial, como forma de minimizar as injustiças cometidas na avaliação para concessão de créditos nos EUA. Na ocasião, para ofuscar o site oficial, companhias começaram a criar sites próprios de divulgação de informações extraoficiais (PASQUALLE, 2015, p. 23). Jogando com as três estratégias – por vezes, o sigilo real ou legal, por outras, a ofuscação –, o setor privado consegue tornar seus algoritmos e instruções de tomada de decisão opacos e, independente do meio, conseguem se manter tão inteligíveis que indivíduos não sabem que ações e em que esferas podem ou não impactar decisões que podem moldar nossas vidas.

Tal opacidade se justifica no setor privado porque as companhias alegam que seus algoritmos representam seu verdadeiro capital social: sistemas são mantidos em segredo já que o resultado ótimo deles dependerá das escolhas realizadas por aquele programa específico. Assim, sistemas com a pretensão de neutralidade, que camuflam, porém, vieses e preconceitos sistêmicos, são dificilmente combatidos, já que os parâmetros utilizados para qualificar pessoas e preferências são desconhecidos. Não é possível questionar seus possíveis defeitos, repercutidos em grande escala pela sociedade. Por isso, autores empregam seus esforços para definir qual a transparência desejada. Divulgar integralmente o código, contrariando o desejo das companhias de mantê-los em segredo para que o valor da tecnologia seja mantido (e não seja replicado por outras empresas que queiram oferecer o mesmo tipo de serviço) não é, necessariamente, uma forma de conferir maior integridade aos sistemas das companhias. Afinal, as centenas linhas de código são inteligíveis para a maior parte da população e, mesmo que fossem compreensíveis (ou que venham a ser no futuro), os vieses se originam de outras fontes que não somente o desenvolvimento do código do modelo em si. Por isso, a busca pela transparência deve ir além da mera disponibilização do código por trás da inteligência artificial.

Desai e Kroll, em *Algorithm and the Law*, buscam oferecer possíveis soluções regulatórios para mitigar os efeitos da opacidade e, para isso, traçam duas dimensões de transparência de acordo com a área do conhecimento que esteja lidando com o problema de

algoritmos. Para o direito, a transparência se relaciona diretamente com a possibilidade de verificar o sistema e seu funcionamento, comportamento e resultado, de forma a permitir que o responsável por resultados que estejam em desacordo com a legalidade (ou com parâmetros estabelecidos como aceitáveis) seja responsabilizado no mundo político-legal. Segundo os autores Desai e Kroll (2017) há uma noção, no mundo jurídico, de que um jurista dotado de um mínimo conhecimento acerca do funcionamento de um sistema, se acompanhado por um expert em ciência da computação, terá plenas condições de fornecer as melhores respostas legais para mitigar processos discriminatórios automatizados.

Para a ciência da computação, na visão dos mesmos autores, a prestação de contas não envolve necessariamente a divulgação de todo o código: o importante é que o algoritmo funcione de acordo com as regras pré-acordadas. A transparência se relaciona, então, com o fornecimento de evidências consistentes que sejam suficientes para verificar o adequado funcionamento da tecnologia. O objetivo é esclarecer quais as os comandos e dados utilizados para informar as tomadas de decisões dos algoritmos, e os motivos pelos quais foram determinadas escolhas foram priorizadas em detrimento de outras por desenvolvedores e operadores de sistema. Dessa forma, os desenvolvedores podem ser responsabilizados por possíveis discriminações (DESAI, KROLL, 2017).

Hoje, não contamos com nenhum desses espectros de transparência, os quais poderiam ser aplicados de forma conjunta. Como mencionado ao longo de diversos exemplos, mesmo quando o algoritmo é desenvolvido para promover justiça, ou então não tem a intenção de promover discriminação, é possível que resultados enviesados podem se originar. Por isso, um nível de transparência que ao menos revele quais os dados oferecidos para a decisão e quais os resultados almejados é necessário que os indivíduos possam saber o que originou um resultado prejudicial e se esse resultado foi ou não justo, de acordo com os resultados almejados pelo sistema (e se esses objetivos são ou não igualmente justos)

2. Mulheres como alvo

Em "Algorithm of oppression", Safiya Noble relata que, ao realizar uma busca no Google com os termos "mulheres negras", os primeiros resultados eram, expressivamente, pornografia. A objetificação dos corpos negros femininos é ponto de partida para sua tese sobre o poder dos algoritmos em reforçar relações de opressão e modelos inatos de estereótipos racistas (NOBLE, 2018).

No último mês (Novembro de 2019), uma matéria circulou nos grandes veículos de comunicação dando notoriedade a outro exemplo: o cartão Apple Card é objeto de uma investigação em curso nos EUA, por conta reprodução de desigualdade entre homens e mulheres na concessão do limite de crédito, que é determinado pelo Goldman Sachs.[22]

Em matéria veiculada no portal Wired em 2017, Matt Reynolds[23], editor de ciência, denunciou enviesamento até mesmo nas classificações de filmes do IMDB, plataforma na qual a avaliação dos filmes é feita com base na média da pontuação atribuída por seus usuários. A média é modulada para minimizar desproporções que poderiam influenciar a nota geral, e.g., pontuações atribuídas por novos usuários são desconsideradas para não criar notas desproporcionais. Notas atribuídas por usuários de uma mesma região também são relativizadas a fim de evitar que um grupo de uma mesma localidade influencie, desproporcionalmente a avaliação de um filme. No entanto, a grande maioria do usuários é homem e os filmes melhores *ranqueados*, à época, eram filmes que visavam o público masculino (i.e., que atendessem às expectativas sociais de papéis masculinos) ou com atores homens como protagonistas[24].

---

[22] GLOBO. Algoritmos reproduzem machismo e racismo por se basearem em práticas discirminatórias humanas. Novembro, 2019 disponível em: <https://oglobo.globo.com/economia/algoritmos-reproduzem-machismo-racismo-por-se-basearem-em-praticas-discriminatorias-dos-humanos-24085081> acesso em: 17/11/2019

[23]WIRED. *You should ignore film ratings on IMDb and Rotten Tomatoes.* Outubro, 2017. disponível em <https://www.wired.co.uk/article/which-film-ranking-site-should-i-trust-rotten-tomatoes-imdb-metacritic> acesso em 29/11/2019

[24] "Take IMDb's top-ranked film for example – *The Shawshank Redemption*. Its score of 9.3 is based on the votes of around 1.86 million IMDb users. 1.2 million of those votes came from men. IMDb does tweak its rankings to lessen the influence of particular demographics, but men often make up over 70 per cent of the voters for any film. And it turns out that men tend to look much more favourably on films with more masculine themes, or male leading actors. A look at the ratings for *Sex and the City* demonstrates how divided the voting audience on IMDb is. Over 29,000 men gave the film an average rating of 5.8, while 43,000 women came up with a score of 8.1. A straight-up averaging of the scores gives it a ranking of 7.4, but IMDb's maths leaves it with a final score of 7."WIRED. *You should ignore film ratings on IMDb and Rotten Tomatoes.* Outubro, 2017. disponível em <https://www.wired.co.uk/article/which-film-ranking-site-should-i-trust-rotten-tomatoes-imdb-metacritic> acesso em 29/11/2019

O mesmo acontece no Rotten Tomatoes, mas de uma outra forma. Nesta plataforma, somente críticos aprovados são levados em consideração, e quanto maior o número de avaliações positivas, melhor ranqueado é o filme (e.g., um filme com 75 avaliações com nota 100 é pior avaliador do que um filme com 111 avaliações com nota 99). No Rotten Tomatoes, a maioria dos críticos é homem – até 2016, segundo a matéria veiculada no Wired, apenas 27% dos críticos mais relevantes eram mulheres. A opinião masculina é, outra vez, priorizada em relação à feminina.

É possível desenvolver um trabalho inteiro apenas com exemplos de vieses de gênero em tecnologias de inteligência artificial. Mulheres, assim como indivíduos que são parte de outros grupos marginalizados socialmente, são o principal alvo também da discriminação algorítmica, uma vez que nossa sociedade é repleta de valores construídos e transmitidos sob a ótica da hegemonia masculina. Esses valores, que fundamentam as desigualdades de gênero que se manifestam em todas as esferas da vida, são replicados em algoritmos. Em vez de abrandar, a inteligência artificial perpetua resultados negativos para mulheres em diversos processos de tomada de decisão. Inclusive (e sobretudo) no mercado de trabalho: empregos são definidos de acordo com o sexo (MACKINNON, 1979).

A dificuldade de ingresso no mercado de trabalho e a divisão do trabalho por gênero tem suas raízes profundas, e consequências estruturais para a sociedade e para mulheres. Essa situação de desigualdade acaba por impactar “as possibilidades de auto definição e as oportunidades disponíveis para as pessoas” (BIROLI, MIGUEL, 2014). Ignorar essas diferenças - ou replicá-las em sistemas automatizados - faz com que seja ainda mais difícil para mulheres superar as barreiras à entrada no mercado de trabalho. Sem emprego, não há dinheiro, e a autonomia financeira de mulheres é um pilar essencial para a sociedade dê um passo em direção à redução de discriminação de gênero.

Para entender as novas formas de discriminação de gênero automatizada no ambiente de trabalho, é preciso compreender primeiro os contornos sociais que limitam a vida de uma mulher, em um mundo marcado pela divisão sexual do trabalho. Alinhada com suas ideias acerca da expressão da opressão sofrida por mulheres e da consolidação do machismo nas estruturas e instituições sociais, irei me apoiar, principalmente, nas ideias levantadas por Catharine MacKinnon sobre a relevância do atributo gênero, no que diz respeito às condições de trabalho. O cenário de discriminação que será apresentado a seguir é refletido e reforçado pelas normas formas de tomadas de decisões por inteligência artificial.

2.1 Mulheres no trabalho (de mulheres)

As origens dos entraves que mulheres encontram para conseguir um emprego remontam à consolidação do liberalismo e à distinção entre o espaço público - visto como a esfera racional, protagonizada por homens - e o espaço privado - esfera emocional, na qual mulheres cumpriam sua função social. A imagem e papel social da mulher foi, durante muito tempo, construída em torno das necessidades masculinas. Autores liberais clássicos dedicaram parte de sua obra para justificar e "fundamentar" a restrição de mulheres ao espaço privado. Rousseau, por exemplo, atribui importância política ao papel da mulher na educação de homens, função precípua de mulheres: o cuidado materno era visto pelo autor como essencial para consolidação de um Estado em que os cidadãos (homens) deliberassem em busca do bem comum (LANGE, et. al, 2002).

Mesmo após a barreira inicial ser rompida e a força de trabalho feminina (majoritariamente branca e de classe média) passar a ser utilizada também no espaço público, estereótipos de gênero se mantiveram: a presença mais expressiva de mulheres continua sendo verificada em cargos associados ao papel social atribuído a elas[25].As mulheres negras e pobres já representavam parte da mão de obra marginalizada e, da mesma forma, passam a desempenhar funções associadas ao seu papel social (cuidados domésticos, por exemplo). Diferente do que ocorre com a maior parte dos homens (em geral, brancos), o mercado de trabalho para mulheres é um ambiente hostil, marcado por segregação horizontal, estratificação vertical e a desigualdade salarial(MACKINNON, 1979), aspectos que serão detalhados a seguir.

2.1.1. Segregação horizontal.

A situação de desvantagem de mulheres no trabalho começa na associação entre áreas de trabalho e gênero. As expectativas sociais de que a mulher deve cumprir um papel determinado pela cultura predominantemente masculina faz com que ocorra uma segregação horizontal, que estabelece a diferenciação *profissão de mulher* vs *profissão*. A ocupação profissional varia de acordo com o critério de gênero. Mulheres que conseguem ingressar em

[25]"Women tend to be employed in occupations that are considered "for women", to be men's subordinates on the job, and to be paid less than men both on the average and for the same job". (MACKINNON, 1979, p. 9)

ramos profissionais que desafiem os padrões sociais impostos a elas – trabalhos de "homem"- passam a ser vistas como exceções, símbolos femininos, demonstrando o peso negativo que o gênero feminino exerce no processo de conseguir empregos (MACKINNON, 1979).

A área de tecnologia da informação – e, sobretudo, inteligência artificial, principal foco deste trabalho - expressa bem a segregação horizontal. Em um futuro próximos, as perspectivas que esse quadro se altere ainda são desanimadoras (WORLD ECONOMIC FORUM, 2018). Segundo os dados do Relatório Global de Diferença de Gênero de 2018 do Fórum Econômico Mundial (WEF, na sigla em inglês), 78% dos profissionais de inteligência artificial são homens. No Brasil, a presença de mulheres cai para 14%.

A associação entre racionalidade-homem-tecnologia faz com que tecnologias de inteligência artificial sejam desenvolvidas praticamente sem diversidade de gênero. A falta de mulheres propicia a introdução de vieses pré-existentes no momento do desenvolvimento da tecnologia, já que homens não conseguem adiantar os possíveis cenários discriminatórios, por sequer compreenderem quais são e como se dão (e por serem os principais atores de promoção de preconceitos contra mulheres).

A divisão sexual do trabalho faz com que os empregos mal pagos e com menor status sejam reservados às mulheres. O restante fica para os homens: "*Women's work is defined as inferior work, and inferior work tends to be defined as work for women*" (MACKINNON, 1979). Assim, criou-se um imaginário no qual "profissões de mulher" tendem a ser marcadas por trabalhos não tão interessantes, complexos, já que as mulheres seriam parte do espaço privado, marcado pelo emocional (em detrimento do racional), predominantemente orientado pelo serviço, envolvimento com criança, e tarefas que homens não estão dispostos a realizar.

Uma pesquisa que tinha por objetivo verificar a reprodução (e ampliação) de estereótipos de gênero em processamento de linguagem por meio de *word-embedding* encontrou resultados que indicam que a tecnologia reflete a presença de segregação horizontal nos dias de hoje. A pesquisa utilizou os modelos públicos do *word2vec* para reconstruir, a partir de associações aritméticas, o contexto de palavras em vetores[26]. Os modelos foram treinados com matérias do Google News, que reuniam 3 milhões de palavras em inglês distintas (BOLUKBASI, et al, 2016, p.2). Uma das dimensões verificadas foi a

[26] Conferir tópico 1.1.2, para fins de melhor compreensão acerca do processo de *word-embedding, apresentado na* exemplificação de vieses técnicos (páginas 9-10).

divisão de ocupações por gênero[27]. Para isso, palavras em que o gênero é sugerido (*businesswoman, sister*) foram utilizadas somente como forma de entender os subespaços de gênero na rede de vetores de palavras (e quais palavras, sobretudo aquelas que possuem gênero neutro, se encontram subespaços de cada gênero). Diferente do exemplo utilizado para verificar as possíveis formas de vieses A partir das matérias, o algoritmo fez correlações extremas entre ocupações e gênero, a fim de verificar quais palavras são mais próximas do termo "ela" (*she*) do que do termo "ele" (*he*): para mulheres (termo ela), as ocupações encontradas nas direções dos vetores que se correlacionam com *she* são recepcionista, dona-de-casa, *socialite*, cabeleireira, estilista, enfermeira, babá; enquanto as ocupação relacionadas à direção do vetor homem eram maestro, capitão, arquiteto, guerreiro, filósofo, financistas, programador: *word-embedding contains biases in their geometry that reflect gender stereotypes present in broader society* (BOLUKBASI, et al, 2016, p. 2). Nesse caso, diferente do exemplo utilizado na explicação de vieses técnicos, o viés não foi introduzido a partir da rotulação do significado de palavras próprias (*machine learning* supervisionado), mas se originou da derivação dos dados, agrupados por variáveis que eram semelhantes nas palavras: os efeitos não eram esperados por programadores. Não foi, previamente, identificado um resultado final ótimo, i.e., a correlação entre as palavras *he* e *she* com profissões que refletem os estereótipos de gênero da divisão sexual do trabalho não eram desejados pelos programadores.

A partir dessas correlações, o sistema desenvolveu outras correlações automaticamente, que representam analogias por binômios, em que a primeira palavra é associada ao vetor mulheres, e a segunda, ao vetor homens (direções *she-he*, na pesquisa). "Analogias são uma forma útil de avaliar tanto a qualidade de uma palavra vetorizada, como verificar os estereótipos dessa qualificação" (BOLUKBASI, et al, 2016)[28]. Os pesquisadores categorizaram as analogias como apropriadas ou estereotipadas, de acordo vieses encontrados na fala de diversos trabalhos que responderam a questionários (*crowdsourcing experiment*). As impressões da qualidade dos vetores estereotipados reforçam as diferenças nos papéis sociais atribuídos a mulheres e homens, verificadas nos vetores de palavras (i.e., representação de palavras em sequências numéricas): dona-de-casa-comerciante, atrevida-esperto, amável-brilhante, costura-carpintaria, feminismo-conservadorismo, cosméticas-

[27] A pesquisa foi realizada utilizando matérias em inglês para o processamento de vetores. A língua ingles é marcada por maior neutralidade de gênero do que o português.
[28] No original: *"Analogies are a useful way to both evaluate the quality of a word embedding and also its stereotypes"* (BOLUKBASI, et al, 2016, p. 3)

farmacêuticos. Algumas analogias são mais sutis que essas, mas revelam como vieses na fala podem ser facilmente despercebidos: vôlei-futebol, *cupcake-pizza,* vocalista-guitarrista (BOLUKBASI, et al, 2016). A dupla função desse exemplo é justamente mostrar como os vieses que se originam a partir de *machine learning* são um reflexo de percepções sociais estereotipadas e quais são os atributos de gênero gerados a partir dessas impressões.

Não obstante ser possível perceber alguma alteração dessa percepção, na prática ainda há forte relação entre gênero e divisão do trabalho[29]. Pode ser que, na sociedade, a ideia de que mulheres devem desempenhar somente o papel historicamente atribuído a elas - de cuidados domésticos - não permeie como antes[30]. Mas poucos homens estão dispostos a abrir mão de seus privilégios para, efetivamente, começar a desempenhar essas funções, ou ceder espaço para força feminina em campos ocupados, tradicionalmente, por eles[31].

2.1.2. Estratificação vertical.

Outra dimensão da divisão sexual do trabalho é a estratificação vertical. Mesmo quando passamos a avaliar homens e mulheres trabalhando na mesma área/atuando na mesma profissão, mulheres ocupam cargos subordinados a homens. Aqui, a desigualdade de gênero ocorre de forma vertical, já que mulheres ocupam posições que dependem de homens para contratação, retenção e promoção (MACKINNON, 1979)

Conforme se extrai de matéria veiculada pela Gênero e Número, em 2017, é possível afirmar que os padrões observados por MacKinnon nos Estados Unidos em 1979 são observados no Brasil de hoje:

> *Segundo levantamento realizado pela agência Volt Data Lab entre abril e maio de 2017, em mais de 400 empresas listadas na Bolsa de Valores de São Paulo (Bovespa), apenas 215 cargos de alta gerência executiva (diretoria ou vice-presidência), dentre 2.043 verificados, são preenchidos por mulheres — meros 10,5%. Somente em 17 das empresas (4,4%) há mulheres como chefes-executivas (CEOs, na sigla em inglês). Dessas, algumas ocupam a presidência em mais de uma empresa do mesmo grupo,*

[29] Segundo Pesquisa realizada pela Oxfam Brasil, em parceria com o Instituto DataFolha, "64% concordam que o fato de ser mulher impacta a renda". A pesquisa levou em consideração uma amostra de 2.086 pessoas em nível nacional, por meio de entrevistas realizadas em pontos de fluxo populacional em 130 municípios distintos. (OXFAM BRASIL e DATAFOLHA, 2019)

[30] Segundo a mesma pesquisa, 86% das pessoas "discordam que mulheres deveriam se dedicar somente a cuidar de casa e dos filhos, e não trabalhar fora". (OXFAM BRASIL e DATAFOLHA, 2019).

[31] Segundo dados do IBGE, enquanto homens dedicam apenas 10,5 de horas semanais a cuidados de pessoas e/ou afazeres domésticos – 10,4 entre brancos, e 10,6 entre negros –, mulheres dedicam 18,1 horas semanais – 17,7 entre mulheres brancas, e 18,6 entre mulheres negras. (IBGE, 2019).

> *e o número total de mulheres CEOs cai para 12. A situação nos Conselhos de Administração não é melhor: dos 2.647 assentos efetivos verificados, apenas 203 eram ocupados por mulheres (7,7%)* (FROEHNER e SPAGNUOLO, 2017)

Mulheres dependem de homens para contratação, supervisão e demissão. A vulnerabilidade feminina é mantida (e às vezes reforçada) mesmo quando, de forma excepcional, mulheres quebram as barreiras horizontais da divisão do trabalho. Segundo os dados apresentados pelo relatório da Mckinsey, apenas 21% dos cargos de maior nível hierárquico (executivos sênior, *C-level* ou *C-suite*) de empresas são ocupados por mulheres - e apenas cerca de 4% são mulheres negras (MCKINSEY, LEAN IN, 2019)[32]. No entanto, os resultados da pesquisa apontam que o principal obstáculo encontrados por mulheres no alcance de posições de liderança é a dificuldade de promoção para o nível de gestor (o segundo nível hierárquico nos dados da pesquisa): a cada 100 homens contratados e promovidos, no primeiro passo, apenas 72 mulheres conseguem alcançar os mesmos resultados; o que faz com que, em dados globais, apenas 38% dos cargos em nível de gestão sejam ocupados por mulheres.

Mulheres seguem negociando salários e promoções em níveis iguais aos homens desde 2015, e não desistem dos seus empregos (nem mesmo para focar em família, já que menos de 2% dos empregados consideram deixar o emprego para tal fim)[33]. Mulheres estão adotando comportamentos que favorecem as chances de promoção. Mas ainda assim, a realidade denunciada por MacKinnon em 1979 segue nos dias de hoje e "*quanto maior o status do trabalho, maior a probabilidade de uma mulher ser paga marginalmente mais próximo ao rendimento do homem, e menor a probabilidade de mulheres ocuparem tais cargos, independentemente da preparação educacional "*(MACKINON, 1979, p.13-14)[34]

2.1.3. Desigualdade salarial.

[32] O estudo realizado pelas companhias Mckinsey and LeanIn foi realizado com o fornecimento de dados ou resposta a pesquisas de Recursos humanos, que contou com a participação de 329 companhias, que empregam mais de 13 milhões de pessoas, e também a partir de pesquisa sobre experiência no trabalho de mais de 68.500 funcionários. (LEANIN, MCKINSEY AND COMPANY, 2019)

[33] "Women are staying in the workforce and doing their part. Again this year, women are not leaving their companies at higher rates than men. Moreover, the vast majority of women and men who plan to leave their company intend to stay in the workforce—and less than 2 percent of employees are planning to leave to focus on family. Women are also asking for promotions and negotiating salaries at the same rates as men, and this has been true since 2015." (LEANIN, MCKINSEY AND COMPANY, 2019)

[34]No original: The higher is the job status, the more likely a woman is to be paid marginally closer to man's wage rate, and the less likely women are to occupy theses positions at all, regardless of educational preparation (tradução da autora). (MACKINNON, 1979, p. 14)

A consequência direta da segregação horizontal e da estratificação vertical é a desigualdade salarial, que faz com que práticas nocivas no ambiente de trabalho sejam mais facilmente aceitas, ao fundamentar a subordinação material. Quando não é justificada pela subordinação hierárquica no trabalho (reflexo da estratificação vertical), a divisão sexual do trabalho desvaloriza as profissões consideradas femininas pela segregação horizontal. Segundo matéria publicada pela Gênero e Número[35], em 82% das carreiras, homens ganham mais do que mulheres em 2016 - "*a visualização foi gerada a partir dos dados estatísticos do Cadastro Geral de Empregados e Desempregados – CAGED, relativos a todos os meses de 2016*" (GÊNERO E NÚMERO, 2017)

O que causa estranhamento é verificar que companhias preferem contratar homens mesmo quando podem contratar mulheres mais qualificadas, e por um salário menor[36] - já que, apesar de claramente discriminatório, tal prática é institucionalizada a fim de manter a subordinação material feminina (MACKINNON, 1979). Ou ainda, observar que, mesmo em áreas em que há escassez de mão de obra qualificada, como a inteligência artificial, a baixa presença de mulheres é verificada, independente da eventual disponibilidade alta de força de trabalho feminina capacitado (WEF, 2018).

Em ambos os cenários, a justificativa meramente econômica (ou de eficiência) para a contratação majoritária de homens perde força, enquanto a relevância do fator "gênero" ganha contornos mais claros. Quais as circunstâncias e motivações que justificam a perpetuação da definição do trabalho pelo gênero?

Catharine Mackinnon elenca de algumas possibilidades. A estrutura monopolista do capitalismo, por exemplo, faz com que seja possível que empresas que não são afetadas pela competição possam "gastar mais" por menos. As empresas que enfrentam maior concorrência, por outro lado, tendem a ter mais mulheres (MACKINNON, 1979, p. 16).

Os próprios incentivos de homens e mulheres ajudam a explicar a relevância do fator gênero (MACKINNON, 1979, p. 16). Homens já ocupam as posições de poder nas empresas - resquício histórico, inclusive, da ausência de mulheres no mercado formal de trabalho -,

---

[35] GÊNERO E NÚMERO. Qual a diferença salarial? disponível em <http://www.generonumero.media/interativos/salario-genero/index.html> acesso em 25/11/2019

[36] Importante ressaltar que apesar de ser uma prática extremamente discriminatória, e que essa diferença salarial entre mulheres e homens desempenhando a mesma função não se justifica (mesmo cargo e mesma posição, já que a ocupação de cargos em posições diferentes faz com que a função desempenhada não seja a mesma), companhias institucionalizam essa forma de discriminação.

fazendo com que tenham alternativas aos trabalhos mal pagos e com menos status (que poderão afetar sua masculinidade, já que questionam seu poder). Por outro lado, mulheres não se consideram aptas a ocupar determinados cargos, uma vez que a marginalização social afeta a autoestima feminina, levando-as a crer que não obterão sucesso em processos seletivos.

No entanto, o elemento central, na concepção de MacKinnon, para a perpetuação da desigualdade entre mulheres e homens no mercado de trabalho é a sexualidade (MACKINNON, 1979, p.16-23). Para homens, é desejável que algumas posições sejam ocupadas por mulheres, de forma que elas possam satisfazê-los, assim como faziam quando o espaço público era reservado apenas a eles. Nesses casos, as mulheres são contratadas como "mulheres" - ou de acordo com o que a sociedade espera que mulheres sejam:

> *É comumente observado que o trabalho de mulheres fora de casa tende a monetizar papéis e tarefas que mulheres tradicionalmente desempenham em casa para homens* [37] (MACKINNON, 1979, p. 18)

Para fins do presente trabalho, o argumento da sexualidade como fator decisivo para divisão sexual do trabalho é importante para reforçar a distinção entre papéis sociais de mulheres e homens e, consequentemente, os processos de socialização de cada um. Essa distinção é replicada em processos automatizados - mesmo quando há a pretensão de neutralidade. Se hoje, na concepção social, isso começa a se reverter, por muito tempo as relações de trabalho traduziam as relações domésticas entre mulheres e homens. Então, conforme argumento, dados históricos e as características desejáveis para cada cargo, moldados pela divisão sexual, representam fatores que levam à perpetuação desses estereótipos e à consequente discriminação perpetrada por tecnologias de inteligência artificial.

2.2 Discriminação automatizada no trabalho.

Quando Cathy O'Neil compara o modelo de contratação dominante no passado com o atual, ela pondera que, no final das contas, a contratação "cara-a-cara" não passava de um simples julgamento de empatia e identificação do contratante (ou outras pessoas com quem

[37] No original: *It is commonly observed that women's employment outside the home tends to monetize the roles and tasks women traditionally perform for men in the home.* (tradução da autora)

trabalhe e que admire) com o candidato ou candidata. (O'NEIL, 2016) [38]. Aqui, afirmo que esse julgamento ainda se perpetua hoje, mas de forma automatizada. O uso da inteligência artificial no processo seletivo é uma tendência atual - já que empresas buscam contratar pessoas de forma mais barata e rápida[39] - e está atingindo um novo estágio, ainda mais desconhecido e inteligível para pessoas, mas com potencial ameaçador.

Companhias usam a inteligência artificial para ordenar currículos (como é o caso da Amazon e do vagas.com), aferir personalidade - por meio de testes ou controle de atividades em redes sociais, fóruns e ferramentas de busca[40] - e selecionar o candidato mais apropriado para o cargo, de forma rápida e menos custosa. Uma vez que o candidato é selecionado, essas companhias continuam utilizando a inteligência artificial para controlar todos os aspectos da vida de seus funcionários[41]. Por meio de todos esses mecanismos não-testados, representações estereotipadas estão sendo mantidas (e invisibilizadas) no mundo do trabalho por algoritmos e grande parte da sociedade sequer sabe disso, dada a falta de transparência dos modelos utilizados.

Empresas movidas principalmente pelos incentivos de reduzir custos e diminuir o tempo de dedicação em processos de seleção utilizam serviços de ordenação de currículo automatizada para descartar, em uma primeira análise computadorizada, a maior quantidade de candidatos que puder. Quanto maior o número de pessoas rejeitadas em um primeiro corte, maior o valor do algoritmo para a companhia, pois mais tempo será poupado em etapas posteriores. Os parâmetros, então, são estabelecidos para cortar pessoas, e não necessariamente selecionar os melhores e mais aptos. Saber quais palavras exatas utilizar no currículo, para que o algoritmo não rejeite de pronto alguém que tenha incluído um termo errado[42], se torna vantagem em processos de contratação.

---

[38] "Candidates then usually faced an interview, where a manager would try to get a feel for them. All too often, this translated into a single basic judgment: Is this person like me (or other I get along with)? The result was a lack of opportunity for job seekers without a friend inside, especially if they came from a different race, ethnic group or religion. Women also found themselves excluded by this insider game." (O'NEIL, 2018, p. 106 e 107)

[39] "Companies faced with tens of thousands of job application don't want to deal with each one individually. It's easier and faster to let software programs crunch a few hundred of variables" (PASQUALLE, 2015. p. 34). "For most companies, those WDMs are designed to cut administrative costs and to reduce the risk of bad hires (or ones that might require more training). The Objective of the filter, in short, is to save money" (O' NEIL, 2016, p. 118)

[40] acreditam haver necessidade de inserir um exemplo aqui?

[41] "Workers routinely surrender the right to object to, or even know, about, surveillance.(...) Technology makes it easy for firms to record workers' keystrokes and telephone conversations, and even to translate speech into text and so, predictive analysts claim, distinguish workers from shirkers" (PASQUALLE, 2015. p. 35)

[42] "so job applicant must craft their resumes with that automatic reader in mind. It's importante, for example, to sprinkle the resume liberally with words the specific job opening is looking for." (O'NEIL, 2015)

O mesmo ocorre com testes de personalidade ou registro de atividades na internet. Saber o que as companhias buscam nas respostas de testes (ou padrões de resposta), ou saber quais tipos de atividade em rede - e o tempo que você gasta em cada uma delas - são ou não desejáveis para uma empresa pode fazer com que você seja contratado. Não saber pode fazer com que você seja desqualificado, muitas vezes, de forma injusta.

No caso do registro de redes sociais, ou registro de atividades em geral na internet, um problema maior surge no que diz respeito à possibilidade de exclusão de mulheres. Cathy O'Neil (2016, p. 120) traz o exemplo da Gild, companhia que monta um perfil de candidatos acordo com as habilidades adicionadas por cada um em suas redes sociais. No entanto, com intuito de fazer uma busca mais completa acerca da pessoa, a companhia utiliza um espectro maior, buscando padrões de comportamento em fóruns e sites diversos. Para montar o perfil, no entanto, e atribuir valor às atividades em redes, utilizam casos de sucesso: uma vez a companhia identificou um talento que passava muito tempo em sites de histórias em quadrinho japonesas; a partir de então, esse passou a ser um atributo recompensado pela companhia. Apenas nesse exemplo, duas problemáticas surgem do uso de registro de atividades para mulheres. O primeiro diz respeito ao tempo livre. Como visto, mulheres gastam mais horas de sua semana desempenhando tarefas domésticas., já que são inerentes ao papel social da mulher. A dupla jornada, no entanto, reduz o tempo disponível para que mulheres sejam igualmente recompensadas por seus registros na internet. Além disso, muitos desses espaços virtuais utilizados como padrões são marcados por práticas sexistas, o que pode provocar desconforto para mulheres[43]. Logo, mulheres estariam sendo penalizadas por se preservarem de preconceitos.

Para tentar aferir como vai ser a personalidade de alguém no trabalho - algo incerto, por natureza, que só pode ser melhor aferido no dia-a-dia profissional, de fato - é necessário fazer simplificações e estabelecer parâmetros. E como falado repetidamente ao longo do trabalho, aproximações tendem a ser inexatas e muitas vezes, injustas (O'NEIL, 2016).

A imprecisão de testes dessa natureza é tal que até mesmo a forma de os responder pode levar a resultados injustos. E ela é moldada pelo nosso processo de sociabilização, com base no que aprendemos. Normalmente, os parâmetros não levam em consideração o processo de socialização diferenciado entre homens e mulheres. Assim, respostas que

[43] "The fact that prospects don't spend six hours discussing manga every evening shouldn't be counted against them. And if, like most of techdom, that manga site is dominated by males and has a sexual tone, a good number of the women in the industry will probably avoid it" (O'NEIL, 2016, p. 121)

refletem a forma - marcada pela divisão sexual de trabalho - pela qual mulheres aprenderam que devem se comportar no ambiente de trabalho, ao longo de toda a história, podem levar a resultados indesejados.

Assim como os demais candidatos, mulheres não sabem que determinado comportamento não é desejado para vaga - pensada para homens. Elas pensam que, como sempre foi em suas experiências anteriores, os contratantes estão em busca de alguém que responda de forma receptiva e solidária a seus chefes (quase sempre, homens) independentemente do que pensam, independentemente de considerarem os caminhos escolhidos pelos seus superiores piores, menos eficientes do que outros. Afinal, ao longo da maior parte da história, ensinaram que mulheres deveriam satisfazer os padrões desejados por homens/seus chefes.

É claro que muitas vezes nem mesmo o contratante tem ciência de que suas aproximações e inferências acerca de padrões verificados em currículos, atividades em rede, ou respostas em testes de personalidade estão sendo sexistas. Tal desconhecimento, no entanto, é fruto da ausência de mulheres, como pontuado anteriormente, que possam antecipar os cenários de discriminação, desconhecido por homens. O problema é que, mesmo nas raras hipóteses em que alguém se dá conta de que um resultado injusto foi perpetrado pelo algoritmo, não há incentivos em consertar o que leva à discriminação. Exceto nas hipóteses em que casos de candidatas injustamente desclassificadas, com bases claramente sexista, tenham se tornado públicos.

É claro que as companhias, como um todo, têm uma preocupação em demonstrar que estão preocupadas com a diversidade de gênero: em 2019, das companhias utilizadas como parâmetro para pesquisa realizada pela consultoria McKinsey, 87% declarava que a consolidação de diversidade de gênero é uma de suas prioridades (MCKINSEY, LEAN IN, 2019). Afinal, não é bom para publicidade de uma empresa demonstrar descaso com a representação feminina, por exemplo, sobretudo daquelas que têm como principais consumidoras as mulheres. No entanto, na hora de colocar em prática, o comprometimento não é o mesmo: apenas 35% das empresas estabeleceram metas de diversidade com relação à representatividade de seus trabalhadores e, talvez por esse motivo, apenas 52% dos empregados afirmam que a companhia na qual trabalham estabelece como prioridade a diversidade de gênero.

Cathy O'Neil (2016) utiliza exemplos do mundo dos esportes para ilustrar modelos matemáticos positivos, a partir dos quais times realizam avaliações de estatísticas históricas para tentar prever jogadas, escalações, ou até mesmo resultados de jogos. Conforme a autora elucida, diferente do que ocorre nas empresas, erros nesses sistemas são revisados constantemente, já que qualquer equívoco pode levar a uma derrota do time. Os incentivos para revisão são claros, já que os efeitos negativos são veiculados ao vivo: com a perda em campo, o time é diretamente afetado, bem como seus dirigentes (que serão penalizados pelo público torcedor).

Em empresas controladas por homens, os incentivos inexistem: os dirigentes não serão diretamente afetados por decisões injustas que discriminem mulheres. Não se elas não souberem em quais bases foram desclassificadas e, caso tomem conhecimento, não dispuserem meios para tornar isso público, a ponto de afetar a imagem da companhia

Assim, tecnologias de contratação automatizadas são utilizadas frequentemente sem que sejam previamente testadas e, quando são, muitas vezes não corrigem erros que possam surgir, ou que são identificados a partir do uso do sistema (vieses decorrentes). Como apresento a seguir, essa situação está atingindo um novo estágio com efeitos ainda mais desconhecidos.

.

3. Mais um passo em direção à discriminação: o exemplo da HireVue

Além do ranqueamento automático de currículos, dos testes de personalidade, ou do controle de atividade em rede, entrevistas preditivas em vídeo estão sendo empregadas sem que as pessoas tenham ciência de como estão sendo avaliadas, o que está sendo avaliado e como elas podem estar sendo rejeitadas em bases injustas. Os potenciais efeitos discriminatórios ainda são incertos e, mesmo assim, a tecnologia já está se espalhando pelo mundo: a HireVue, empresa que presta esse tipo de serviço, já atende mais de 700 clientes em diversos países. Utilizada também pela Goldman Sachs, a empresa acaba de ter uma reclamação apresentada contra ela também nos Estados Unidos[44]. Fundada em 2004[45], a HireVue é uma companhia de desenvolvimento de tecnologias cujos objetivos são aprimorar, facilitar e agilizar processos de contratação de pessoal:

> *HireVue uses a combination of proprietary voice recognition software and licensed facial recognition software in tandem with a ranking algorithm to determine which candidates most resemble the ideal candidate*[46].

A avaliação do sistema é expressa por meio da ordenação dos candidatos, de acordo com a pontuação atribuída por um sistema de inteligência artificial. A intenção da HireVue é que a tecnologia desenvolvida por ela sirva apenas como orientação para decisão do contratante. No entanto, apesar da ordenação automática não ser mandatória, claramente exerce um grande peso. Empresas querem estar nos primeiros resultados de pesquisa do Google, pois sabem que as pessoas dificilmente irão consultar páginas seguintes em busca de opções melhores, já que os consumidores acreditam nas predições dos algoritmos, mesmo sem conhecê-los. Da mesma forma, não é desejável para um candidato não estar entre os primeiros resultados: seus empregadores não irão verificar os últimos da listagem. Afinal, para que serviria o *ranqueamento* automático se não para desqualificar o maior número de pessoas de forma mais rápida, e acelerar o processo de contratação?

---

[44] GLOBO. Algoritmos reproduzem machismo e racismo por se basearem em práticas discirminatórias humanas. Novembro, 2019 disponível em: <https://oglobo.globo.com/economia/algoritmos-reproduzem-machismo-racismo-por-se-basearem-em-praticas-discriminatorias-dos-humanos-24085081> acesso em: 17/11/2019

[45] FAST COMPANY. This Bot judges how much you smile during your job: Predictive AI applications like HireVue screen job candidates on around 25,000 data points, breaking down their words, voice, and facial expressions. 15/01/2019 disponível em: <https://www.fastcompany.com/90284772/this-bot-judges-how-much-you-smile-during-your-job-interview> acesso em 01/11/2019

[46] BUSINESS INSIDER. I tried the software that uses AI to scan job applicants for companies like Goldman Sachs and Unilever before meeting them - and it's not as creepy as it sound. 29/08/2019 disponível em <https://www.businessinsider.com/hirevue-ai-powered-job-interview-platform-2017-> acesso em 03/11/2019

O exemplo da HireVue, que será pormenorizado nesta seção, é útil para visualização de como uma tecnologia utilizada para pontuar/avaliar pessoas é capaz de reproduzir discriminação de gênero diante da incorporação de vieses de naturezas distintas. Além disso, os efeitos dos resultados enviesados fornecido pelos sistemas da HireVue refletem as características perigosas do uso de sistemas de inteligência artificial: a aplicação é feita em escala, os resultados discriminatórios se justificam em um feedback loop pernicioso e as instruções do algoritmo carecem de transparência. A seguir, passo a explicar, de forma breve, como a tecnologia da HireVue é empregada em processos seletivos e como a empresa desenvolveu seus sistemas de entrevistas preditivas em vídeo.

3.1. Como funciona a HireVue

A entrevista em vídeo é apenas uma das etapas de contratação quando os serviços da HireVue são empregados em processos seletivos. Desde a inscrição para a vaga, feita por meio de um formulário online, grande parte dos campos são preenchidos automaticamente pelas informações contidas em outras redes, como o *Linkedin*. Após um processo de seleção automatizado das informações curriculares, outros estágios/processos computadorizados identificados como perniciosos ao longo do trabalho, como jogos de personalidade[47], definem as chances de um candidato ter ou não sucesso na sua candidatura.

Até a última fase, os candidatos são avaliados somente por máquinas, as quais têm a pretensão de garantir, além da eficiência, maior diversidade na contratação (novamente, esse é apenas o objetivo declarado da companhia que desenvolveu o algoritmo). *"It's important to note that HireVue Assessments don't replace all person-to-person interviews"*[48]. A entrevista final é única fase presencial e somente participam, no entanto, candidatos com recursos suficientes para entender quais os parâmetros que devem seguir para que não sejam descartados por máquinas. Conforme apresentei no tópico relativo às formas discriminação automatizada no ambiente de trabalho, os candidatos não sabem quais são os parâmetros desejados pela companhia nos testes de personalidade e na atividade no *LinkedIn*, ou quais são os comportamentos desejados durante entrevista. Esses parâmetros desconhecidos são

[47] Assim como os testes de personalidade, buscam tentar prever como será o comportamento de um candidato no dia a dia de trabalho, por meio de jogos interativos, em vez de perguntas.
[48] LARSEN, Loren. Artificial INtelligence, Hiring Assessments, Video Interviewing. HireVue, 2018. disponível em <https://www.hirevue.com/blog/train-validate-re-train-how-we-build-hirevue-assessments-models>

padronizados pela máquina, reduzindo as chances de assegurar uma mínima diversidade nas entrevistas.[49]

As entrevistas em vídeo da HireVue, em um primeiro momento, consistiam em apenas a um serviço *on-demand*, cujo objetivo precípuo era possibilitar a avaliação de uma piscina de talentos maior em menos tempo. Sem a necessidade de agendamentos de entrevista., os entrevistados poderiam fazê-las onde e quando fosse mais conveniente, assim como os avaliadores poderiam verificá-las e avaliá-las no horário de preferência[50]. No entanto, esse serviço ainda possuía falhas. Muitas pessoas desistiam dos processos seletivos ainda por conta da demora dos empregadores em avaliar as entrevistas e, além disso, as companhias estavam contratando não os melhores, mas os que candidatos que estavam mais dispostos a enfrentar todo o processo.

Foi só então que a HireVue começou a utilizar os vídeos arquivados de processos seletivos antigos e combiná-los com *machine learning*, para realizar predições acerca do sucesso de cada candidato no dia-a-dia de trabalho, caso fosse aprovado no processo seletivo. Para viabilizar uma primeira aplicação desse novo estágio de avaliação automatizada, a HireVue informou seus algoritmos com entrevistas registradas por ela ao longo de todo o período em que prestava somente serviços de entrevistas *on-demand* – formato de avaliação que ainda exigia que os avaliadores assistissem a todos os vídeos, para informar as decisões que iriam tomar posteriormente). Ou seja, as entrevistas utilizadas como alimentar a inteligência artificial da HireVue eram submetidas a crivos enviesados.

Segundo Larsen, Chief Technology Officer (CTO) da HireVue, a companhia se assegurou de que um indicador claro para diferenciar os candidatos ruins dos bons estivesse presente desde o desenvolvimento do seu primeiro algoritmo. Direcionada por esse indicador, hoje a HireVue desenvolve modelos com um mínimo grau de personalização para cada processo seletivo e elabora perguntas diferentes para cada vaga, levando em consideração as habilidades exigidas para execução do emprego. Essas perguntas são utilizadas para avaliar uma população inicial – grupo teste –, que podem ser tanto possíveis candidatos, como com pessoas que trabalham na empresa contratante e que sejam

[49]HIREVUE. Unilever's Recruiting Process disponível em : <https://www.hirevue.com/resources/unilevers-recruiting-process>

[50] LARSEN, Loren. Reducing Bias and Widening the candidate pool: why built HireVue assessments. HIREVUE, 2018. disponível em: <https://www.hirevue.com/blog/reducing-bias-and-widening-the-candidate-pool-why-we-built-hirevue-assessments> acesso em 25/11/2019

consideradas casos de sucesso dentro da companhia. As avaliações do grupo serão consideradas como referencial para o modelo. Em seguida, os resultados dos grupos testes são observados para verificar a presença de vieses e realizar eventuais correções, antes de implementar o sistema. Periodicamente, a HireVue monitora os resultados, para assegurar que as avaliações estejam de acordo com os objetivos de seus clientes (os contratantes) e garantir que a predição é precisa[51].

Dado o funcionamento da tecnologia, há margem para incorporação de vieses desde o desenvolvimento do primeiro sistema. Isso porque a HireVue utiliza dados históricos de decisões tomada por seres humanos para desenvolver o modelo de predição das entrevistas em vídeo – e a própria companhia reconhece que essa forma de tomada de decisão é enviesada. Conforme argumento no presente trabalho, esses vieses poderão refletir padrões discriminatórios de gênero, pautados na diferença de papéis sociais de mulheres e homens[52].

Afinal, o referencial utilizado para as entrevistas preditivas de cada empresa contratante são, em parte, candidatos prospectivos que representam potenciais talentos para elas– e que tendem refletir representações estereotipadas de gênero, pautadas na divisão sexual do trabalho. De outro lado, dada a baixa diversidade de gênero em companhias (seja pela segregação horizontal, ou pela estratificação vertical), haverá uma escassez de dados femininos na hipótese em que os próprios empregados das contratantes são utilizados como referenciais, o que também poderá propiciar resultados negativos e injustos às mulheres.

Nem mesmo o monitoramento periódico é suficiente para minimizar as chances de incorporação de vieses de dados nos processos automatizados, uma vez que sempre será orientada por decisões humanas, em última instância: os candidatos que de fato obtiveram sucesso na contratação (formalizada pelo HireVue, através de contratos eletrônicos) foram, necessariamente, submetidos a um crivo presencial. E, conforme passaremos a analisar, o uso de dados históricos para informar a decisão dos algoritmos não é a única forma que abre margem para a incorporação de representações discriminatórias de gênero na inteligência artificial: a tecnologia da HireVue é passível de introdução também de vieses pré-existentes, técnicos e decorrentes.

---

[51] LARSEN, Loren. *Train, validate, re-train: how we build HireVue assessments models*. HIREVUE, 2018. disponível em: <https://www.hirevue.com/blog/train-validate-re-train-how-we-build-hirevue-assessments-models> acesso em 25/11/2019

[52]

3.2 Vieses nas entrevistas preditivas em vídeos

A ferramenta de entrevistas preditivas em vídeo analisa cerca de 25 mil dados a cada 20 minutos de vídeo, a partir da escolha de palavras do entrevistado, voz e expressões faciais, coletados com o objetivo de ranquear as entrevistas. A avaliação dos candidatos é expressa por uma nota, que varia entre 0 e 100 e expressões faciais representam menos de 30% desta nota – informação incerta, conforme verificaremos a seguir. O restante, é feito com base, principalmente, no uso da linguagem do candidato de suas habilidades verbais

3.2.1. Linguagem.

.Cerca de 350 características são analisadas na captação da linguagem do entrevistado (linguagem mais ou menos técnica, quão longa é a sentença, pronome utilizado).[53] Só no processamento dessa informação, é possível identificar vieses de diversas origens. Como vimos, vieses técnicos são inerentes às tecnologias de NPL. Nosso modo de falar é repleto de valores, e diversas palavras representam vieses de gênero sem que nós sequer nos demos conta. A fala reproduz o machismo imbricado na sociedade: o vocabulário da língua inglesa, por exemplo, comporta mais palavras que sexualizam mulheres, mesmo diante da existência de um universo muito maior de palavras que se refiram a homens (BOLUKBASI, et al, 2016, p. 3). E isso é processado pela máquina. É fácil perceber discriminação de gênero mesmo no emprego de termos em diálogos prosaicos do nosso cotidiano, repletos de gírias que favorecem atributos masculinos, e que serão identificados pelos sistemas informatizados.

Logicamente, ninguém irá empregar uma linguagem com esse nível de informalidade em uma entrevista de emprego, que é a natureza dos dados utilizados como grupos testes pela HireVue[54]. Mas as noções de favorecimento masculino imbricadas em tais gírias são reproduzidas (e suavizadas) também no emprego da linguagem mais formal. Se observarmos exemplos no outro extremo da formalidade da língua, é possível verificar que, dificilmente, o processamento de linguagem para posterior classificação será isento de vieses. Como verificado anteriormente, mesmo as pesquisas realizadas com sistemas de *word-embedding*

[53]THE DAILY TELEGRAPH. *AI used for first time in job interviews in UK to find best applicants.* 27/09/2019. disponível em: <https://www.telegraph.co.uk/news/2019/09/27/ai-facial-recognition-used-first-time-job-interviews-uk-find/> acesso em 05/10/2019

[54] Imagino que a HireVue utilize modelos primários para produzir redes de vetores de palavras com dados textuais que tenham uma mínima pretensão de neutralidade, mas a companhia não informa.

treinados com artigos do *Google News* revelam estereótipos de gênero – "*in a disturbing way*" (BOLUKBASI, et al, 2016).

A escolha de perguntas, as respostas e as representações numéricas de palavras são fatores que reforçam resultados enviesados. Por exemplo, as **perguntas** abrem margem para vieses decorrentes, que emergem da frustração das expectativas dos programadores com relação de valores de suas entrevistadas; as **respostas**, por sua vez, podem refletir vieses pré-existentes, dado o processo de socialização das mulheres; já as **representações numéricas das palavras,** praticamente de forma inevitável, apresentam vieses de técnicos e pré-existentes, já que levam em consideração modelos primários de vetorização de palavras, que consideram proximidades geométrica delas.

Não por outro motivo, em um de seus e-books, "2 ways to take video interviewing to the next level", a própria HireVue reconhece que perguntas que levam em consideração experiências de trabalho apresentam alto potencial em conter vieses, assim como anos de educação e escolha da universidade. No entanto, as companhias podem ou não optar por utilizar esse tipo de pergunta e, mesmo que não optem, é possível que candidatos falem de suas experiências passadas. Se, por exemplo, a correlação entre as palavras recepcionista e mulher não forem neutralizadas no próprio algoritmo, é possível que o próprio sistema aprenda a diferenciar homens e mulheres e, caso aprenda, poderá aprender também que mulheres, no histórico daquela companhia, não eram desejadas – ou eram, mas somente para desempenhar determinadas funções alinhadas com seu papel social.

3.2.2. Voz.

As entrevistas em vídeo da HireVue levam em consideração informações como tom e a velocidade da voz também[55]. São diversas as formas pelas quais vieses podem se originar a partir do processamento desses dados, mas apresentarei somente a os vieses inerentes ao reconhecimento da voz. Tecnologias de texto-para-fala (*speech synthesis*) por muito tempo operaram melhor para homens do que para mulheres (FEAST, 2019; HENTON, 1999). Até que grande parte passou a ter voz de mulheres, mesmo não soando tão naturais, e sim mais robóticas: a disseminação de assistentes digitais fez com que o emprego dessa tecnologia em

[55] THE DAILY TELEGRAPH. *AI used for first time in job interviews in UK to find best applicants.* 27/09/2019. disponível em: <https://www.telegraph.co.uk/news/2019/09/27/ai-facial-recognition-used-first-time-job-interviews-uk-find/> acesso em 05/10/2019

nosso cotidiano seja predominante feminina. Afinal de contas, a Alexa e a Siri desempenham o papel de secretária, tipicamente atribuídos a mulher (não seria tão amigável para os homens se ouvirem como assistentes[56]).

Da forma similar, a discriminação de gênero pode ser percebida em tecnologias de reconhecimento de fala automático (*automatic speech recognition*, em inglês), i.e., fala-para-texto. Nessa hipótese, no entanto, não há discriminação direta – em que o programador adota uma escolha (voz de mulher para assistentes digitais) a fim de reproduzir conscientemente papéis de gênero estereotipados. Um sistema que reconhece mal a voz de mulheres pode levar a "traduções" erradas de sua fala e palavras que não foram efetivamente ditas (ou palavras que foram ditas) podem ser erroneamente identificadas (ou omitidas). Isso pode ocorrer diante de dados insuficientes de falas de mulheres, ou do desenvolvimento de um sistema preparado apenas para homens[57] (HENTON, 1999). E como verificamos, em diversos estágios do desenvolvimento do sistema da HireVue, os grupos testes tendem a refletir dados históricos.

Nesse cenário, inadequações no sistema podem excluir, de forma ainda mais intensa, mulheres que estejam sendo entrevistadas qualquer que não seja sua língua nativa (o mesmo ocorre para homens estrangeiros também). Se a voz não é bem reconhecida, o tom e a velocidade também não serão processados da melhor maneira possível. E as inferências que se originam a partir do processamento dessas informações a partir serão incorretas e, provavelmente, injustas.

3.2.3. Expressões faciais.

Outro aspecto considerado pelas entrevistas em vídeo são as expressões faciais. O algoritmo avalia a quantidade de vezes que olhos se abrem ou fecham, o levantar ou cerrar de sobrancelhas, e até mesmo o sorriso[58]. Quando a rede de hotéis Hilton utilizou os serviços

[56] UOL NOTÍCIAS. Mulheres digitais Porque todas as assistentes virtuais têm vozes femininas?. disponível em: <https://www.uol/noticias/especiais/assistentes-de-voz-x-feminismo.htm#tematico-1 >

[57]"(...) female voices historically have been marginalized acoustically (and hence, disregarded in phonetic theory) owing to inadequacies in analytic hardware. That should be obviou to anyone who has wrestled with interpreting spectrograms of female voices. Until recently, the sound spectrograph has been the most frequently-used tool in acoustic speech analysis, and other instruments (such as narrow-band spectrum analyzers and pitch extraction tools) are still imperfect in analyzing females speech". (HENTON, 1999)

[58] THE DAILY TELEGRAPH. *AI used for first time in job interviews in UK to find best applicants.* 27/09/2019. disponível em: <https://www.telegraph.co.uk/news/2019/09/27/ai-facial-recognition-used-first-time-job-interviews-uk-find/> acesso em 05/10/2019

da HireVue para contratação, a empresa optou por dar maior relevância para a quantidade de vezes que o candidato sorriu ao longo da entrevista – parâmetro eliminatório do processo seletivo da rede de hotéis. Esse critério foi desenvolvido tendo como referência o que se demonstrou ter sido relevante para emprego a partir do grupo teste utilizado para informar o modelo[59].

Nessa hipótese, mulheres que fogem do papel social atribuídos a elas são excluídas, sem saber que os motivos para tal foram a insuficiência da quantidade de sorrisos. Como apresentado por Mackinnon, diante da construção social da imagem do que deve ser a mulher no ambiente de trabalho, algumas características passaram a ser desejadas para cada emprego, como era o caso de "gestos submissos" para cargos de telefonistas (MACKINNON, 1979). No processo seletivo do Hilton, mulheres que fogem desse padrão são desclassificadas de pronto. Neste exemplo, em que a ocupação envolve tarefas compatíveis com o papel social da mulher, a discriminação se dá de forma contrária ao que foi verificado quando avaliamos os testes de personalidade, mas ambas reforçam representações estereotipadas de gênero. Assim, noções sexistas acerca do papel social de cada gênero – e as formas pelas quais esses valores se relaciona com categorias de trabalho – são reforçadas – e aplicadas em escala – pelo algoritmo. Como isso poderá se reverter, se a discriminação é invisível?

A análise de expressões faciais talvez seja o critério mais problemático de entrevistas preditivas em vídeo, até mesmo porque depende diretamente de outros fatores externos. A luz poderá prejudicar a precisão da digitalização da imagem, bem como a qualidade da câmera utilizada. Se mulheres recebem menos, também é possível que tenham recursos tecnológicos mais precários para fazer as entrevistas e a baixa precisão de suas avaliações levará a resultados injustos. Entretanto, a falta de precisão claramente discriminatória ocorre não por conta de limitações dos instrumentos utilizados, mas por limitações técnicas fruto de reprodução de vieses sociais pela máquina.

Uma pesquisa realizada por Buolamwini e Gebru apontou para a identificação errônea de mulheres por tecnologias de análise facial, diante da escassez de dados femininos - sobretudo de mulheres negras. A pesquisa foi realizada com três diferentes tecnologias

[59]FAST COMPANY. *This Bot judges how much you smile during your job: Predictive AI applications like HireVue screen job candidates on around 25,000 data points, breaking down their words, voice, and facial expressions*. 15/01/2019 disponível em: <https://www.fastcompany.com/90284772/this-bot-judges-how-much-you-smile-during-your-job-interview>

comerciais de identificação de gênero e tipo de pele por meio de reconhecimento facial (BUOLAMWINI, GEBRU, 2018). Ou seja, Na aplicação comercial (objeto de estudo da pesquisa), que tem como principal foco aqueles que têm maior capacidade de consumo, mulheres negras são as principais discriminadas. Na aplicação no mercado de trabalho, mulheres (ainda sub representadas, sobretudo como casos de sucesso, que seriam utilizados como parâmetros) podem ser desfavorecidas diante de uma performance menos precisa de entrevistas realizadas com elas.

É claro que essas suposições são incertas, assim como o próprio funcionamento da tecnologia, como veremos adiante. Os exemplos de discriminação que podem se originar a partir do uso da HireVue são apenas hipóteses que se fundamentam em outras oportunidades em que tecnologias de inteligência artificial foram comprovadas discriminatórias, por refletirem vieses - na maior parte das vezes, de forma não intencional. Por mais incertas que sejam, essas suposições são plausíveis e até mesmo as mais absurdas podem ser reais, dada a opacidade característica de modelos matemáticos perniciosos. Considerando que a HireVue ainda não entrega um relatório para os candidatos (com o resumo da avaliação individual, identificação dos parâmetros universais e dos indicadores de performance), é possível supor até mesmo que vieses intencionais sejam responsáveis pela reprodução de representações estereotipadas.

Não poderiam as empresas contratantes, por exemplo, solicitar que referenciais sexistas sejam utilizados para orientar decisões do sistema da HireVue? Conforme *MacKinnon* (1979) elucida, "mulheres tendem a ser avaliadas economicamente de acordo com a percepção que homens têm sobre o potencial de elas serem assediadas sexualmente"(MACKINNON, 1979, p.23)[60]. Segundos os resultados da pesquisa mencionada, realizada pela *McKinsey*, 41% das mulheres entrevistadas reportaram ter sofrido assédios sexuais durante sua carreira, seja ao ser tocada de formas sexuais inapropriadas, ao ouvir piadas sexistas, ou ao ser alvos de tentativas de sexo não consentido. E dentre as mesmas mulheres entrevistadas, apenas 50% acreditam que reportar casos de assédio sexual é efetivo (LEANIN e MCKINSEY, 2019).

Além disso, já existem pessoas que empregam seus esforços para desenvolver tecnologias capazes até mesmo de determinar a orientação sexual de uma pessoa a partir de

[60] No original: [The observations suggests that]*Women tend to be economically valued according to men's perceptions of their potential to be sexually harassed* (MACKINNON, 1978, p. 23)

uma fotografia, com bases questionáveis e aparentemente repleta de vieses para suas predições[61]. Ou seja, há disponibilidade tecnológica e disposição moral para desenvolver uma tecnologia de inteligência artificial que avalie mulheres forma insatisfatória. Aquelas que desviarem do papel social atribuído a elas (de satisfação dos desejos masculinos) serão penalizadas, seja por não atenderem aos padrões de beleza estipulados por grupos de poder, seja por serem consideradas lésbicas por sistemas automatizados.

As possibilidades de discriminação são inúmeras, e as certezas de que a HireVue não as utiliza sob qualquer hipótese são mínimas. Não por outro motivo, conforme já mencionado, foi apresentada uma reclamação contra a companhia perante as autoridades competentes nos Estados Unidos. Entretanto, enquanto o processo segue pendente de análise, e a reprodução de representações estereotipadas e discriminações de gênero são plausíveis, a tecnologia segue sendo utilizada em larga escala.

3.3 Perversidade na tecnologia da HireVue

A reclamação apresentada contra a empresa evidencia o potencial discriminatório de sua tecnologia, já que questiona a presença de vieses de gênero em seus algoritmos. Mas o descaso dos demais atores do mercado com a discriminação é tal que os possíveis efeitos segregacionistas não servem como desincentivo para seu uso. Assim, a implementação da tecnologia de entrevista em vídeo em mais de 700 companhias promove uma padronização nos valores que serão dominantes - essencialmente, masculinos. As brechas para mulheres obterem êxito em driblar o status quo, tornam-se, então, menores. Isso não necessariamente significa que menos mulheres conseguirão ingressar no mercado, mas sim que determinadas características em mulheres serão valorizadas em detrimento de outras, o que poderá variar para cada vaga. E, a partir do uso disseminado de uma só tecnologia para diversos processos de contratação, as representações discriminatórias promovidas por ela são empregadas em escala.

Considerando a tecnologia de NPL, por exemplo, necessária para traduzir as falas das entrevistas em informações compreensíveis para máquinas, não é razoável esperar que a companhia utilize, para cada uma das mais de 700 companhias que atende, códigos diferentes na determinação de sequências numéricas que representam cada palavra – e,

[61] "(...) *gay men and women tended to have gender-atypical facial morphology, expression and grooming styles*" (WANG, KOSINSKI, 2017)

consequentemente, no estabelecimento das respectivas correlações com outras com o mesmo significado, quando aplicadas em determinado contexto. Os mesmos rótulos serão aplicados mundialmente.

Há, de fato, um certo grau de flexibilização para cada companhia. Quais critérios serão ou não determinantes, quais "talentos" da companhia serão utilizados como padrões para informar decisões (i.e., qual a base de dados será utilizada como grupo teste para informar como as máquinas devem escolher): tudo isso pode ser determinado pela companhia que contrata o serviço da HireVue. No entanto, tudo isso também depende de decisões humanas, e podem dar origens a vieses de dados, ou vieses decorrentes, por não considerar a mudança de valores da sociedade como um todo (que acredita que mulheres não devem mais desempenhar papéis domésticos), mas somente os valores dos tomadores de decisão das companhias (em sua maioria, homens, por conta da estratificação vertical). Além disso, como vimos, todos esses critérios personalizáveis pela companhia serão marcados por noções já sexistas, levando em consideração o histórico das empresas (que até pouco tempo, não tinham qualquer pretensão de contratar mais mulheres).

Conforme matéria divulgada no *Business Insider*, realizada após Richard Feloni ter testado a tecnologia da HireVue[62], a maior parte das companhias que emprega os serviços d HireVue utiliza entrevistas preditivas em vídeo como uma das primeiras etapas de seleção para cargos de estágio inicial. Ou seja, o corte inicial é feito por máquinas, para cargos de menor relevância. Como vimos, o final do processo depende de aprovação humana: o emprego da tecnologia pode ser encarado como uma justificativa e base para escolhas posteriores enviesadas.

Assim, em um próximo ciclo, as decisões de contratação em processos que já empregaram a tecnologia da HireVue (dotadas de vieses reproduzidos por inteligência artificial, e por vieses humanos - que decidem a contratação final) serão utilizadas para orientar novas decisões automatizadas. As possibilidades de que um resultado discriminatório seja apresentado contra mulheres são reforçadas. Isso porque mulheres entrevistadas digitalmente e reprovadas na fase presencial serão usadas como parâmetro negativo para alimentar o sistema, e apenas aquelas que tenham atendido aos critérios

[62] A oportunidade de testar a companhia foi possível a partir do contato de Richard Feloni com o CTO da HireVue.

(possivelmente) discriminatórios da máquina e tenham burlado o crivo igualmente sexista de seres humanos enviesados serão utilizadas como casos de sucesso.

A própria Unilever, apresentada como caso de estudo de promoção de diversidade a partir do uso da tecnologia[63], emprega entrevistas preditivas em vídeo apenas para cargos de entrada (no caso, representantes de atendimento ao cliente). Logo, a estratificação vertical pode ser mantida e, se utilizarem futuras apenas os dados dos candidatos que permaneceram na empresa por mais tempo para orientar decisões futuras, ou informações daqueles empregados que receberam qualquer tipo de promoção, decisões automatizadas podem, em um futuro próximo, voltar a aumentar a desigualdade. Não por outro motivo, a HireVue não especifica em quais cargos ocorreu o aumento de 16% em diversidade, nem quais os indicadores de aferição de diversidade.

Considerando que a Unilever utiliza o software sobretudo em contratação para cargos de atendimento ao cliente, é de fato provável que o emprego da tecnologia promova o aumento de mulheres na companhia. No entanto, o aumento da diversidade é acompanhado da segregação horizontal: a função de atendimento reflete o papel social da mulher, e se o aumento foi verificado apenas nessa classe, não há diversidade sendo promovida, e sim o reforço de representações estereotipadas de gênero. A empresa, provavelmente, não irá alterar todo o código da tecnologia caso perceba que discriminações desse tipo se deram contra um grupo específico de pessoas no emprego das entrevistas preditivas em vídeo para um processo de contratação específico.

Para agravar, o funcionamento dessa tecnologia, desenvolvida em 2012[64] pela HireVue, é incerto. Como verificamos, dada a opacidade dos parâmetros de decisão, tudo o que é sabido acerca do uso de inteligência artificial em entrevistas de vídeos preditivas é o que consta na divulgação dos *white papers* da própria companhia (dispersos em diversos documentos diferentes, o que dificulta ainda mais a verificação[65]), ou de informações divulgadas à imprensa através de seus representantes, as quais não são uníssonas. Por exemplo, enquanto em informações oferecidas ao *The Daily Telegraph*, a companhia alega que linguagem e

[63]HIREVUE. Unilever Success Story. disponível em <https://www.hirevue.com/wp-content/uploads/2019/02/Unilever-Success-Story-PDF.pdf>

[64] HIREVUE. *Artificial Intelligence, Hiring Assessments, Video Interviewing.* disponível em https://www.hirevue.com/blog/reducing-bias-and-widening-the-candidate-pool-why-we-built-hirevue-assessments>

[65]HIREVUE. *2* ways to take video interviewing to the next level. disponível em <https://www.hirevue.com/wp-content/uploads/2019/04/2-Ways-to-Take-Video-Interviewing-to-the-Next-Level-HireVue-eBook.pdf>

escolha de palavras representam mais de 80% da avaliação do candidato, em entrevista concedida para *Fast Company*, o CTO alega que 30% da nota é dada tendo como referência a análise de expressões faciais. Os números não fecham, e as informações são conflitantes.

A opacidade intencional (por meio do emprego tanto do sigilo real e legal, como da ofuscação, verificada no fornecimento de informações conflitantes), característica de algoritmos discriminatórios, é identificada no serviço prestado pela HireVue. Indivíduos eventualmente injustiçados pelas decisões orientadas pelas predições apresentadas pelos algoritmos não sabem os motivos pelos quais receberam um resultado negativo. E mesmo que venham a saber, não possuem os meios adequados para compreender de que forma aquele resultado injusto se deu.

Conclusão

Estamos entrando em uma nova fase de discriminação de gênero, em que representações estereotipadas são reproduzidas por sistemas de inteligência artificial, e ainda não sabemos como lidar com essas novas formas de promoção de preconceitos. Sabemos quais foram os impactos negativos da consolidação de uma divisão sexual do trabalho, pautado na imagem social de que a mulher deve satisfazer as necessidades do homem, protagonista do espaço privado, ao longo do período histórico em que somente o espaço privado era feminino. Não obstante os efeitos da replicação em escala desses vieses serem desconhecidos, diversas companhias empregam tecnologias de decisão automatizada em seus processos de seleção – e muitas já foram consideradas discriminatórias.

Dados históricos são usados para moldar os parâmetros considerados desejáveis para cada contratação. Mesmo quando as informações que indicam gênero são neutralizadas, atributos considerados femininos podem etiquetados como indesejados para determinados cargos (e desejados para cargos que reforçam estereótipos de gênero). Um dos motivos que pode levar a essa forma de viés é a ausência de diversidade nos dados oferecidos para orientar as decisões - já que, historicamente, decisões de contratação eram realizada por humanos orientados pela divisão sexual do trabalho.

Há casos em que, excepcionalmente, companhias atuam com cuidado na seleção do grupo teste, para que um conjunto de pessoas diversas entre si seja utilizado para orientar decisões automatizadas. Ainda assim, é possível que esses padrões sejam reproduzidos pela própria noção cristalizada - na fala, nas instituições, nas relações pessoas – de atributos considerados por humanos de valores/características próprias de cada profissão – os vieses pré-existentes, sociais ou individuais. Não é que o desenvolvedor acredite que mulheres sejam mais desejáveis como secretárias, mas sim que as características socialmente atribuídas às mulheres não são almejadas para ocupar um cargo de programador, por exemplo. Assim, mulheres e homens acabam reproduzindo padrões de gênero determinados pela sociedade - para não serem penalizados por ela.

Há ainda que relembrar de que a decisão de quem irá ocupar os principais cargos de liderança (estratificação vertical) não é feita por máquinas, mas sim por pessoas, que possuem valores construídos sob a ótica machista. Como a matemática Cathy O' Neil alerta, a intensidade dos impactos da aplicação não-testada e não-questionada de modelos matemáticos é maior nas massas pobres e/ou oprimidas. Para a autora, premissas perversas

encobertas por expressões matemáticas (inteligíveis por grande parte da sociedade) não afetam os privilegiados: diferente do que ocorre com as massas historicamente reprimidas, eles não são processados por máquinas.

Assim, caso sejam utilizados os casos de sucesso como referenciais para decisões automatizadas (i.e., pessoas que tenham se saído bem em entrevista e alcançaram cargos altos na empresa), os parâmetros sempre serão masculinos. Isso ocorre pois, de um lado, as decisões de promoção ainda são tomadas por humanos e, em sua maioria, homens, de outro, processamento da massa que não se encontra no topo hierárquico (de onde mulheres são excluídas) é feito por máquinas que utilizam os parâmetros de sucesso – majoritariamente, masculinos.

Dada a natureza enraizada dos vieses, e as formas pelas quais eles são reproduzidos, dificilmente processos de revisão dos modelos matemáticos levará à mitigação de padrões discriminatórios[66]. A maior parte desses sistemas, como vimos, é desenvolvido para penalizar pessoas que não possuam determinados atributos, e não para identificar possíveis talentos e oferecer ajudar. Caso a promoção de diversidade fosse o verdadeiro objetivo dos sistemas automatizados, atributos de gênero não seriam, necessariamente, neutralizados, mas levados em consideração para que as empresas possam fornecer os recursos necessários para reparar a sub-representação que leva à perpetuação de discriminação por meio de inteligência artificial. Encaro a baixa representatividade feminina como uma importante barreira para o uso de modelos automatizados de seleção de pessoas, seja pela baixa presença de mulheres no desenvolvimento de algoritmos, que possam adiantar os cenários de discriminação, pela escassez de dados femininos, ou pela falta de liderança feminina, que possa impulsionar outras mulheres, a fim de subverter os papéis consolidados de gênero.

Sem esse primeiro passo em direção à maior representatividade, o uso de tecnologias de inteligência artificial não irá deixar de representar, ao meu ver, mais uma forma de discriminação feminina, que reforça o valor do atributo gênero como fator discriminatório no ambiente de trabalho. Conforme os dados do IBGE apontam (cf. nota de rodapé 31), vimos o discurso que mulheres somente são aptas a ocupar cargos cujas tarefas sejam compatíveis com seu papel social parece estar enfraquecendo no imaginário social. Além disso, posturas

---

[66] *"Phrenology was a model that relied on pseudoscientific non-sense to make authorative pronouncements, and for decades it went untested. Big Data can fall into the same trap. Models like the ones that red-lighted Kyle Behm and blackballed foreign medical students at St. George's can lock people out, even when the Science inside them is little more than a bundle of untested assumptions.*" (O' NEIL, 2016, p. 122)

claramente discriminatórias adotadas por companhias podem prejudicar a imagem das mesmas. No entanto, dada a natureza de tecnologias de inteligência artificial, o seu emprego em processos de seleção pode voltar a consolidar papeis sociais marcados por estereótipos de gênero. Essas instruções algorítmicas enviesadas dificilmente são denunciadas e posteriormente combatidas, tendo em vista a opacidade e inteligibilidade dos programas, que impedem a comprovação da discriminação. Nesses cenários, avanços tecnológicos representam retrocessos sociais – e nenhum benefício de otimização de tempo e quantidade de candidatos avaliados poderia servir como justificativa para manutenção do emprego dessa tecnologia.

No presente trabalho, aventei os possíveis prejuízos do emprego indiscriminado de tecnologias de inteligência artificial em processos de tomadas de decisão que envolvam aspectos sensíveis da vida de uma mulher, sem pretensão de fornecer soluções normativas para garantir o uso justo dessas tecnologias. Após verificar o potencial negativo para mulheres, entretanto, concluo que ainda não alcançamos um grau de neutralidade na sociedade a ponto de garantir que o uso de tecnologias possa promover maior diversidade de gênero no ambiente do trabalho. Pelo contrário, o alcance de seu funcionamento é tal que as representações estereotipadas possam ser retomadas e reforçadas. São diversas as origens de vieses e ética digital ainda não é tímida no combate de incorporações de discriminação.

Referências

ANGWIN, Julia. LARSON, Jeff. MATTU, Surya. KIRCHNER, Lauren. MACHINE BIAS. *There's software used across the country to predict future criminals. And it's biased against blacks*. Pro Publica, 2016. disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> acesso em: 20/10/2019

ANGWIN, Julia. LARSON, Jeff. TERRIS, Terry. *How Machines Learn to be Racist.* disponível em: <https://www.propublica.org/article/breaking-the-black-box-how-machines-learn-to-be-racist?word=Trump> acesso em junho/2019

BIROLI, Flavia. MIGUEL, Luis Felipe. Feminismo e Política: Uma introdução. São Paulo. Boitempo: 2014.

BOLUKBASI, Told. et al. *Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings*. 30th Conference on Neural Information Processing Systems. Barcelona, Spain, 2016.

BUSINESS INSIDER. *I tried the software that uses AI to scan job applicants for companies like Goldman Sachs and Unilever before meeting them - and it's not as creepy as it sound*. 29/08/2019 disponível em <https://www.businessinsider.com/hirevue-ai-powered-job-interview-platform-2017-> acesso em 03/11/2019

BUOLAMWINI, Joy. GEBRU, Timnit. *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification.* in.: Proceedings of Machine Learning Researc*h.* 2018

COX, Michael, ELLSWORTH, David. *Application-Controlled Demand Paging for Out-of-Core Visualization.* Phoenix, 1997. disponível em <https://www.nas.nasa.gov/assets/pdf/techreports/1997/nas-97-010.pdf>

DASTIN, Jeffrey. Amazon scraps secret AI recruiting tool that showed bias against women. disponível em: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G> , acesso em 03/06/2019.

DESAI, Deven. KROLL, Joshua. *Trust but verify: a guide to algorithms and the law.* In. **Harvard Journal of Law and Technology**. Vol, 31, n 1. Harvard Univertiy, 2017.

FAST COMPANY. *This Bot judges how much you smile during your job: Predictive AI applications like HireVue screen job candidates on around 25,000 data points, breaking down their words, voice, and facial expressions*. 15/01/2019 disponível em: <https://www.fastcompany.com/90284772/this-bot-judges-how-much-you-smile-during-your-job-interview> acesso em 10/11/2019

FEAST, Josh. *4 ways to address Gender Bias in AI.* In: **Harvard Business Review**, 2019. disponível em: <https://hbr.org/2019/11/4-ways-to-address-gender-bias-in-ai> aceso em 20/11/2019

FOLHA DE SÃO PAULO. *151 pessoas são presas por reconhecimento facial no país; 90% são negras. Sem dados, pesquisadores fizeram levantamento em cinco estados brasileiros. 22/11/2019* disponível em <https://www1.folha.uol.com.br/cotidiano/2019/11/151-pessoas-sao-presas-por-reconhecimento-facial-no-pais-90-sao-negras.shtml> acesso em 22/11/2019

FROEHNER, Bruna e SPAGNUOLO, Sérgio. Mulheres avançam em ritmo lento ao topo do mundo corporativo. Gênero e Número, 2017. disponível em: <http://www.generonumero.media/mulheres-avancam-comando-mundo-corporativo> acesso em 15/11/2019.

HAO, Karen; STRAY, Jonathan. Can You Make AI fairer than a judge? Play our courtroom algorithm game. The US criminal legal system uses predictive algorithms to try to make judicial process less biased. But there's a deeper problem. In.: MIT Technology Review. 2019. disponível em: <https://www.technologyreview.com/s/613508/ai-fairer-than-judge-criminal-risk-assessment-algorithm/?utm_medium=tr_social&utm_source=instagram&utm_campaign=site_visitor.unpaid.engagement> acesso em: 20/10/2019.

HENTON, Caroline. *Where* is Female Synthetic Speech? In.: Journal of the International Phonetic Association. Cambridge University Press, 1999. disponível em <https://www.jstor.org/stable/44526232?read-now=1&seq=2#page_scan_tab_contents>

HIREVUE. Unilever's Recruiting Process disponível em : <https://www.hirevue.com/resources/unilevers-recruiting-process>

HIREVUE. *2* ways to take video interviewing to the next level. disponível em <https://www.hirevue.com/wp-content/uploads/2019/04/2-Ways-to-Take-Video-Interviewing-to-the-Next-Level-HireVue-eBook.pdf>

HIREVUE. Unilever Success Story. disponível em <https://www.hirevue.com/wp-content/uploads/2019/02/Unilever-Success-Story-PDF.pdf>

HUFF, Charles; COOPER, Joel. *Sex Bias in Educational Software: the effect of designers` stereotype on the software they design*. In **Journal of Applied Social Psychology**, 1987, p. 519-532.

GLOBO. Algoritmos reproduzem machismo e racismo por se basearem em práticas discriminatórias humanas. Novembro, 2019 disponível em: <https://oglobo.globo.com/economia/algoritmos-reproduzem-machismo-racismo-por-se-basearem-em-praticas-discriminatorias-dos-humanos-24085081> acesso em: 17/11/2019

IBGE. Estatística de Gênero: Indicadores sociais das mulheres no Brasil. In.: Estudos e Pesquisa - Informação Demográfica e Socioeconômica, edição 38. IBGE, 2018. disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv101551_informativo.pdf>

LAGNE, Lynda, et al. *Feminist Interpretations of Jean-Jacques Rousseau*. The Pennsylvania State University Press, 2002.

LARSEN, Loren. Reducing Bias and Widening the candidate pool: why built HireVue assessments. HIREVUE, 2018. disponível em: <https://www.hirevue.com/blog/reducing-bias-and-widening-the-candidate-pool-why-we-built-hirevue-assessments> acesso em 25/11/2019

LARSEN, Loren. *Train, validate, re-train: how we build HireVue assessments models*. HIREVUE, 2018. disponível em: <https://www.hirevue.com/blog/train-validate-re-train-how-we-build-hirevue-assessments-models> acesso em 25/11/2019

LEANIN, e MCKINSEY AND COMPANY. Women in the workplace, 2019. disponível para download em : <https://womenintheworkplace.com/>

MATOS, Helena Ferreira. *O viés em Machine Learning: Perspectivas Regulatórias*. In: **Horizonte Presente Tecnologia e Sociedade em Debate**. Belo Horizonte: Casa do DIreito; FGV - Fundação Getúlio Vargas, 2019 (p. 569)

NOBLE, Safyia. *Algorithms of oppression: How search engines reinforce racism*. New York, NY, US: New York University Press, 2018

O'NEIL, Cathy. *Weapons of Math Destruction: How Big Data increases Inequality and Threatens Democracy*. Nova York: Crown, 2018.

OXFAM BRASIL, DATAFOLHA. *NÓS E AS DESIGUALDADES: Percepções Sobre Desigualdades no Brasil*. 2019. disponível em: <https://oxfam.org.br/um-retrato-das-desigualdades-brasileiras/pesquisa-nos-e-as-desigualdades/pesquisa-nos-e-as-desigualdades-2019>

PASQUALE, Frank. *The Black Box Society: The secret Algorithms that control Money and Information*. Cambridge, Massachusetts, London, England: Harvard University Press, 2015.

SOSNICK, Michael. *Exploring Fairness and Bias in Algorithms and Word Embedding*. School of Engineering and Applied Science, University of Pennsylvania, dec/2017.

THE DAILY TELEGRAPH. *AI used for first time in job interviews in UK to find best applicants*. 27/09/2019. disponível em: <https://www.telegraph.co.uk/news/2019/09/27/ai-facial-recognition-used-first-time-job-interviews-uk-find/> acesso em 05/10/2019

UOL NOTÍCIAS. Mulheres digitais Porque todas as assistentes virtuais têm vozes femininas?. disponível em: <https://www.uol/noticias/especiais/assistentes-de-voz-x-feminismo.htm#tematico-1 >

WANG, Yilun. KOSINSKI, Michal. *Deep neural networks are more accurate than humans at detecting sexual orientation from facial 10 images.* American Psychological Association, 2017.

WORLD ECONOMIC FORUM. *The Global Gender Report*. 2018. disponível em: <http://www3.weforum.org/docs/WEF_GGGR_2018.pdf> acesso em 1 0/11/2019.