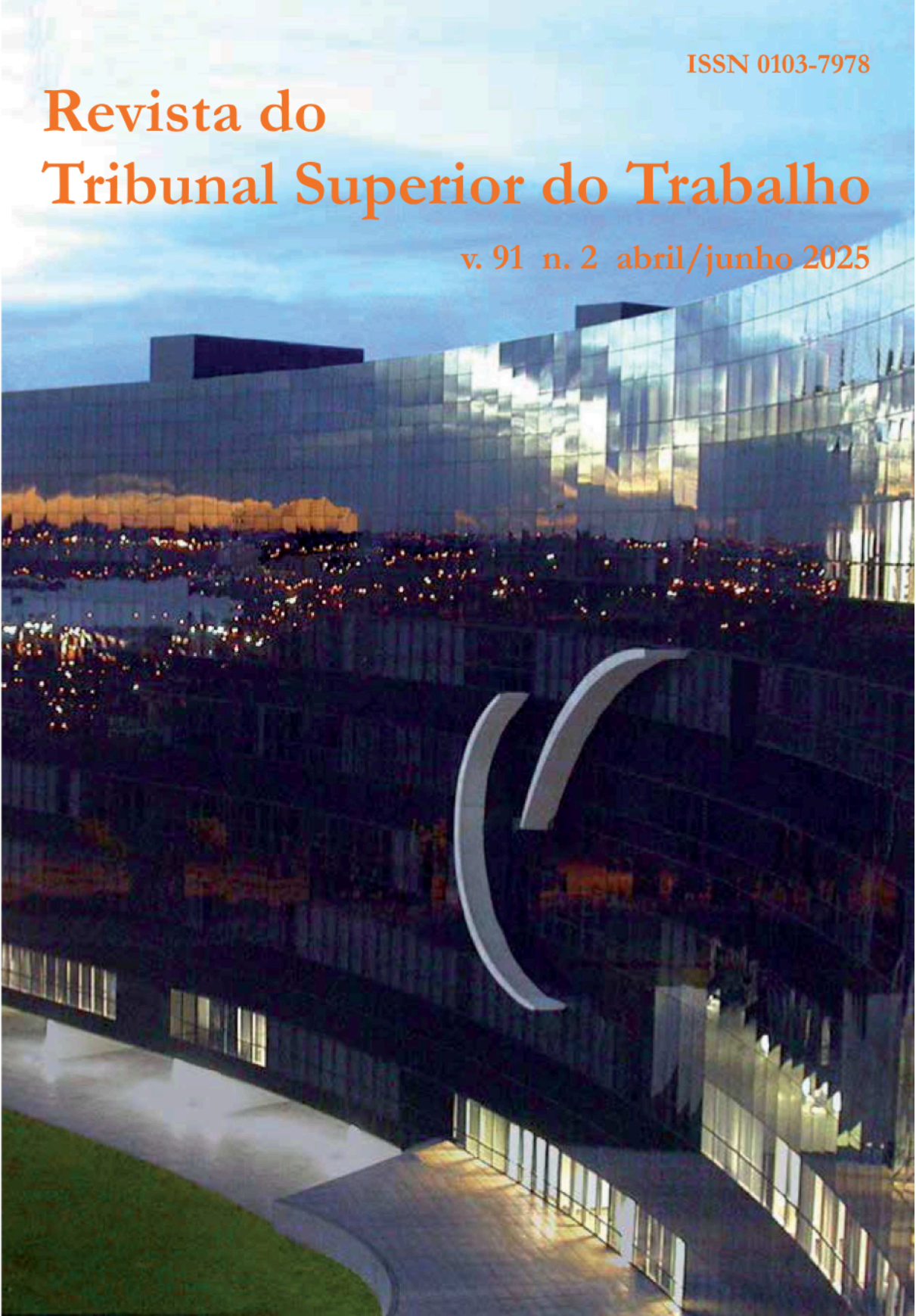
ISSN 0103-7978

Revista do Tribunal Superior do Trabalho

v. 91 n. 2 abril/junho 2025



Revista do Tribunal Superior do Trabalho

QUALIS B2

PODER JUDICIÁRIO

JUSTIÇA DO TRABALHO

TRIBUNAL SUPERIOR DO TRABALHO

Revista do Tribunal Superior do Trabalho

QUALIS B2

v. 91 n. 2 abril/junho 2025

Rua Dezoito de Novembro, 423 – Conj. 203 – CEP 90240-040 – Porto Alegre-RS
comercial@lex.com.br – www.lex.com.br

ISSN 0103-7978

Revista do Tribunal Superior do Trabalho / Tribunal Superior do Trabalho. – Vol. 21, n. 1 (set./dez. 1946) – Rio de Janeiro : Imprensa Nacional, 1947-.
v.
Trimestral.
Irregular, 1946-1968; suspensa, 1996-1998; trimestral, out. 1999-jun. 2002; semestral, jul. 2002-dez. 2004; quadrimestral, maio 2005-dez. 2006.
Continuação de: Revista do Conselho Nacional do Trabalho, 1925-1940 (maio/ago.).
Coordenada pelo: Serviço de Jurisprudência e Revista, 1977-1993; pela: Comissão de Documentação, 1994-.
Editores: 1946-1947, Imprensa Nacional; 1948-1974, Tribunal Superior do Trabalho; 1975-1995, LTr; out. 1999-mar. 2007, Síntese; abr. 2007- jun. 2010, Magister; jul. 2010- , Lex.
ISSN 0103-7978

1. Direito do Trabalho. 2. Processo Trabalhista. 3. Justiça do Trabalho – Brasil. 4. Jurisprudência Trabalhista – Brasil. I. Brasil. Tribunal Superior do Trabalho.

CDU 347.998.72(81)(05)

Equipe Editorial e Científica: Comissão de Documentação e Memória – Presidente: Ministra Maria Cristina Irigoyen Peduzzi; Membros: Ministro Alexandre Luiz Ramos; Ministro Sergio Pinto Martins; Ministra Liana Chaib (Suplente)

Organização e Supervisão: Kassandra Trindade Clatworthy – Coordenadora de Documentação

Revisão em língua portuguesa: José Geraldo Pereira Baião – Coordenadoria de Documentação

Capa: Ivan Salles de Rezende (sobre foto de Marta Crisóstomo)

Editoração Eletrônica: Lex Editora S/A

Tiragem: 700 exemplares

Os artigos publicados nesta Revista não traduzem necessariamente a opinião institucional do Tribunal Superior do Trabalho. A publicação dos textos obedece ao propósito de estimular o debate sobre questões jurídicas relevantes para sociedade brasileira e de refletir as várias tendências do pensamento jurídico contemporâneo. A avaliação de artigos ocorre de maneira contínua durante todo o ano. Os textos devem ser enviados em formato word para o seguinte endereço: revista@tst.jus.br.

Tribunal Superior do Trabalho
Setor de Administração Federal Sul
Quadra 8, lote 1, bloco "B", mezanino
70070-600 – Brasília – DF
Fone: (61) 3043-3056
E-mail: *revista@tst.jus.br*
Internet: *www.tst.jus.br*

Lex Editora S.A.
Rua Dezoito de Novembro, 423 – Conj. 203
90240-040 – Porto Alegre-RS
Fone: (51) 3191-3033
Assinaturas:
comercial@lex.com.br
www.lex.com.br

Composição do Tribunal Superior do Trabalho

Sumário

DISCRIMINAÇÃO ALGORÍTMICA NAS RELAÇÕES DE TRABALHO

ALGORITHMIC DISCRIMINATION IN LABOR RELATIONS

Sergio Torres Teixeira[1]
Alexandre Freire Pimentel[2]
Camila Crasto Pugliesi[3]

RESUMO: Com o advento da Quarta Revolução Industrial, as tecnologias do *Big Data*, do *Machine Learning* e da Inteligência Artificial passam a ser utilizadas de maneira combinada permitindo que algoritmos estejam em constante reanálise de padrões de interesses com base em um banco de dados previamente estabelecido pelo programador do código-fonte. Entretanto, os padrões algorítmicos podem se revelar enviesados e capazes de produzir uma nova espécie de dano na dinâmica laboral: a "discriminação algorítmica". O presente trabalho foi realizado por meio de uma pesquisa quantitativo-qualitativa, investigando o fenômeno em questão, e fundamentando por meio de casos concretos trazidos em artigos científicos e doutrinas que tratam do assunto. Foi realizada a análise dos dados trazidos das pesquisas, a respeito do tema, por meio de técnica dedutiva, ressaltando os principais pontos estudados, de acordo com os resultados obtidos.

PALAVRAS-CHAVE: algoritmos; discriminação; plataformas digitais; quarta revolução industrial.

ABSTRACT: With the advent of the Fourth Industrial Revolution, Big Data, machine learning and Artificial Intelligence technologies are now used in combination, allowing algorithms to constantly re-analyze patterns of interest based on a database previously established by the programmer of the source code. However, algorithmic standards may prove to be envied and capable of producing a new kind of damage in labor dynamics: "algorithmic discrimination". This study was carried out through a quantitative-qualitative research, investigating the phenomenon in question, and substantiating through concrete cases brought up in papers and doctrines dealing with the subject. The data gathered from the research on the subject was analyzed using a deductive technique, highlighting the main points studied according to the results obtained.

KEYWORDS: algorithms; discrimination; digital platforms; fourth industrial revolution.

---

1 *Doutor em Direito; professor titular da Unicap; professor associado da FDR/UFPE; desembargador do TRT6; pesquisador-líder do Grupo Logos: Processo, Hermenêutica e Tecnologia.*

2 *Doutor em Direito; professor adjunto da Unicap; professor adjunto da FDR/UFPE; desembargador do TJPE; pesquisador-líder do Grupo Logos: Processo, Hermenêutica e Tecnologia.*

3 *Mestranda em Direito pela Universidade de Lisboa; pesquisadora do Grupo Logos: Processo, Hermenêutica e Tecnologia; advogada.*

Recebido em: 28/04/2025
Aprovado em: 05/05/2025

1 Introdução

A tecnologia tem se tornado, a cada dia mais, inerente ao desenvolvimento das atividades do ser humano. No direito, houve a ruptura paradigmática instaurada com a adoção do processo eletrônico.

Um exemplo prático foram as milhares de pilhas de "papéis" que passaram a ser substituídas pelo armazenamento de documentos e informações de modo digital, o que, inegavelmente, trouxe ganhos incalculáveis, facilitando a vida dos que precisavam manejar os processos físicos, se tornando uma tarefa muito mais prática.

Dessa forma, percebe-se que a Inteligência Artificial, apesar de parecer para alguns apenas ficção científica, já faz parte do cotidiano dos indivíduos, ainda que desconheçam, por exemplo, das escolhas que a Netflix faz para os usuários, das sugestões nas redes sociais de produtos que muitas vezes nem foram pesquisados, das ferramentas de busca, etc.

As alterações advindas da tecnologia da chamada revolução digital 4.0 atingiram diretamente o modo de viver dos indivíduos.

Atingida, também, foi a área trabalhista. É sabido que programas de *software* de inteligência artificial são a cada dia mais utilizados para otimizar o gerenciamento da atividade dos trabalhadores, desde o processo seletivo até o aferimento de produtividade.

Na esfera trabalhista, algumas alterações que podem ser citadas são: o aumento da utilização do teletrabalho; a existência de empresas digitais, sem sede física, que admitem e assalariam trabalhadores; e a adoção de programas de inteligência artificial e de aplicativos para dirimir as relações laborais.

Entretanto, a criação de ferramentas tecnológicas não deve ser feita de maneira desenfreada, tampouco deve dispensar a avaliação humana. O uso de algoritmos na admissão e na avaliação dos trabalhadores pode gerar situações discriminatórias.

Para frear esses tipos de situações, as empresas devem se atentar aos princípios trabalhistas no momento de programar os *softwares* de inteligência artificial e, sobretudo, agir com transparência algorítmica.

Sendo assim, o presente trabalho visa a refletir sobre a transparência algorítmica como um direito do trabalhador, contextualizando-o com os direitos à intimidade, à privacidade, à autodeterminação informativa e à não discriminação nas relações laborais.

Dessa maneira, este trabalho se utilizará do método de pesquisa quantitativo-qualitativa, investigando o fenômeno em questão, e fundamentando através de casos concretos trazidos por meio de artigos científicos e doutrinas

que tratam do assunto, uma vez que houve o enfoque principalmente nas pesquisas bibliográfica e documental sobre a questão da inteligência artificial na aplicabilidade do Direito do Trabalho, bem como casos no qual essa utilização gerou o que tratamos ao longo do trabalho de discriminação algorítmica, se utilizando de livros, artigos e reportagens.

Ademais, com o intuito de comprovar a legitimidade das alegações apresentadas neste trabalho, serão utilizados pontos de vista de ilustres doutrinadores, principalmente do direito estrangeiro, que contribuem ou contribuíram para o entendimento claro e harmônico para a compreensão do Direito Digital e do Trabalho e a ligação entre ambos os temas.

2 Inteligência artificial, ciberespaço e uso dos algoritmos

Devido a sua inserção em todas as esferas da atividade humana, a revolução da tecnologia da informação tem relação direta com a complexidade da nova economia, sociedade e cultura em formação[4].

Nossa espécie, *homo sapiens* (homem sábio), foi denominada pelo fato de nos diferenciarmos das demais espécies do reino animal pela presença de autoconsciência, racionalidade e sapiência. Desde os primórdios, procuramos entender como pensamos e agimos.

A inteligência artificial (IA) vai além do que nós, *homo sapiens*, somos capazes de fazer: ela tenta não apenas compreender, mas também construir entidades inteligentes.

A IA é um dos campos mais recentes nas ciências e na engenharia. Sua história começou logo após a Segunda Guerra Mundial, e o próprio nome foi denominado em 1956. Hoje, a IA abrange uma enorme variedade de subcampos, do geral (aprendizagem e percepção) até tarefas específicas, como jogos de xadrez, demonstração de teoremas matemáticos, criação de poesia, direção de um carro em estrada movimentada e diagnóstico de doenças.

Ainda no contexto da IA, como meio disruptivo que culminou em grandes mudanças na vida das pessoas, a cibercultura está intimamente ligada a ela e se define como a cultura atual, sendo influenciada pelas tecnologias digitais e seu impacto na vida civil.

Ela surgiu a partir do desenvolvimento da internet e da tecnologia digital e, portanto, do uso exagerado da rede de computadores e de todos os meios tecnológicos que dão suporte a este tipo de conexão.

4 CASTELLS, Manuel. *A sociedade em rede*. A era da informação: sociedade, economia e cultura. 6. ed. Rio de Janeiro: CIP-Brasil, 1999. v. 1.

Esses meios, hoje, são cruciais para a comunicação, além do desenvolvimento da indústria do entretenimento e do comércio eletrônico. Dessa maneira, essa forma de cultura, em resumo, nada mais é que uma grande ligação, disseminação e interação entre, praticamente, todas as formas de cultura existentes em todo o mundo.

A cibercultura, por sua vez, é naturalmente uma forma de cultura surgida junto com o desenvolvimento das tecnologias digitais. Ela surgiu naturalmente junto com o desenvolvimento das tecnologias digitais, que estão ganhando cada vez mais espaço entre a sociedade moderna, levando a sua maior presença em todo o mundo, sendo nada mais que uma grande ligação, disseminação e interação entre, praticamente, todas as formas de cultura existentes em todo o mundo.

O ciberespaço, por seu turno, pode ser conceituado como o local virtual onde a cibercultura acontece por meio das manifestações e interações dos usuários de cada plataforma disponível. É um espaço de comunicação que não precisa da estrutura física, para que os seus usuários se comuniquem e, até mesmo, se relacionem.

Segundo Pierre Lévy, esse espaço e as interações que ali ocorrem "[...] conduzem diretamente à virtualização das organizações que, com a ajuda das ferramentas da cibercultura, tornam-se cada vez menos dependentes de lugares determinados, de horários de trabalho fixos e de planejamentos a longo prazo"[5].

Dessa forma, com o uso exacerbado do ciberespaço e com a realização de diversas necessidades básicas e essenciais dos indivíduos nesse ambiente, que eram feitas de forma física – como pagamentos, transações econômicas e financeiras, troca de mensagens, ligações, compras, aulas, reuniões, encontros entre outros, os quais foram substituídos por *home banking*, cartões inteligentes, voto eletrônico, *pages*, *palms*, imposto de renda via rede, inscrições via internet – o caráter virtual é levado em consideração, visto que passou a gerar uma grande importância na vida física das pessoas no contexto atual.

Assim, a extensão do ciberespaço acompanha e acelera uma virtualização não só da economia, mas da sociedade como um todo.

É comum pensar que os algoritmos surgiram após os grandes avanços da tecnologia, no fim do século passado, mas o conceito deles é antigo, da mesma forma que a matemática, sendo autônomo em relação à digitalização contemporânea, pois existe independentemente de qualquer computador, disco rígido ou outro substrato físico sobre o qual possa ser implementado.

---

5 LÉVY, Pierre. *Cibercultura.* São Paulo. Editora 34: 1999.

Eles são utilizados desde o início da civilização egípcia, quando as pessoas projetavam fórmulas para resolver problemas diários como a próxima enchente do rio Nilo, por exemplo.

Trata-se de uma operação projetada por uma sequência específica de etapas que são escritas para resolver um determinado problema ou para executar uma tarefa projetada automaticamente.

O algoritmo se conceitua como o átomo de cada processo de computação e objetiva mediar as atividades humanas a fim de diminuir a quantidade de procedimentos repetitivos ou exaustivos que agora realizamos indissociavelmente por meio de algoritmos, como uma pesquisa no Google ou a busca de uma rota no GPS, por exemplo.

A sua elevada implementação nas mais diversas atividades cotidianas da atualidade deve-se à somatória de três fatores principais:

Primeiramente, o aumento da capacidade de processamento dos computadores, o qual acelerou a velocidade da execução de tarefas complexas: vivemos o aumento contínuo da capacidade de processamento e a queda dos preços do *hardware*, o que permite uma interação cada vez maior e mais rápida entre os dispositivos e a informação disponível em rede[6].

Em seguida, a chegada do *Big Data* nos sistemas de informática. O armazenamento barato de quantidades gigantescas de dados deu aos algoritmos a possibilidade de identificar padrões imperceptíveis ao olho humano.

Reúnem-se e transformam-se, assim, dados isolados em aspectos do mundo nunca quantificados: trata-se de uma tecnologia fundamentalmente interconectada, cujo valor é extraído pelos padrões que podem ser endereçados em conexões entre informações caracterizadas por volume, variedade e velocidade.

Em terceiro lugar, o aprendizado de máquina – enquanto modalidade de IA – permite que algoritmos sejam criados e modificados por eles mesmos, representando um vínculo autonutritivo e duradouro entre as máquinas (dispositivos eletrônicos), os humanos e o *software*.

Os resultados fornecidos pelos algoritmos são chamados de *outputs*, enquanto os *inputs* que possibilitam as operações são os dados de entrada. Assim como uma fórmula matemática recebe valores numéricos para realizar o cálculo, um algoritmo recebe dados com o objetivo de processá-los e obter um resultado, o *output* desejado[7].

---

6 SCHWAB, Klaus. *A quarta revolução industrial*. São Paulo: Edipro, 2019.

7 DA ROCHA, Lorena *et al.* Discriminação algorítmica no trabalho digital. *Rev. Dir. Hum. Desenv. Social*.

Desta maneira, quanto maior a disponibilidade de conjuntos de dados e mais aprimorada a tecnologia de aprendizado de máquina, maior é o poder dado aos algoritmos para mediar nossa experiência com o mundo ao nosso redor, e mais capazes se tornam de substituir humanos na tomada de decisões.

Algoritmos inteligentes atuam não apenas para melhorar processos automatizados e maximizar estratégias comerciais, mas também para criar outras formas subjetivas de interação que envolvem análises avaliativas complexas de perfis – e, portanto, de trabalhadores – avaliação de características, personalidade, inclinações e propensões de uma pessoa, conforme sua orientação sexual, estados emocionais, opiniões políticas e pessoais, sua capacidade e habilidade para empregos ou funções específicas, entre outros aspectos sob as lentes do *Big Data*.

O aprendizado de máquina é projetado para se apoiar no constante aprimoramento do algoritmo e na conseguinte ininterrupta redefinição dos seus parâmetros, de tal feita que o controle algorítmico sobre a ação humana não se limita mais às experiências ensinadas por meio dos conjuntos de dados de treinamento e rotinas analíticas pré-programadas.

Neste contexto, o conceito de Quarta Revolução Industrial foi dado em 2016 por Klaus Schwab, fundador do Fórum Econômico Mundial, na sua obra que trata especificamente desse tema: "A Quarta Revolução Industrial gera um mundo no qual os sistemas de fabricação virtuais e físicos cooperam entre si de uma maneira flexível a nível global"[8].

Em 2019, foi publicado o relatório *Diversity, Equity and Inclusion 4.0* do Fórum Econômico Mundial, o qual mostrou que o início da presente década chamou a atenção para os fatores que estão entre as seguintes mudanças que serão vistas na sociedade: a aceleração da Quarta Revolução Industrial, as mudanças do mercado de trabalho em relação à adaptabilidade ao trabalho digital e um apelo abrangente por maior inclusão, equidade e justiça social[9].

O mundo trabalhista tem mudanças da mesma forma como a humanidade as tem, se transformando à medida que o ser humano desenvolve novos desejos, novas necessidades e novas formas de comunicação.

Uma das mudanças que geram mais reflexos na esfera do trabalho se trata do impacto que a tecnologia tem gerado no labor. Fala-se em empregos que tendem a "acabar" devido à automatização de processos e à alta utilização da IA, como muitos profissionais se preocupam, de outro lado, outros trabalhadores

8 SCHWAB, *op. cit.*

9 WORLD ECONOMIC FORUM. *HR4.0*: shaping people strategies in the Fourth Industrial Revolution. Geneva: World Economic Forum, 2019.

são altamente dependentes das plataformas digitais no seu labor, podendo se falar em uma grande dependência tecnológica.

Ainda estamos nos primeiros anos da Quarta Revolução Industrial. Apesar disso, *influencers*, *tiktokers*, moderadores de conteúdo, *gamers*, *crowdworkers*, *youtubers*, *vloggers*, *podcasters*, *pinners*, *memers* e blogueiros já são algumas das categorias que têm o ciberespaço como seu verdadeiro meio ambiente de trabalho.

Assim como outras esferas da vida dos usuários dessas novas tecnologias, pode-se dizer que esse processo "disruptivo" irá sofrer diversas consequências no aspecto trabalhista, de tal forma que, no futuro, os setores da economia estarão dependentes do algoritmo.

As *Big Techs*, grandes empresas de tecnologia que predominam no mercado, não podem mais permanecer isentas de responsabilidade pelas consequências antiéticas da utilização de recursos como algoritmos na automação de processos internos.

Juan Gustavo Corvalán nos traz que "una inteligencia artificial bien 'entrenada', con acceso al flujo informativo, simplifica y facilita exponencialmente las actividades de una organización y puede obtener resultados que serían imposibles de lograr con los cerebros humanos"[10].

De fato, no mundo do trabalho, a inteligência artificial tem condições de simplificar e facilitar bastante o processo na dinâmica laboral; algoritmos sofisticados são a base da IA, que permitem o aprendizado rápido e contínuo através do processamento, coleta, análise e pesquisa de dados.

Por outro lado, existem muitas críticas quando se fala a respeito da alta implementação de algoritmos na dinâmica de trabalho da *gig economy*, a qual se conceitua como um arranjo alternativo de emprego, uma forma de trabalho em que as pessoas exercem uma atividade *freelancer* e recebem separadamente por cada projeto/serviço.

Sobre os algoritmos nas relações de trabalho, deve ser abordado o viés racista que esses códigos podem manifestar na experiência do usuário nas redes sociais virtuais, as quais atualmente servem para uma multiplicidade de finalidades e, principalmente, como locais de conexão de pessoas a prestadores de serviços, lojas e os mais diversos bens adquiríveis, caracterizando um factual mercado de trabalho.

---

10 CORVALÁN, Juan Gustavo. Inteligencia artificial: retos, desafíos y oportunidades – Prometea: la primera inteligencia artificial de Latinoamérica al servicio de la Justicia. *Revista de Investigações Constitucionais*, Curitiba, v. 5, n. 1, p. 295-316, jan./abr. 2018.

Estes trabalhadores são internautas que possuem um amplo "público" de seguidores e, portanto, são remunerados por marcas e empresas para recomendar suas iniciativas comerciais, como meio de divulgação mais lucrativo.

No entanto, recentemente, coletivos de blogueiras, influenciadoras, modelos e criadoras de conteúdo *online* negras passaram a trazer à tona práticas de *shadowbanning* e de disparidade salarial das quais se perceberam vítimas ao comparar seus ganhos financeiros com o histórico de contratos de suas colegas brancas.

A remuneração que essa categoria de trabalhadores digitais recebe por suas interações nas redes sociais virtuais é baseada em métricas de alcance e taxas de engajamento de seus perfis, as quais são definidas por algoritmos.

No contexto da Quarta Revolução Industrial e, especificamente, da *gig economy*, algoritmos podem ser implementados na dinâmica laboral desde a ocasião da contratação, durante todo o acompanhamento das atividades dos trabalhadores até a rescisão contratual e, mais ainda, nas novas modalidades de trabalho em que o vínculo de emprego ainda não é reconhecido, embora factualmente caracterizado, como ocorre com os trabalhadores criadores de conteúdo *online*.

Vivemos hoje em um contexto de Economia Digital, as categorias de trabalhadores digitais têm recebido atenção especial de organismos internacionais como a Organização Internacional do Trabalho (OIT), a Organização para a Cooperação e Desenvolvimento Econômico, a Conferência das Nações Unidas sobre Comércio e Desenvolvimento e a Comissão Europeia.

Em todos os relatórios publicados pelos referidos órgãos é reconhecido o impacto dos algoritmos não só na organização do trabalho em plataformas digitais, mas na economia global como um todo.

Essas novas tecnologias podem mudar processos de recrutamento e facilitar as operações de gerenciamento de desempenho: desde o rastreamento de localização, avaliações de usuários e criação de perfis, os algoritmos da *gig economy* são capazes de conectar empresas e trabalhadores e consumidores em plataformas, aplicativos e *sites*, seja para atuar puramente *online* no trabalho digital, em atividades ao vivo *off-line* ou em regimes híbridos.

Em relação às práticas de recrutamento, as empresas de tecnologia têm desenvolvido *softwares* que automatizam a análise de currículos, classificando candidatos, fazendo ofertas e determinando níveis salariais. Em um primeiro momento, o abandono dos processos seletivos tradicionais pelas *Big Techs* teve como finalidade diversificar o grupo de candidatos contratados.

O mesmo Juan Gustavo Corvalán já apontado, ao tratar da transparência, assinala que: "la inteligencia artificial debe ser transparente en sus decisiones, lo que significa que se pueda inferir o deducir una 'explicación entendible' acerca de los criterios en que se basa para arribar a una determinada conclusión, sugerencia o resultado"[11].

A partir do momento em que uma pessoa analisa os currículos de candidatos, ela está implicitamente tentando prever quais candidatos terão um bom desempenho e quais não e, neste processo, pode ser influenciada (in) ou conscientemente por tendências discriminatórias.

Na teoria, o uso de algoritmos levaria a resultados mais consistentes, não enviesados, ao mesmo tempo que isentaria as empresas de realizar entrevistas de seleção dispendiosas.

No entanto, autores como Juan Gustavo Corvalán afirmam que os vieses permaneceram presentes nos processos seletivos baseados em algoritmos, resultando em consequências antiéticas. De fato, a implantação da IA teve como objetivo minimizar discriminações na contratação. Na prática, a complicação trazida pelo artifício é a de que as pessoas cujas características não coincidem com as representadas no banco de dados de treinamento do algoritmo selecionador são tendencialmente excluídas dos *outputs* do processo seletivo.

Dessa forma, os candidatos não são classificados para os cargos, uma vez que a máquina entende que sua candidatura é menos atrativa. Um dos exemplos mais notórios dessa negligência algorítmica foi testemunhado no início de 2019, quando a IA da Amazon se revelou manifestamente sexista em seu processo seletivo automatizado.

O *software* havia sido criado em 2014 como meio de classificar currículos e selecionar automaticamente os candidatos mais talentosos; no entanto, o sistema foi treinado em um banco de dados apresentado pelos resultados dos proponentes contratados ao longo de dez anos, um grupo majoritariamente composto por homens.

Dessa forma, em função dos dados de treinamento não representativos, a tecnologia manifestou uma aprendizagem tendenciosa e não neutra em relação ao gênero: o algoritmo rapidamente aprendeu a favorecer candidatos do sexo masculino sobre os do sexo feminino, penalizando os currículos que incluíam a palavra "mulheres".

---

11 CORVALÁN, *op. cit.*

A "ferramenta sexista de IA da Amazon"[12] chama a atenção para as preocupações sobre o quão confiável e não discriminatória a implantação de algoritmos em processos seletivos pode ser.

À medida que o Vale do Silício desenvolve artifícios baseados no aprendizado de máquina, as empresas precisam considerar o risco de que os algoritmos podem ser enviesados, inserindo sexismo, racismo, homofobia, xenofobia ao reproduzir preconceitos arraigados em códigos, destacando implicitamente as disparidades da sociedade em seus processos internos.

Tal risco também ficou demonstrado quando, em uma experiência, candidatos a empregos com nomes que soam brancos, como Emily, receberam 50% mais ligações de retorno algoritmicamente selecionadas do que aqueles com nomes que soam afro-americanos, como Lakisha[13].

Foi constatado também que empresas, como a Amazon, utilizam algoritmos para controlar a produtividade dos empregados, cujos desligamentos são decididos por um *software* inteligente que "descarta" os trabalhadores mais "lentos" na execução de suas tarefas.

A média do tempo gasto pelos empregados é calculada a partir dos *scanners* pessoais que eles usam para a expedição dos produtos de suas prateleiras e esteiras. Todavia, entre os trabalhadores havia mulheres grávidas, cujo tempo de execução das tarefas é maior devido à sua condição e à maior frequência de utilização do banheiro, de modo que o algoritmo as classificou entre as mais ineficientes e as despediu, o que gerou ações trabalhistas por discriminação.

Dessa maneira, como exemplificam Kleinberg *et al.*, ao utilizar modelos de aprendizado de máquina para a contratação de um empregado, uma empresa busca um bom funcionário[14].

Mas o que seria um bom funcionário? Seria um funcionário que vende mais, que nunca se atrasa ou aquele cujos clientes mais frequentemente retornam à loja? Ou seria aquele mais qualificado? E quais seriam as qualificações que seriam mais preditivas para *performance* do empregado? São perguntas para as quais não há uma única reposta e que são permeadas pela subjetividade humana.

Os autores apontam que, por exemplo, a definição, feita pelo desenvolvedor, de que um bom funcionário é aquele que possui mais horas trabalhadas,

12 MATEUS, Kátia. *Como a Amazon foi atraiçoada pelo algoritmo sexista*. Disponível em: https://expresso.pt/economia/2018-12-09-Como-a-Amazon-foi-atraicoada-pelo-algoritmo-sexista. Acesso em: 15 abr. 2023.

13 CHANDER, A. The racist algorithm? *Michigan Law Review*, v.115, n. 6, p. 1022-1045, 2017.

14 KLEINBERG, Jon; LUDWIG, Jens; MULLAINATHAN, Sendhil; SUNSTEIN, Cass. Discrimination in the age of algorithms. *Journal of Legal Analysis*, v. 10, p. 113-174, 2018.

sem levar em conta que mulheres estatisticamente trabalham menos que homens em razão da licença-maternidade, pode levar a um impacto discriminatório.

Constatou-se também a discriminação em razão de gênero, em prejuízo das mulheres, no trabalho *on-demand* por meio de aplicativos. Em razão do acúmulo de tarefas domésticas e de cuidado com pessoas da família, as mulheres possuem em geral menos tempo para permanecer à disposição do aplicativo, inclusive em períodos noturnos, os quais, em razão da alta demanda e do chamado preço dinâmico, são mais bem remunerados.

Ademais, em tais períodos há maior risco à vida e à integridade física dos trabalhadores, em virtude, por exemplo, de possíveis roubos e assédio por parte dos passageiros e clientes, o que também afasta as mulheres.

O resultado é que estas acabam recebendo uma remuneração mais baixa e são preteridas pelo algoritmo, que lhes reserva corridas e tarefas menos lucrativas[15].

É igualmente preocupante, em particular, a implementação de *softwares* para promover algoritmos a assumirem o poder do empregador de monitorar os funcionários, sancionando-os e encerrando a relação de trabalho. Como o estudioso Prassl descreve em seu artigo *What if your boss is an algorithm?*, automatizar as decisões do empregador por meio de artifícios codificados é um ponto de virada revolucionário para o gerenciamento de recursos humanos baseado em dados[16].

A ascensão da análise algorítmica da rotina laboral é perfeitamente ilustrada na dinâmica típica da *gig economy*: a dos trabalhadores digitais que dependem de recomendação, revisão, reputação e mecanismos de classificação para gerenciar e avaliar sua força de trabalho.

O art. 6º, XI, da Lei Geral de Proteção de Dados dispõe que as atividades de tratamento de dados pessoais deverão observar a não discriminação, da impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos[17].

Vale ressaltar, neste sentido, que as estruturas de *compliance* empresariais não podem fechar os olhos para a necessidade de garantir a conformidade das decisões automatizadas com o ordenamento jurídico.

---

15 PUBLICA. A uberização do trabalho é pior pra elas. *Publica*, São Paulo, 28 maio 2019.

16 ADAMS-PRASSL, Jeremias. What if your boss was an algorithm? The Rise of Artificial Intelligence at Work, 2019. *Comparative Labor Law & Policy Journal*. Disponível em: https://ssrn.com/abstract=3661151. Acesso em: 10 abr. 2023.

17 BRASIL. *Lei Geral de Proteção de Dados*. Disponível em https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.

A atenção aos vieses discriminatórios embutidos nas decisões automatizadas é ainda mais urgente no contexto da *gig economy*, em que os algoritmos são determinantes da renda dos trabalhadores.

O trabalho por meio das plataformas digitais foi criado na segunda década do século XXI como uma modalidade de serviços caracterizada pela dependência dos meios eletrônicos *online* (plataformas, aplicações e *sites*) que permitem o desenvolvimento da atividade da empresa e dos trabalhadores, conectando clientes a prestadores de serviços.

As plataformas escolhem ofertas mais convenientes por meio de critérios baseados em modelos padrões de trabalhador que têm como características alta disponibilidade e pouca resistência aos comandos dos algoritmos.

Em relação especificamente aos trabalhadores digitais que laboram como criadores de conteúdo *online*, suas receitas dependem exclusivamente de como sua *performance* é medida pelos algoritmos das redes, ou seja, os números de seguidores, porcentagem de engajamento, curtidas, cliques e compartilhamento alcançados pelos obreiros.

É por meio da taxa de acessos dos usuários ao conteúdo dos trabalhadores que plataformas como TikTok, Instagram, YouTube, Facebook, Twitter, Snapchat monetizam o conteúdo em rede e lucram com a publicidade imbricada acessada.

É também a partir dessas estatísticas que blogueiros e influenciadores se tornam visíveis para as plataformas de *marketing* e marcas que os procuram em busca de parcerias e contratos publicitários.

A categoria de trabalhadores digitais vive momento semelhante àquele vivenciado pelos operários das indústrias do século XIX, em que a mudança na organização do trabalho da casa para a fábrica se operou em razão da necessidade do uso de uma tecnologia da qual os obreiros não eram titulares.

Sabe-se que foi desta mudança que surgiram as primeiras reivindicações trabalhistas, pois os empregadores da indústria – os proprietários da tecnologia indispensável para o trabalho – estavam em uma condição privilegiada que lhes permitia abusos sobre os trabalhadores.

Na Quarta Revolução Industrial, a reorganização do trabalho em plataformas não parece modificar a necessidade de o trabalhador recorrer a uma tecnologia alheia – antes o maquinário fabril, agora, as redes sociais virtuais, plataformas de diversas naturezas e os seus respectivos algoritmos –, que lhes permite produzir, ainda que de maneira subordinada e dependente dos ditames dos detentores da inteligência operacional algorítmica.

3 A preocupação com a ética no desenvolvimento da inteligência artificial

Machine learning, traduzido como aprendizado de máquina, pode ser definido como um processo automatizado de descoberta de correlações, relacionamentos ou padrões entre variáveis de uma base de dados, utilizado frequentemente para realizar predições ou estimar algum resultado[18].

Na prática, é um processo baseado na otimização de uma função de erro (ou custo), na qual o algoritmo se torna mais preciso na resolução do problema à medida que sucessivas tentativas fornecem um *feedback* (retroalimentação) que levam ao aperfeiçoamento.

Algoritmos de *machine learning* supervisionados são usados quando uma variável *target* (alvo) é conhecida. Um exemplo de utilização de *machine learning* ocorre na seleção de empregados, na qual se busca saber qual seria a *performance* (variável alvo) de um candidato caso fosse contratado.

Para isso, o algoritmo é treinado com dados dos empregados que já passaram pela empresa, predizendo a *performance* do candidato a partir do desempenho que outros empregados com perfil similar obtiveram no passado.

No processo de *machine learning*, diversos estágios passam por tomadas de decisão humana que podem provocar o surgimento de vieses no algoritmo, tais como a definição de *target variables* (variáveis de saída, variáveis alvo, resultado almejado) e *class labels* (rótulo de classe, ou seja, a maneira como os resultados serão interpretados); a escolha da base de dados que será usada para treinar o algoritmo (*training data*) – que pode ser enviesada em razão de discriminações históricas ou em razão de sub-representação ou sobrerrepresentação de grupos específicos na base de dados – e a seleção das variáveis de entrada que serão acessadas pelo algoritmo para buscar padrões e correlações[19].

Dessa forma, caso se use o modelo de aprendizagem de máquina para máquina para a contratação de um empregado, a empresa que recruta busca por um bom funcionário. E daí surgem pontos. O que seria um bom funcionário? O que não se atrasa, o mais qualificado, o mais querido pelos clientes?

Essas perguntas são muito subjetivas, e, por exemplo, se for apontado que um bom funcionário é o que trabalha por mais horas, deve ser levado em conta que mulheres estatisticamente trabalham menos que homens em razão da licença-maternidade, senão haverá um impacto discriminatório.

---

18 LEHR, Davida; OHM, Paul. De playing with the Data: What Legal Scholars Should Learn About Machine Learning. *UC Davis Law Review*, v. 51, 2017, p. 671.

19 BARROCAS, Solon; SELBST, Andrey. Big data's disparate impact. *California Law Review*, v. 104, n. 671, 2016.

Outro exemplo de possível impacto discriminatório é quando o desenvolvedor define, como uma das variáveis de entrada, a reputação das universidades em que o candidato se formou, podendo sistematicamente prejudicar estudantes de baixa renda, haja vista que estes normalmente não frequentam as universidades reconhecidas como de maior prestígio, em que pese o fato de que eles possam ter um desempenho acadêmico similar ao de um aluno de renda superior.

Para Barocas e Selbst, a escolha das variáveis pelos programadores pode introduzir vieses nos algoritmos ao serem selecionadas variáveis que são mais preditivas para membros de certos grupos ou que servem como *proxies* para membros de um grupo, tais como código postal, realização ou não de intercâmbio[20].

Dessa forma, a remoção do gênero ou raça das variáveis de entrada não garante que o processo de *machine learning* não faça correlações baseadas em *proxies*, causando um impacto discriminatório, ainda que não intencional.

Kleinberg *et al.* apontam que, na fase de determinação de quais variáveis coletadas serão consideradas preditoras para alcançar o resultado almejado, o papel do ser humano possui menor impacto[21].

Isto se deve ao fato de que a seleção, pelo algoritmo, de quais varáveis terão alto grau de predição, dada a escolha da base de dados utilizada no treinamento e do resultado (variáveis de saída já rotuladas), "é basicamente apenas uma questão estatística de quais variáveis são mais correlacionadas com o resultado".

Ocorre que David Lehr e Paul Ohm explicam que é um erro focar nossas atenções apenas sobre esses estágios do aprendizado de máquina, visto que estágios finais de *machine learning – data cleaning* (limpeza de dados), *summary statistics review* (análise exploratória), *data partitioning* (particionamento de dados), *model selection* (seleção de modelo/algoritmo), *model training* (treinamento de modelo) e *model deployment* (exposição do modelo ao uso real) – oferecem oportunidades para remediar os vieses causados pelos estágios anteriores, quais sejam, a definição da variável de saída e suas categorias, coleta e categorização da base de dados e seleção das variáveis de entrada que serão acessadas pelo algoritmo, mas que nem sempre são utilizadas[22].

---

20 *Idem.*

21 KLEINBERG, Jon; LUDWIG, Jens; MULLAINATHAN, Sendhil; SUNSTEIN, Cass. *Discrimination in the age of algorithms*, 2019.

22 LEHR, Davida; OHM, Paul. Playing with the data: what legal scholars should learn about machine learning. *UC Davis Law Review*, v. 51, 2017, p. 671.

Kleinberg *et al*. também nos alertam para o fato de que "só se pode prever resultados usando observações de treinamento sobre aqueles para quem observamos o resultado"[23].

Seria o caso de uma empresa em que mulheres, historicamente, não se candidatam, de forma que a base de dados seria composta por amostras insuficientes para estimar as previsões de *performance* para candidatas, mas poderia prever os resultados para o grupo com maior presença na amostra (homens) com melhor precisão.

Dessa forma, resta claro que a objetividade e eliminação das discriminações pretendidas com a substituição da tomada de decisão humana pela decisão automatizada ou a utilização da inteligência artificial como forma de apoio à decisão humana não são necessariamente atingidas.

Dado o importante papel que as decisões humanas possuem na construção do algoritmo, seria irresponsável, e até mesmo perigoso, confundir a programação orientada a dados com a garantia de não discriminação e ausência de vieses.

Conclui-se, portanto, que, embora os vieses dos resultados obtidos nos modelos de *machine learning* supervisionados estejam comumente associados à base de dados, há também casos em que os vieses podem emergir antes mesmo da coleta de dados, ou em outros estágios do processo de *machine learning*, em função das decisões tomadas pelos desenvolvedores.

Identificar a influência da subjetividade humana no *design* e na configuração do algoritmo não é uma tarefa fácil e, frequentemente, a discriminação só se torna aparente quando um uso problemático do modelo surge.

Dessa forma, tamanha é a responsabilidade do desenvolvedor de tecnologias no contexto da ética da inteligência artificial.

É de grande relevância compreender a importância da participação dos cientistas de dados na garantia da ética na inteligência artificial. A objetividade e extinção das discriminações pretendidas com a substituição da tomada de decisão humana pela decisão automatizada ou a utilização da inteligência artificial como forma de apoio à decisão humana não são necessariamente atingidas.

As decisões humanas possuem um importante papel na construção do algoritmo, seria irresponsável, e até mesmo perigoso, confundir a programação orientada a dados com a garantia de não discriminação e ausência de vieses[24].

---

23 KLEINBERG *et al.*, *op. cit.*

24 KLEINBERG, Jon; LUDWIG, Jens; MULLAINATHAN, Sendhil; SUNSTEIN, Cass. Discrimination in the age of algorithms. *Journal of Legal Analysis*, v. 10, p. 113-174, 2018. p. 138.

Além disso, deve-se considerar que a opacidade e a imprevisibilidade associadas aos modelos de *machine learning* impõem dificuldades à concepção tradicional de responsabilidade do desenvolvedor, em razão do *gap* entre controle do desenvolvedor e o comportamento do algoritmo.

Dessa forma, embora os vieses dos resultados obtidos nos modelos de *machine learning* supervisionados estejam comumente associados à base de dados, há também casos em que os vieses podem emergir antes mesmo da coleta de dados, ou em outros estágios do processo de *machine learning*, em função das decisões tomadas pelos desenvolvedores.

Identificar a influência da subjetividade humana no *design* e na configuração do algoritmo não é uma tarefa fácil e, frequentemente, a discriminação só se torna aparente quando um uso problemático do modelo surge[25].

Ainda neste contexto, é preciso considerar que as decisões tomadas pelo desenvolvedor, além de gerarem impactos no desempenho do algoritmo e, consequentemente, na sociedade, não podem ser vistas como neutras. Dilemas morais surgirão em momentos em que os programadores precisem tomar decisões que não possuem soluções fáceis e, muitas vezes, serão os únicos agentes a entenderem as vantagens e os perigos gerados pela tecnologia.

Ben Green, ao tratar da atuação dos cientistas de dados, conclui que estes "devem se reconhecer como atores políticos engajados em construções normativas de sociedade e, como convém à prática política, avaliar seu trabalho de acordo com seus impactos na vida das pessoas"[26].

O autor refuta os argumentos comumente invocados para evitarem posições políticas no que se refere ao trabalho, quais sejam, "eu sou apenas um engenheiro" (engenheiros apenas desenvolvem a tecnologia, não determinam como a tecnologia será usada), "nosso trabalho não é tomar decisões políticas" (quanto mais neutro, melhor), "nós não podemos deixar o perfeito ser inimigo do bom" (embora não sejam perfeitas, ferramentas desenvolvidas por cientistas de dados contribuem para a sociedade de forma positiva, portanto, deveríamos focar em apoiar o desenvolvimento dessas tecnologias em vez de discutir sobre o que seria a solução perfeita).

Quanto ao primeiro argumento – "eu sou apenas um engenheiro" –, o autor defende que, embora a tecnologia não se enquadre na noção convencional de política, há de se considerar que seu uso frequentemente molda aspectos da

25 MITTELSTADT, Brent Daniel *et al*. The ethics of algorithms: mapping the debate. *Big Data & Society*, v. 3, Issue2, July-December 2016, p. 2.

26 GREEN, Ben. *Data science as political action*: grounding data science in a politics of justice. July, 2020, p. 1. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abs-tract id=3658431. Acesso em: 10 abr. 2023.

sociedade, assim como a lei, as eleições e as decisões judiciais, na medida em que podem modificar o comportamento das pessoas e as estruturas de poder.

Quanto ao segundo argumento – "nosso trabalho não é tomar decisões políticas" – o autor questiona a suposta neutralidade dos cientistas de dados. Para o autor, a neutralidade é uma meta inatingível, dado que é impossível se envolver na ciência ou política sem ser influenciado por valores e interesses antecedentes.

Além disso, a neutralidade, embora possa parecer ser apolítica, costuma ser uma posição fundamentalmente conservadora, de aquiescência a valores sociais e políticos dominantes e que contribuem para a conservação do *status quo*. A produção científica, mesmo quando conduzida sob o modelo da objetividade, requer desenvolver perguntas, hipóteses, protocolos e objetivos que são moldados pelos contextos sociais que o geraram.

Quanto ao terceiro argumento – "nós não podemos deixar o perfeito ser inimigo do bom" –, o autor afirma que a ciência de dados carece de teorias robustas acerca do que os conceitos de "perfeito" e "bom" realmente implicam, isto é, não há definição praticável do que pode ser considerado o bem comum a fim de guiar o trabalho dos cientistas de dados.

Além disso, há que se discutir a relação entre o perfeito e o bom, não estando ainda claro se esses esforços para fazer o bem estão, na verdade, fazendo o bem de forma consistente, isto é, é necessário discutir-se a relação entre intervenções algorítmicas e impactos sociais.

Um último ponto diz respeito à existência de vieses entre os próprios desenvolvedores. No Brasil, por exemplo, segundo pesquisa realizada pelas organizações *Thoughtworks* e *Pretalab*, entre os meses de novembro de 2018 e março de 2019, as mulheres correspondiam a apenas 31,7% dos profissionais que trabalhavam com Tecnologia da Informação[27].

Além disso, em 64,9% das empresas, as mulheres representavam apenas 20% da equipe de tecnologia e, em 32,7% dos casos, não havia qualquer pessoa negra na composição das equipes de trabalho.

Esta baixa diversidade entre os profissionais de desenvolvimento de *software* segue a tendência internacional, prejudicando a construção de diferentes perspectivas e formas de raciocínio, colocando nas mãos de um grupo social específico decisões que geram diversos impactos sociais e, por vezes, negativos.

Ocorre que é necessário discutir como a abordagem ética tem respondido a estes problemas, considerando a natural preocupação com a ética no desenvolvimento da inteligência artificial.

---

27 THOUGHTWORKS; PRETALAB. *Quem coda br*. Disponível em: https://www.pretalab.com/report-quem-coda. Acesso em: 10 abr. 2023.

Diversas iniciativas internacionais passaram a abordar a incorporação de questões éticas ao desenvolvimento da IA ao longo dos últimos anos.

É o caso, por exemplo, da criação do K&L Gates *Endowment for Ethics and Computational Technologies* na Carnegie Mellon University em 2016, a fim de fomentar pesquisa e a educação sobre as questões éticas e políticas que surgem dos avanços em inteligência artificial e outras tecnologias computacionais[28].

Além dessa, há a publicação da Declaração de Montreal para o Desenvolvimento Responsável da Inteligência Artificial (*Montreal Declaration for a Responsible Development of Artificial Intelligence*), em 2017, com o objetivo de estimular o debate público e encorajar o desenvolvimento da IA sob uma orientação progressiva e inclusiva[29].

O lançamento da *Ethics and Governance of AI Initiative* em 2017, através de uma parceria entre o MIT Media Lab e o Harvard Berkman-Klein Center for Internet and Society, com o objetivo de garantir que as tecnologias de automação e aprendizado de máquina sejam pesquisadas, desenvolvidas e implantadas de uma forma que reivindique valores sociais de equidade, autonomia humana e justiça também foi uma dessas iniciativas internacionais[30].

Na *Conference on Beneficial AI*, organizada pela Future of Life Institute em 2017, se discutiu um conjunto de princípios que ficaram conhecidos como Asilomar Principles[31]; da publicação das Orientações Éticas para uma IA de Confiança pelo Grupo de peritos de alto nível sobre a inteligência artificial da Comissão Europeia em 2019[32]; da atualização do Código de Ética e Conduta Profissional da ACM (Association for Computing Machinery) em 2018, reafirmando a obrigação dos profissionais em utilizar suas habilidades para benefício da sociedade[33] e, em abril de 2021, da apresentação da Proposta para regulamentação das tecnologias de inteligência artificial (Artificial Intelligence Act) da Comissão Europeia[34].

---

28 CARNEGIE MELLON UNIVERSITY. *Carnegie Mellon University Announces K&L Gates Professorships*. Disponível em: https://www.cmu.edu/news/stories/archives/2018/april/kl-gates-professorships.html. Acesso em: 10 abr. 2023.

29 UNIVERSITÉ OF MONTRÉAL. *Montréal Declaration Responsible AI*. Disponível em: https://www.montrealdeclaration-responsibleai.com. Acesso em: 10 abr. 2023.

30 AIETHICSINITIATIVE. *The ethics and governance of artificial intelligence initiative*. Disponível em: https://aiethicsinitiative.org. Acesso em: 01 set. 2021.

31 FUTURE OF LIFE INSTITUTE. *Asilomar AI Principles*. Disponível em: https://futureoflife.org/ai-principles. Acesso em: 10 abr. 2023.

32 COMISSÃO EUROPEIA. *Orientações éticas para uma IA de confiança.* Disponível em: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-pt/ format-PDF. Acesso em: 10 abr. 2023.

33 ASSOCIATION FOR COMPUTING MACHINERY. *ACM Code of Ethics and Professional Conduct*. Disponível em: https://www.acm.org/code-of-ethics. Acesso em: 10 abr. 2023.

34 COMISSÃO EUROPEIA. *Proposta de regulamento da inteligência artificial*. Disponível em: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0004.02/DOC_1&format=PDF. Acesso em: 10 abr. 2023.

Dessa forma, passado o contexto da abordagem ética nos principais documentos, analisemos as limitações dessa abordagem, bem como a utilização de Códigos de Conduta Ética nas estruturas das empresas.

4 Limites da inteligência artificial e a implementação de códigos de conduta ética nas empresas

Muito se fala sobre o estabelecimento de princípios éticos capazes de orientar o desenvolvimento de programas de inteligência artificial e a proliferação de documentos no cenário internacional que possuem esta abordagem. Há autores que veem essa ênfase na ética como uma resposta limitada para os problemas com que nos deparamos, como foi visto no capítulo anterior.

Por exemplo, Ben Green, ao defender que (i) a ética da ciência dos dados baseia-se em uma divisão artificial entre tecnologia e sociedade, quando, na verdade, não se trata de tecnologias que podem ser projetadas para ter resultados bons ou ruins; (ii) a ética da ciência dos dados raramente é acompanhada por mecanismos capazes de assegurar que engenheiros sigam os princípios éticos ou sejam responsabilizados em caso de inobservância; e (iii) a ética da ciência dos dados carece de substrato normativo explícito[35].

O autor conclui que a abordagem ética fornece estruturas úteis para ajudar os cientistas de dados a refletir sobre sua prática e os impactos de seu trabalho, mas não resolve questões concernentes a quais impactos são desejáveis e como negociar entre perspectivas, tampouco como guiar a inovação tecnológica para esses fins.

No mesmo sentido, Greene, Hoffmann e Stark[36], ao analisarem declarações públicas de princípios éticos por instituições independentes de 2015 a 2018, apontaram que a discussão sobre ética nos sistemas de inteligência artificial se aproxima mais de uma discussão sob a perspectiva da ética corporativa convencional, do que da promoção de justiça política e social.

Além disso, a responsabilidade moral dos desenvolvedores não é foco das principais iniciativas internacionais, que se preocupam mais com o projeto ético do algoritmo (*ethics by design*).

No entanto, os autores atribuem bastante relevância ao reconhecimento de que tecnologias digitais são objeto legítimo de preocupação ética, isto é, que valores humanos são incorporados ao *design* dessas tecnologias.

---

35 GREEN, *op. cit.*

36 GREENE, Daniel; HOFFMANN, Anna Lauren; STARK, Luke. Better, nicer, clearer, fairer: a critical assessment of the movement for ethical artificial intelligence and machine learning. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

A abordagem "ethics by design" diz respeito aos métodos, algoritmos e ferramentas necessários para dotar as tecnologias de inteligência artificial da capacidade de raciocinar sobre os aspectos éticos de suas decisões e garantir que estas se comportem dentro de determinados limites morais[37].

No entanto, isso nos remete a questões tais como "em que medida sistemas de inteligência artificial podem compreender a realidade social em que operam?"; "sistemas de inteligência artificial devem ser tratados como entidades éticas?"; "como a moral, os valores sociais e legais fazem parte do processo de *design*?".

Quaisquer que sejam as respostas a estas perguntas, não é possível afastar o papel central do desenvolvedor na construção desse *design*.

Além disso, há de se considerar que a maior parte dos documentos e iniciativas que tratam de diretrizes e princípios éticos na implementação e utilização de IA são emitidos por organizações localizadas nos Estados Unidos, União Europeia e Reino Unido, o que nos leva a preocupações com a negligência do conhecimento local, pluralismo cultural e justiça global.

Por fim, Thilo Hagerdoff aponta que, em muitos casos, os documentos que abordam a ética no desenvolvimento de IA carecem de mecanismos de *enforcement*, não havendo consequências para violações, consistindo apenas em estratégias de *marketing* para algumas empresas[38].

Segundo o autor, experimentos empíricos mostraram que as diretrizes éticas tiveram influência pouco significativa na tomada de decisão dos desenvolvedores, sendo a ética frequentemente considerada como algo externo, excedente ou algum tipo de complemento para questões técnicas, imposta por instituições externas à comunidade técnica. O autor ainda aponta que carecem aos desenvolvedores sentimento de *accountability* ou visão moral do significado de seus trabalhos.

A *AI4People*, primeiro fórum global a discutir o impacto social da inteligência artificial, recomendou especificamente aos responsáveis pelas tomadas de decisões políticas: apoiarem o desenvolvimento de códigos de conduta autorreguladores das profissões relacionadas à inteligência artificial e análise de dados, com deveres éticos específicos.

Segundo o documento, este tipo de regulação estaria em consonância com o que já ocorre com outros profissionais de áreas socialmente sensíveis, tais como médicos ou advogados. É preciso que requisitos profissionais sejam

37 DIGNUM, Virginia *et al*. *Ethics by design*: necessity or curse? p. 1-2. Disponível em: https://aperto.unito.it/retrieve/handle/2318/1688083/469151/p60-dignum.pdf. Acesso em: 10 abr. 2023.

38 FLORIDI, Luciano *et al*. AI4People – an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machine*, v. 28, p. 689-707, 2018, p. 705.

estabelecidos, não somente no que se refere às obrigações técnicas, mas também ao seu impacto na sociedade[39].

Dessa forma, levando-se cientistas de dados e demais profissionais envolvidos com a inteligência artificial a uma maior responsabilidade pessoal e prestação de contas, não somente na esfera da empresa.

Torna-se bastante importante que haja nos cursos de engenharia disciplinas que versem sobre ética com enfoque não somente em questões éticas normalmente enfrentadas por engenheiros civis, mecânicos ou eletricistas, mas também em dilemas éticos que surgem no contexto da engenharia de *software*.

O desenvolvimento de códigos de conduta ética também é visto como um dos métodos para concretização de uma IA de confiança, conforme Orientações Éticas para uma IA de Confiança (Ethics Guidelines for Trustworthy AI), segundo as quais as organizações e as partes interessadas podem desenvolver códigos de conduta e documentos de política interna da empresa para contribuírem para construir uma IA de confiança[40].

Da mesma forma, a proposta europeia de Regulamento da Inteligência Artificial também incentiva a elaboração de códigos de conduta destinados a fomentar a aplicação voluntária dos requisitos obrigatórios aplicáveis aos sistemas de IA de risco elevado.

A implementação de códigos de conduta ética pode contribuir para que as empresas aproveitem os benefícios gerados pela utilização de IA, como utilização de *machine learning* supervisado em processos decisórios, ao mesmo tempo que mitigam riscos do seu mau uso.

Contudo, para que isso ocorra, sua construção deve ser feita com base na análise de riscos, isto é, os códigos devem ser elaborados somente após identificação, análise, avaliação e tratamento dos riscos, de um ponto de vista não apenas técnico/computacional, mas sob a perspectiva dos valores, missão e objetivos da empresa, considerando os impactos sociais de suas atividades.

Utilizando-se ainda do exemplo da aplicação de *machine learning* para seleção de candidatos, como no início do artigo: ao estabelecer diretrizes para a definição do nível de equidade almejado para o modelo, uma empresa pode controlar, detectar e corrigir vieses que são passíveis de ocorrer estágios de *machine learning*, garantindo assim que o uso da tecnologia contribuirá para que a seleção do empregado ocorra de forma justa.

---

39 *Idem.*

40 COMISSÃO EUROPEIA. *Proposta de regulamento da inteligência artificial.* Disponível em: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0004.02/DOC_1&format=PDF. Acesso em: 10 abr. 2023.

Para isso, o código de conduta ética, assim como ocorre na implementação de programas de *compliance*, deve estar articulado aos outros mecanismos/práticas de conformidade da empresa, indo além da gestão de riscos (considerando-se a tensão entre os potenciais benefícios e danos causados pelos modelos de *machine learning*), englobando também a previsão de aplicação de medidas disciplinares em caso de violação, o estabelecimento de procedimentos que assegurem a pronta interrupção de irregularidades, o monitoramento contínuo das atividades dos desenvolvedores e da tecnologia (com mecanismos de controle como testes, auditorias), o treinamento dos desenvolvedores e o apoio da alta liderança.

É necessário ainda que sejam atribuídas responsabilidades para dar efetividade e concretude ao código, além de recursos humanos e financeiros para averiguar sua devida observância pelos profissionais.

A implementação de códigos de conduta possibilita a redução do risco de violação da privacidade de dados e segredo comercial, contribui para o alinhamento da ética à inovação, diminui o descompasso existente entre regulação e inovação e não exige o conhecimento de informações sobre a lógica interna da caixa-preta dos algoritmos para que seja cumprida.

A empresa IBM identifica como princípios da abordagem ética da inteligência artificial: "o propósito da IA de aumentar a inteligência humana", "dados e *insights* pertencem a quem os produziu" e "tecnologia deve ser transparente e explicável"[41].

Por sua vez, os pilares para tecnologias de inteligência artificial são os seguintes aspectos: explicabilidade, equidade, robustez, transparência e privacidade[42].

Quanto ao princípio da equidade (*fairness*), tido como tratamento equitativo de indivíduos ou grupos de indivíduos, entende-se que, quando adequadamente calibrado, o sistema de IA pode ajudar seres humanos a fazerem escolhas mais justas, combatendo os preconceitos humanos e promovendo a inclusão. Além disso, a empresa destaca como vieses conscientes ou inconscientes podem tornar a saída (*output*) do sistema injusta, devido às limitações técnicas de seu *design* e expectativas culturais, sociais ou institucionais.

Por fim, a empresa evidencia a importância da diversidade nas equipes de desenvolvimento de IA: "inclusão significa trabalhar para criar uma equipe

41 IBM. *Artificial intelligence*. Disponível em: https://www.ibm.com/artificial-intelligence/ethics. Acesso em: 10 abr. 2023.

42 IBM. *Pilares*. Disponível em: https://www.ibm.com/artificial-intelligence/ai-ethics-focus-areas Acesso em: 10 abr. 2023.

de desenvolvimento diversificada e buscar as perspectivas das organizações que atendem às minorias e comunidades afetadas".

Para tratar de virtudes específicas, o sistema de IA deve possuir, bem como prover diretrizes para desenvolvedores construírem e treinarem IA, a empresa elaborou o documento denominado "Everyday Ethics for AI"[43].

Segundo o documento, *designers* e desenvolvedores de sistemas de IA devem entender as considerações éticas da atividade que realizam.

O documento é claro ao colocar que o sistema inteligente deve ser centrado no ser humano e desenvolvido de maneira a estar alinhado com os valores e princípios éticos da sociedade ou comunidade que afeta, cabendo aos profissionais estarem atentos a questões éticas durante a concepção, construção e manutenção da IA.

Para estabelecimento de uma estrutura ética para desenvolvimento dos sistemas de IA, a empresa foca em cinco áreas: a*ccountability*, alinhamento de valores, explicabilidade, equidade e direitos de proteção dos dados dos usuários.

No que se refere à equidade, o documento alerta para a possibilidade de que os vieses dos desenvolvedores sejam incorporados aos sistemas de IA tratando de diversos vieses inconscientes que podem impactar a atividade dos desenvolvedores, além de recomendar práticas para mitigação dos mesmos vieses.

Verifica-se, portanto, que ao identificar os pilares sobre os quais as tecnologias de inteligência artificial se assentam e, ao estabelecer as diretrizes no documento "Everyday Ethics for AI", a empresa colocou em evidência o papel dos desenvolvedores na construção de uma IA que promova a equidade.

Neste sentido, a empresa também criou um *kit* de ferramentas de código aberto para ajudar os desenvolvedores a examinar, relatar e reduzir a discriminação e os vieses em modelos de *machine learning* durante toda a vida útil da aplicação, inclusive com tutorais específicos para definição de *scores* de crédito e despesas médicas, compartilhando dez algoritmos de mitigação de vieses que estão no estado da arte.

São eles: pré-processamento otimizado, reajuste de pesos, remoção de vieses contraditórios, classificação com rejeição de opinião, remoção de impactos díspares, aprendizado de justa representação, remoção de vieses, pós-processamento calibrado para equalização de chances, pós-processamento para equalização de chances e classificador meta-ajustado para equidade.

---

43 IBM. *Everyday for artificial inteligence*. Disponível em: https://www.ibm.com/watson/assets/duo/pdf/everydayethics.pdf. Acesso em: 10 abr. 2023.

Além destes algoritmos, a empresa aborda métricas utilizadas para medir se grupos e pessoas recebem tratamento similar, como, por exemplo, a distância de Manhattan, a distância euclidiana, a diferença estatística pareada e a diferença da média das chances.

Observa-se que, apesar destas métricas e algoritmos atuarem em vários estágios de *machine learning*, nos estágios de coleta de dados, definição da variável-alvo e da sua rotulação e a definição das varáveis de entrada continuam sendo tarefas usualmente realizadas pelo ser humano, para as quais é mais difícil a aplicação de redução de vieses[44].

A empresa expressa ainda a necessidade de testes de justiça, preconceito, robustez e segurança, além de medidas corretivas necessárias, antes da venda e implantação do sistema de IA ou após o início de sua operação.

Para garantir a conformidade com a abordagem ética, a empresa criou o *AI Ethics Board*, órgão de governança central, interdisciplinar, que deve supervisionar as avaliações de riscos e as estratégias de mitigação de danos, promovendo uma cultura de IA ética, responsável e confiável na empresa.

Por fim, não deixa de ser interessante ver como a postura da empresa modificou-se ao longo do tempo no que se refere à aplicação de tecnologias de reconhecimento facial, tão contestada em razão dos vieses apresentados.

Em carta dirigida ao Congresso Americano em 2020, o CEO se pronunciou se opondo a qualquer tecnologia de reconhecimento facial, para vigilância em massa, perfil racial, violações de direitos humanos básicos e liberdades, ou qualquer propósito que não seja consistente com valores e princípios da empresa, decidindo não mais desenvolver, oferecer ou pesquisar tecnologia de reconhecimento facial[45].

5 Conclusões

O presente trabalho teve como principal finalidade explorar *Big Data*, *Machine Learning* e *Inteligência Artificial*, que passaram a ser usadas de maneira conjunta, e permitiram que os algoritmos estivessem em constante reanálise de padrões de interesses com base em um banco de dados, e analisar se esses padrões podem revelar uma nova forma de dano na dinâmica do trabalho: a "discriminação algorítmica".

---

44 IBM. *AI fairness 360*. Disponível em: http://aif360.mybluemix.net/?_ga=2.38832867.1167512373.1630614877-197245064.1630614877. Acesso em: 03 set. 2021.

45 THE VERGE. *IBM will no longer offer, develop, or research facial recognition technology*. Disponível em: https://www.theverge.com/2020/6/8/21284683/ibm-no-longer-general-purpose-facial-recognition-analysis-software. Acesso em: 01 set. 2021.

Como consequência do avanço da tecnologia, a área trabalhista também foi impactada. Hoje, programas de *software* de inteligência artificial são a cada dia mais utilizados para otimizar o gerenciamento da atividade dos trabalhadores, desde o processo seletivo até o aferimento de produtividade.

Ocorre que a criação de ferramentas tecnológicas não deve ser feita de maneira desenfreada, tampouco deve dispensar a avaliação humana. O uso de algoritmos na admissão e na avaliação dos trabalhadores pode gerar situações discriminatórias.

Para frear esses tipos de situações, as empresas devem se atentar aos princípios trabalhistas no momento de programar os *softwares* de inteligência artificial e, sobretudo, agir com transparência algorítmica.

Assim, a partir das divisões dos capítulos, o presente trabalho se esforçou para esclarecer, primeiramente, um pouco dos conceitos de inteligência artificial, indicando suas funções, como funciona o ciberespaço, o uso dos algoritmos e sua relação com o trabalho atualmente.

Em seguida, tratou-se da preocupação com a ética no desenvolvimento da inteligência artificial, demonstrando o caso concreto da Amazon, que criou um *software* em 2014 como meio de classificar currículos e selecionar os candidatos mais talentosos, ocorre que o sistema foi treinado em um banco de dados composto por um grupo majoritariamente de homens.

Assim, em função dos dados de treinamento não representativos, a tecnologia manifestou uma aprendizagem tendenciosa e não neutra em relação ao gênero.

Além disso, foi possível demonstrar, também, a influência do desenvolvedor de *software*, pois as suas decisões, além de gerarem impactos no desempenho do algoritmo e, consequentemente na sociedade, não podem ser vistas como neutras. Dilemas morais surgirão em momentos em que os programadores precisem tomar decisões que não possuem soluções fáceis e, muitas vezes, serão os únicos agentes a entenderem as vantagens e os perigos gerados pela tecnologia.

Por fim, o estudo tratou de explicar a importância da implementação de Códigos de Conduta Ética na estrutura das empresas, usando o caso concreto da empresa IBM.

Dessa forma, foi possível concluir que a "discriminação algorítmica" é um tipo de dano que pode acontecer nas relações de trabalho atuais, mas através de desenvolvedores de *software* comprometidos com um banco de dados diverso e com empresas que se preocupem em implementar Códigos de Conduta Ética em sua estrutura, teremos uma diminuição desse problema.

Referências

ADAMS-PRASSL, Jeremias. What if your boss was an algorithm? The Rise of Artificial Intelligence at Work, 2019. *Comparative Labor Law & Policy Journal*. Disponível em: https://ssrn.com/abstract=3661151. Acesso em: 10 abr. 2023.

ASSOCIATION FOR COMPUTING MACHINERY. *ACM Code of Ethics and Professional Conduct*. Disponível em: https://www.acm.org/code-of-ethics Acesso em: 10 abr. 2023.

BARROCAS, Solon; SELBST, Andrey. Big data's disparate impact. *California Law Review*, v. 104, n. 671, 2016.

BRASIL. *Lei Geral de Proteção de Dados.* Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.

CASTELLS, Manuel. *A sociedade em rede*. A era da informação: sociedade, economia e cultura. 6. ed. Rio de Janeiro: CIP-Brasil, 1999. v. 1.

CHANDER, A. The racist algorithm? *Michigan Law Review*, v.115, n. 6, p. 1022-1045, 2017.

COMISSÃO EUROPEIA. *Orientações éticas para uma IA de confiança.* Disponível em: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-pt/format-PDF. Acesso em: 10 abr. 2023.

COMISSÃO EUROPEIA. *Proposta de Regulamento da Inteligência Artificial*. Disponível em: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0004.02/DOC_1&format=PDF. Acesso em: 10 abr. 2023.

CORVALÁN, Juan Gustavo. Inteligencia artificial: retos, desafíos y oportunidades – Prometea: la primera inteligencia artificial de Latinoamérica al servicio de la Justicia. *Revista de Investigações Constitucionais*, Curitiba, v. 5, n. 1, p. 295-316, jan./abr. 2018.

CUNHA DE SOUZA, Francisco Rodrigo. *EAD*: cibercultura, tecnologias educacionais e educação. Disponível em https://meuartigo.brasilescola.uol.com.br/educacao/ead-cibercultura-tecnologias-educacionais-educacao.htm. Acesso em: 04 abr. 2023.

FLORIDI, Luciano *et al*. AI4People – an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machine*, v. 28, p. 689-707, 2018, p. 705.

FUTURE OF LIFE INSTITUTE. *Asilomar AI Principles.* Disponível em: https://futureoflife.org/ai-principles. Acesso em: 10 abr. 2023.

GREEN, Ben. *Data science as political action*: grounding data science in a politics of justice. July, 2020, p. 1. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abs-tract id=3658431. Acesso em: 10 abr. 2023.

GREENE, Daniel; HOFFMANN, Anna Lauren; STARK, Luke. Better, nicer, clearer, fairer: a critical assessment of the movement for ethical artificial intelligence and machine learning. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

JANNOTTI DA ROCHA, Cláudio *et al*. Discriminação algorítmica no trabalho digital. *Rev. Dir. Hum. Desenv. Social*, Campinas, 2020.

KLEINBERG, Jon; LUDWIG, Jens; MULLAINATHAN, Sendhil; SUNSTEIN, Cass. *Discrimination in the age of algorithms*, 2019.

KLEINBERG, Jon; LUDWIG, Jens; MULLAINATHAN, Sendhil; SUNSTEIN, Cass. Discrimination in the age of algorithms. *Journal of Legal Analysis*, v. 10, p. 113-174, 2018.

LEHR, Davida; OHM, Paul. Playing with the data: what legal scholars should learn about machine learning. *UC Davis Law Review*, v. 51, 2017, p. 671.

LÉVY, Pierre. *Cibercultura*. São Paulo: Editora 34, 1999.

MATEUS, Kátia. *Como a Amazon foi atraiçoada pelo algoritmo sexista*. Disponível em: https://expresso.pt/economia/2018-12-09-Como-a-Amazon-foi-atraicoada-pelo-algoritmo-sexista. Acesso em: 15 abr. 2023.

MITTELSTADT, Brent Daniel *et al*. The ethics of algorithms: mapping the debate. *Big Data & Society*, v. 3, Issue2, July-December 2016, p. 2.

PUBLICA. A uberização do trabalho é pior pra elas. *Publica*, São Paulo, 28 maio 2019.

SCHWAB, Klaus. *A quarta revolução industrial*. São Paulo: Edipro, 2019.

SHUENQUENER, Valter; LIVIO GOMES, Marcus. *Inteligência artificial e aplicabilidade prática no direito*. Brasília: Conselho Nacional de Justiça, 2022, p. 315-317.

THOUGHTWORKS; PRETALAB. *Quem coda br*. Disponível em: https://www.pretalab.com/report-quem-coda. Acesso em: 10 abr. 2023.

TRIBUNAL REGIONAL DO TRABALHO DA 3ª REGIÃO. *Revolução digital*: impactos no direito e processo do trabalho. Belo Horizonte, 2020, p. 315-327.

WORLD ECONOMIC FORUM. *HR4.0:* shaping people strategies in the Fourth Industrial Revolution. Geneva: World Economic Forum, 2019.

Como citar este texto:
TEIXEIRA, Sergio Torres; PIMENTEL, Alexandre Freire; PUGLIESI, Camila Crasto. Discriminação algorítmica nas relações de trabalho. *Revista do Tribunal Superior do Trabalho*, Porto Alegre, v. 91, n. 2, p. 116-142, abr./jun. 2025.