

# Governamentalidade algorítmica e novas práticas punitivas

*Algorithmic governmentality and new punitive practices*  
*Gubernamentalidad algorítmica y nuevas prácticas punitivas*

*Ana Clara Santos Elesbão*<sup>1</sup>

Pontifícia Universidade Católica do Rio Grande do Sul. Brasil

*Augusto Jobim do Amaral*<sup>2</sup>

Universidade de Coimbra. Portugal

*Felipe da Veiga Dias*<sup>3</sup>

Universidad de Sevilla. España

Revista Derechos en Acción ISSN 2525-1678/ e-ISSN 2525-1686

Año 6/Nº 20 Invierno 2021 (21 junio a 20 septiembre) p. ej.: 125-180

<https://doi.org/10.24215/25251678e549>

Recibido: 01/09/2021

Aprobado: 15 /09/2021

<https://orcid.org/0000-0003-0394-0295>

<https://orcid.org/0000-0003-0874-0583>

<https://orcid.org/0000-0001-8603-054X>

<sup>1</sup> Mestra em Ciências Criminais pelo Programa de Pós-Graduação em Ciências Criminais da Pontifícia Universidade Católica do Rio Grande do Sul, pesquisa financiada pela Capes. Bacharelada em Ciências Jurídicas e Sociais pela Escola de Direito da Pontifícia Universidade Católica do Rio Grande do Sul. Vinculada ao Grupo de Pesquisa em Criminologia, Cultura Punitiva e Crítica Filosófica. <https://orcid.org/0000-0003-0394-0295>. E-mail: [anaelesbaos@gmail.com](mailto:anaelesbaos@gmail.com).

<sup>2</sup> Pós-Doutor em Filosofia Política pela Università degli Studi di Padova – Itália. Doutor em História do Pensamento pela Universidade de Coimbra – Portugal – e Doutor em Ciências Criminais pela PUCRS. Professor do Programa de Pós-Graduação em Ciências Criminais e do Programa de Pós-Graduação em Filosofia da PUCRS. <http://orcid.org/0000-0003-0874-0583>. E-mail: [guto\\_jobim@hotmail.com](mailto:guto_jobim@hotmail.com).

<sup>3</sup> Pós-doutor em Ciências Criminais pela PUC/RS. Doutor em Direito pela Universidade de Santa Cruz do Sul (UNISC) com período de Doutorado Sanduíche na Universidad de Sevilla (Espanha). Professor do Programa de Pós-Graduação em Direito da Faculdade Meridional (IMED) – Mestrado. Professor do curso de Direito da Faculdade Meridional (IMED) – Passo Fundo – RS. Brasil. Coordenador do Grupo de Pesquisa “Criminologia, Violência e Controle”. Advogado. <http://orcid.org/0000-0001-8603-054X>. E-mail: [felipevdias@gmail.com](mailto:felipevdias@gmail.com).

**Resumo:** diante do novo contexto sociotécnico em que algoritmos e dados configuram novas formas de controle social, este trabalho tem como objetivo apontar o modo como se configuram novas práticas punitivas., buscando responder a questão de saber: de que maneira os algoritmos configuram novas práticas punitivas? Para tanto, a pesquisa explora através da realização de uma revisão bibliográfica o modo como sistemas informáticos em rede têm servido como método multifacetado de categorização e classificação social cujo objetivo é o de gerenciar populações influenciando pessoas, canalizando escolhas e determinando oportunidades. Tal método multifacetado funciona estimando probabilidades para antecipar futuros passíveis de intervenção, ensejando um novo modo de gestão algorítmica que funciona conforme um mecanismo de triagem que condena o presente ao futuro antecipado e coloca em jogo o próprio campo de experiência e de ação possíveis dos sujeitos. Esse novo modo de gestão reforça desigualdades ao restringir e condicionar oportunidades em função de interesses privados ou mesmo governamentais, favorecendo os afortunados e castigando os já menos favorecidos ao codificar injustiças do passado em sistemas integrados de pontuação que se comportam como profecias autorrealizáveis com amplo impacto de discriminação sistematizada. Dessa forma, os algoritmos ajudam a criar o entorno que justifica suas suposições produzindo um dano generalizado ao instaurar uma dinâmica punitiva que abrange expansivamente todas as instâncias de julgamento humano em que previsões algorítmicas podem ser utilizadas.

**Palavras-chave:** Algoritmos; Práticas Punitivas; Controle Social; Big Data.

**Abstract:** due to the new sociotechnical context in which algorithms and data configure new forms of social control, this work aims at the commitment or the way in which new punitive practices are configured, seeking to answer the question: how will algorithms configure new punitive practices? Therefore, the research explores by conducting a literature review or how the computer systems on the network served as a multifaceted method of categorization and social classification whose objective is to manage populations

influencing people, channeling escorts and determining opportunities. Such multifaceted method works by estimating probabilities to anticipate future intervening passages, teaching a new algorithmic management mode that works according to a screening mechanism that condemns the present to the anticipated future and places in the same field of experience and action with two subjects. This new mode of management reinforces inequalities or restricts and conditions opportunities based on private or government interests, favoring the fortunate and punishing the less fortunate or codifying past injustices in integrated pointological systems that behave like self-fulfilling prophecies with the impact of systematic discrimination. In this way, the algorithms will help to create the environment that justifies their assumptions, producing widespread damage or establishing a punitive dynamic that expansively encompasses all human growth instincts in which algorithmic predictions can be used.

**Keywords:** Algorithms; Punitive Practices; Social Control; Big Data.

---

## Introdução

Os processos de algoritmização da vida têm se apresentado como problemática privilegiada situada no cruzamento entre a vida humana e as novas tecnologias digitais, colocando desafios multidimensionais inéditos a enfrentar. Algoritmos têm sido amplamente utilizados em uma nova dinâmica expositiva de gerenciamento de dados para administrar e exercer controle sobre fluxos de informações, engendrando novas práticas punitivas vinculadas principalmente aos interesses econômicos de grandes corporações.

Nesse contexto de aceleração tecnológica, um processo de *datificação* dissipa gradativamente os limites entre *online* e *offline*, na medida em que cada vez mais toda a informação extraída da experiência vital é transformada em dado para fins de quantificação. Suas implicações de uso, apesar do

estatuto de neutralidade e objetividade que reivindicam, demonstram vieses e preconceções danosas expressas em lógicas de distribuição discriminativa de benefícios e encargos com consequências punitivas.

Parte do desafio de entender a lógica algorítmica de punição passa justamente por compreender que as formulações matemáticas utilizadas no processamento automatizado das decisões são elaboradas por seres humanos que imprimem nelas todos os tipos de valores, muitos dos quais promovem racismo, sexismo e falsas noções de meritocracia classista, tornando a problemática ainda mais gravosa e insidiosamente discriminatória.

Assinalar o sentido dessa compreensão implica apontar as tensões que perpassam tais práticas forjando suas linhas de força através de molduras técnicas pré-definidas desde expectativas de gênero, sexualidade, raça, classe e tantos outros delineadores normativos dados a distribuir socialmente as possibilidades e as restrições de acesso, bem como a sequestrar o ritmo vital em favor da gestão das multiplicidades.

Ante esse contexto, este trabalho tem como objetivo explorar e descrever o modo como tais lógicas operam socialmente, respondendo a seguinte questão: de que maneira os algoritmos produzem novas dinâmicas sociais? Para responder a esta questão, o trabalho

## **1. Governo algorítmico e as ocultações a serviço do controle social**

O desenvolvimento de tecnologias de “aprendizado de máquina” tem colocado em circulação algoritmos teoricamente capazes de “aprender” indutivamente a partir de correlações baseadas em conjuntos abundantes de dados utilizados para treinar sistemas de computação a criarem modelos internos “úteis para mundo”, aplicáveis às mais diversas fina-

lidades (Osonde Osoba, William Welser, 2017, p. 4-5)<sup>4</sup>. No entanto, embora sejam descritos como parte do ramo da ciência da computação popularmente conhecido como “inteligência artificial”, essa caracterização é deveras ilusória, tendo em vista que os procedimentos de tratamento e mineração de dados não “ensinam” máquinas a “pensar” como humanos, mas aplicam matemática a grandes quantidades de dados para inferir deles probabilidades com base na identificação de correlações e padrões (Viktor Mayer-Schonberger, Kenneth Cukier, 2013, p. 10)<sup>5</sup>.

Essa nova lógica baseia-se na suposição apriorística de que padrões de similitude representam tendências de repetição, de modo que seu objetivo é identificar correspondências entre elementos compatíveis, utilizá-las para inferir constatações, e otimiza-las prospectivamente para fins de predição (Bernard Harcourt, 2015, p.156)<sup>6</sup>. Bernard Harcourt descreve tais procedimentos de correlação como um salto da lógica baseada em probabilidades estatísticas individualizadas para uma nova racionalidade baseada em similitudes e correspondências que já não leva em consideração quaisquer tipos de causalidades entre as variáveis em questão (Bernard Harcourt, 2015, pp. 145-165)<sup>7</sup>.

<sup>4</sup> OSOBA, Osonde; WELSER, William (2017). *An Intelligence in Our Image: the risks of bias and errors in artificial intelligence*. Rand Corporation, Santa Monica.

<sup>5</sup> MAYER-SCHONBERGER, Viktor; CUKIER, Kenneth (2013). *Big data: a Revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt Books and Media, Boston, Nova York.

<sup>6</sup> HARCOURT, Bernard (2015). *Exposed: desire and disobedience in the digital age*. Harvard University Press, Cambridge, Londres.

<sup>7</sup> Com respeito aos usos estatísticos, Harcourt oferece uma contribuição genealógica pertinente: durante o séc. XIX até o início do séc. XX, a estatística tradicional lastreava-se em métodos atuariais de categorização de pessoas e grupos para previsão de comportamentos com base em análises probabilísticas. O uso de tabelas atuariais estendeu-se rapidamente do domínio da indústria de seguros para diversos outros domínios, em especial a categorização preditiva destinada à identificação de “delinquentes”. Esses desenvolvimentos fundamentavam-se na delimitação de categorias e grupos e na sua subdivisão para fins de comparação em função de uma média. Tratava-se, portanto, de uma

Nesse novo contexto, processos de tomada de decisão tem sido cada vez mais atribuídos a máquinas sofisticadas de processamento codificado de dados em que sistemas algorítmicos automatizados são utilizados para a elaboração de perfis preditivos<sup>8</sup>, entre outras soluções de sucesso. Esses

---

predição fundamentada no pertencimento categorial, de modo que o somente poderia ser compreendido em comparação com as categorias às quais pertencia. A emergência da lógica atuarial foi possibilitada pelo surgimento das estatísticas e da análise probabilística, bem como pelo desenvolvimento de métodos empíricos voltados à previsão do comportamento humano e às suas práticas correlatas de controle social. Constituem, em última análise, procedimentos que assumem a incerteza dos acontecimentos e à quantifica segundo cálculos de probabilidade em termos de risco, dando ensejo ao que parece ser um conhecimento seguro sobre o futuro, suficientemente hígido para instrumentalizar e fundamentar medidas tomadas no presente e produzir formas de controle social efetivo sobre o comportamento humano. Em meados desse mesmo século, o advento da computação possibilitou a emergência de uma abordagem algébrica baseada em cálculos estatísticos multivariados segundo os quais é possível estabelecer adequadamente as relações causais envolvidas na produção de um determinado fenômeno. Introduziu-se, assim, uma nova racionalidade baseada não mais na comparação categorial de grupos, mas nas relações individuais entre as variáveis em questão, compondo uma lógica que visa inspirar processos de tomada de decisões com base em constatações de ordem causal. Instrumentalizada por novos procedimentos de análise probabilística, os métodos algébricos modificaram os métodos atuariais baseados na categorização de grupos para uma lógica mais sensível às variáveis individualizadas. O surgimento dessa nova abordagem está vinculado ao contexto da Guerra Fria, em que um tipo distinto de raciocínio denominado “análise de sistemas” foi aperfeiçoado e passou a ser amplamente aplicado a questões que vão desde estratégias de defesa nuclear até as políticas domésticas de combate à criminalidade. A partir disso, o controle estatístico seguiu se ampliando, expandindo-se dramaticamente para os mais diversos domínios, em especial no contexto norte-americano. Assim, a busca incansável pela ampliação dos estoques de dados e por maior poder de computação ensejou o desenvolvimento de novos tipos de análises estatísticas complexas até então inconcebíveis, e em velocidades inéditas A conjugação e o desenvolvimento desses elementos levaram o séc. XXI a inaugurar uma nova racionalidade cuja lógica está cada vez mais desvinculada das abordagens demográficas, bem como da preocupação com os vínculos causais; uma lógica que difere das formas de racionalidade que a precederam, portanto, mas que deriva delas de modo a preservar alguns de seus elementos ao rearticula-los e atualiza-los com base nas novas técnicas e tecnologias disponíveis. Enfim, pois, a lógica algorítmica baseada em novas práticas estatísticas decisionais (HARCOURT, Bernard (2015). *Exposed: desire and disobedience in the digital age*. Harvard University Press, Cambridge, Londres).

<sup>8</sup> Trata-se de algoritmos utilizados na elaboração de perfis para predição de comportamentos nos mais diversos âmbitos, como consumo, riscos, fidelização, definição de novas clientelas, entre outros. As práticas de *profiling* consistem basicamente na extração de correlações

modelos definem cada vez mais o modo como a informação e a publicidade devem ser distribuídas na *web*, quais bairros devem ser policiados em determinado perímetro urbano, quem deve receber um empréstimo e sob que condições, quais famílias devem ser beneficiadas por programas sociais, quem deve ser selecionado para uma vaga de emprego, quem deve ser punido mais severamente pelo sistema de justiça, entre outras definições.

No entanto, apesar dos múltiplos e amplos efeitos de destacada relevância que produzem, os procedimentos algorítmicos são muitas vezes inacessíveis aos mais afetados por seus resultados. Conforme largamente desenvolvido por Frank Pasquale, os algoritmos funcionam como “caixas pretas” inescrutáveis cujos procedimentos são misteriosos, ou seja, é possível analisar seus dados de entrada (*inputs*) e seus dados de saída (*outputs*), mas muitas vezes não é possível analisar os processos segundo os quais uns transformam-se nos outros, tampouco quais critérios informam a análise (Frank Pasquale, 2015, p. 9)<sup>9</sup>.

Muitos dos algoritmos-chave que afetam a vida pública de maneira significativa são considerados segredos proprietários ou comerciais, criando uma relação de assimetria entre aqueles que estão submetidos à sua análise e as corporações ou instituições governamentais interessadas no processo, o que representa desafios complexos na compreensão de seus métodos. Assim, avaliadores de crédito, companhias de Internet, grandes bancos, entre outras corporações e instituições governamentais, utilizam-se dos dados das pessoas comuns para tomar decisões cruciais sobre suas vidas (ao convertê-

---

automatizadas entre os dados disponíveis para a identificação de semelhanças e projeção de expectativas futuras (ROUVROY, Antoinette; BERNS Thomas (2015). “Governamentalidade algorítmica e perspectivas de emancipação: o dispar como condição de individuação pela relação?” em *Revista Eco Pós: Tecnopolíticas e Vigilância*, v. 18, n. 2, pp. 36-56. Disponível em [https://revistaecopos.eco.ufrj.br/eco\\_pos/article/view/2662](https://revistaecopos.eco.ufrj.br/eco_pos/article/view/2662)).

<sup>9</sup> PASQUALE, Frank (2015). *The black box society: the secret algorithms that control money and information*. Cambridge, Londres: Harvard University Press.

-los em pontuações, classificações, cálculos de risco e listas de observação, etc.), acarretando-lhes consequências vitais com implicações políticas importantes, muitas vezes sem qualquer explicação (Frank Pasquale, 2015, pp. 4-5)<sup>10</sup>.

Respaldados por uma opacidade que os torna imunes a qualquer tipo de verificação, os problemas decorrentes de procedimentos algorítmicos tendem a ser tratados como questões exclusiva e objetivamente técnicas, nublando ainda mais o entendimento e o debate sobre seus efeitos (Frank Pasquale, 2015, p. 9)<sup>11</sup>. Essa opacidade torna especialmente difícil a questão de saber em favor de que interesses os modelos operam, se são informados por critérios ou concepções discriminatórias, se seus procedimentos estão baseados em operações justas ou injustas, etc.

Desse modo, os algoritmos acabam legitimados a proferir vereditos silenciosos sobre a vida comum dos indivíduos sem deixar margem para formas válidas e regulares de contestação; bem como a influenciar suas decisões mais triviais *online* sem que sequer se apercebam. Sua importância fundamental define o novo regime de verdade digital em que, segundo Pasquale, “o Facebook define quem somos, a Amazon define o que queremos, o Google define o que pensamos, [...] o sistema financeiro define o que temos e os sistemas de reputação definem, cada vez mais, nossas oportunidades” (Frank Pasquale, 2015, p. 15)<sup>12</sup>:

Uma pontuação de crédito ruim pode custar ao mutuário centenas de milhares de dólares, mas ele nunca entenderá exatamente como foi calculada. Uma empresa de análise preditiva pode classificar alguém como um trabalhador de

<sup>10</sup> PASQUALE, Frank (2015). *The black box society: the secret algorithms that control money and information*. Cambridge, Londres: Harvard University Press.

<sup>11</sup> PASQUALE, Frank (2015). *The black box society: the secret algorithms that control money and information*. Cambridge, Londres: Harvard University Press.

<sup>12</sup> PASQUALE, Frank (2015). *The black box society: the secret algorithms that control money and information*. Cambridge, Londres: Harvard University Press.



“alto custo” ou “não confiável”, mas nunca lhe contar sobre a decisão [...]. Os mecanismos de recomendação da Amazon e do YouTube simulam uma familiaridade automatizada, sugerindo gentilmente ofertas que entendem que gostaríamos (Frank Pasquale, 2015, pp. 4-5, tradução nossa)<sup>13</sup>.

No entanto, à medida que reclamações, pesquisas e vazamentos dão dicas ocasionais sobre o que se passa no interior das “caixas pretas” de processamento algorítmico, surge uma imagem de procedimentos de mineração, tratamento e análise de dados descontextualizados e fora de controle (Frank Pasquale, 2015, p. 21)<sup>14</sup>, cujas regras codificadas promovem interesses ocultos e muitas vezes escusos, produzindo efeitos efetivamente nocivos e especialmente desfavoráveis para determinados grupos<sup>15</sup>.

<sup>13</sup> PASQUALE, Frank (2015). *The black box society: the secret algorithms that control money and information*. Cambridge, Londres: Harvard University Press.

<sup>14</sup> PASQUALE, Frank (2015). *The black box society: the secret algorithms that control money and information*. Cambridge, Londres: Harvard University Press.

<sup>15</sup> Nesse sentido, cf. entre outros: ALI, Muhammad; *et al* (2019). “Discrimination through optimization: How Facebook’s ad delivery can lead to skewed outcomes” em arXiv.org. Disponível em: <https://arxiv.org/abs/1904.02095>. BENJAMIN, Ruha (2019). *Race after technology*. Abolitionist Tools for the New Jim Code. Polity Press, Cambridge; BUOLAMWINI, Joy; GEBRU, Timnit. (2018). “Gender Shades: Intersectional Accuracy Disparities” em Commercial Gender Classification em *Proceedings of Machine Learning Research*, v. 81, pp. 77-91; BRAYNE, Sarah (2021). *Predict and Surveil: Data, Discretion, and the Future of Policing*. Oxford University Press, Nova York; BROWNE, Simone (2015). *Dark Matters: on the surveillance of blackness*. Duke University Press, Durham e Londres; CARRERA, Fernanda. “Racismo e sexismo em bancos de imagens digitais: análise de resultados de busca e atribuição de relevância na dimensão financeira/profissional” em SILVA, Tarcizio (org.). *Comunidades, algoritmos e ativismos digitais: olhares afrodiáspóricos*. Literarua, São Paulo; CARRERA, Fernanda; CARVALHO, Denise (2019). “Algoritmos racistas: uma análise da hiper-ritualização da solidão da mulher negra em bancos de imagens digitais” em *Galáxia Revista do Programa de Pós-Graduação em Comunicação e Semiótica*, n. 43, pp. 99-114. Disponível em <https://revistas.pucsp.br/index.php/galaxia/article/view/41614>; ENSIGN, Danielle; *et al* (2018). “Runaway feedback loops in predictive policing” em *Proceedings of Machine Learning Research*, v. 81, pp. 1-12. Disponível em <http://proceedings.mlr.press/v81/ensign18a.html>; EUBANK, Virginia (2018). *Automating Inequality*. How high-tech tools profile, police, and punish the poor. St. Martins Press, Nova York; JURNO, Amanda; D’ANDREA, Carlos (2018). “Algoritmos e cosmopolíticas: a política de censura à nudez no Facebook e o

Para David Lyon, o novo tipo de vigilância cotidiana de dados levado a cabo pelos sistemas informáticos em rede tem servido como método multifacetado de categorização e classificação social cujo objetivo é o de gerenciar popula-

---

regime de poder dos algoritmos” em *PISEAGRAMA*. Disponível em <https://piseagrama.org/algoritmos-e-cosmopoliticas/>; NAKAMURA, Lisa (2007). *Digitizing race: visual cultures of the Internet*. University of Minnesota Press, Minneapolis, Londres; NOBLE, Safiya Umoja (2016). “A Future for Intersectional Black Feminist Technology Studies” em *Scholar and Feminist Online*. Disponível em <https://sfonline.barnard.edu/traversing-technologies/safiya-umoja-noble-a-future-for-intersectional-black-feminist-technology-studies/>; NOBLE, Safiya Umoja (2018). *Algorithms of oppression: how search engines reinforce racism*. New York University Press, Nova York; NOBLE, Safiya Umoja (2012). “Missed connections: what search engines say about women” em *Bitch Media*, n. 54, pp. 36-41. Disponível em <https://safiyaunoble.com/missed-connections-search-engines-say-women-spring-2012/>; NOBLE, Safiya Umoja (2014). “Teaching Trayvon” em *The Black Scholar*, v. 44, n. 1, pp. 12-29. Disponível em <http://www.jstor.org/stable/10.5816/blackscholar.44.1.0012>; NOBLE, Safiya Umoja (2019). “Google Search: Hiper-visibility as a Means of Rendering Black Woman and Girls Invisible” em *InVisible Culture: Na Eletronic Journal for Visual Culture*, v. 19. Disponível em: <http://ivc.lib.rochester.edu/google-search-hyper-visibility-as-a-means-of-rendering-black-women-and-girls-invisible/>; OSOBA, Osonde; WELSER, William (2017). *An Intelligence in Our Image: the risks of bias and errors in artificial intelligence*. Rand Corporation, Santa Monica; O’NEIL, Cathy (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown Publishers, Nova York.; RITCHIE, Marnie (2020). “Fusing race: the phobogenics of racialized surveillance” em *Surveillance & Society*, v. 18, n. 1, pp. 12-29. Disponível em <https://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/13131>; SILVA, Tarcizio (2019). “Racismo Algorítmico em Plataformas Digitais: microacessões e discriminação em código” em *Simpósio Internacional LAVITS: assimetrias e (in)visibilidades: vigilância gênero e raça*, 6. Anais eletrônicos. Disponível em <http://lavits.org/anais-do-vi-simposio-internacional-lavits-assimetrias-e-invisibilidades-vigilancia-genero-e-raca/?lang=pt>; SILVA, Tarcizio (2019). “Teoria racial crítica e comunicação digital: conexões contra a dupla opacidade” em *Congresso Brasileiro de Ciências da Comunicação*, 42. Anais eletrônicos; SILVA, Tarcizio (2020). “Visão computacional e racismo algorítmico: branquitude e opacidade no aprendizado de máquina” em *Revista da Associação Brasileira de Pesquisadores/as Negros/as (ABPN)*, v. 12, n. 31, pp. 428-448. Disponível em <http://abpnrevista.org.br/revista/index.php/revistaabpn1/article/view/744?fbclid=IwAR1uD0ab3TQcKGAseuaVemudHgBWe-Ep5aKIdDpIG-9tN59Jf8b3BuHt5kQ>; SPEICHER, Till, *et al* (2018). “Potential for discrimination in online targeted advertising” em *Proceedings of Machine Learning Research*, v. 81, pp. 1–15. Disponível em <http://proceedings.mlr.press/v81/speicher18a.html>; SWEENEY, Latanya (2013). “Discrimination in Online Ad Delivery” em arXiv.org. Disponível em <https://arxiv.org/abs/1301.6822>; WACHTER-BOETTCHE, Sara (2017). *Technically Wrong: sexist apps, biased algorithms, and other threats of toxic tech*. W.W. Norton & Company, Nova York, Londres; ZARSKY, Tal Z (2014). “Understanding discrimination in the scored society” em *Washington Law Review*, v. 89, n. 4, pp. 1375-1412. Disponível em [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2550248](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2550248).

ções influenciando pessoas, canalizando escolhas e determinando oportunidades:

A vigilância hoje classifica as pessoas em categorias, atribuindo-lhes valor ou risco, de maneiras que têm efeitos reais em suas chances de vida. Formas de discriminação profunda ocorrem, tornando a vigilância não apenas uma questão de privacidade pessoal, mas de justiça social [...]. Considerar a vigilância como método de classificação social é focar as categorias sociais e econômicas e os códigos computacionais através dos quais dados pessoais são organizados com o objetivo de influenciar e gerenciar pessoas e populações. [...] A chamada exclusão digital não é mais apenas uma questão de acesso à informação. A própria informação pode ser o meio de criar divisões (David Lyon, 2005, pp. 1-2, tradução nossa)<sup>16</sup>.

Nesse registro, conforme apontado por Safiya Noble, o uso quase onipresente de *software* orientado por algoritmos exige uma inspeção mais detalhada de quais valores são priorizados em tais sistemas automatizados de decisão, tendo em vista que tendem a reforçar condições discriminatórias em função de fatores como raça, classe e gênero, ao criar formas de *redlining* tecnológico que estão em ascensão (Safiya Noble, 2018, p. 10)<sup>17</sup>. Segundo autora, embora as práticas de *redlining* sejam mais comuns nos círculos imobiliários e bancários, a discriminação está embutida nos usos cotidianos da tecnologia e da Internet em códigos computacionais dos

<sup>16</sup> LYON, David (2005). "Surveillance as social sorting: computer codes and mobile bodies" em LYON, David, *Surveillance as social sorting: privacy, risk, and digital discrimination*. Routledge, Nova York, Londres.

<sup>17</sup> *Redlining* é um termo em língua inglesa utilizada para designar práticas discriminatórias, principalmente nos círculos imobiliário e bancário, criando condições proibitivas de acesso para pessoas negras ou latinas, por exemplo. Noble propõe um uso mais alargado do termo para designar práticas discriminatórias amplamente disseminadas por algoritmos de processamento de dados, em especial na Internet (NOBLE, Safiya Umoja (2018). *Algorithms of oppression: how search engines reinforce racism*. New York University Press, Nova York).

quais as pessoas estão cada vez mais dependentes, por escolha ou não (Safiya Noble, 2018, p. 10)<sup>18</sup>.

É assim que, nos termos de Harcourt, “as vidas digitais passam a assemelhar-se assustadoramente à existência monitorada dos sujeitos carcerários supervisionados” (Bernard Harcourt, 2015, p. 26, tradução nossa)<sup>19</sup>, na medida em que todas as instâncias da vida comum passam a estar expostas e suscetíveis a vereditos algorítmicos de múltiplos tipos que codificam uma miríade de regras e instruções padronizadas e processam-nas em frações de segundos, ampliando o escopo das práticas punitivas e alastrando suas consequências para todos os campos da vida comum.

Ao concentrar a atenção em seus efeitos mais que em seus pressupostos discursivos de neutralidade, objetividade e eficiência, o que se evidencia é a urgência de apontar os danos, incluindo com especial relevância o seu potencial para produzir punições direcionadas, notadamente vinculadas a práticas discriminativas. Nesse sentido, superadas desde logo as ingenuidades que projetam na técnica algum escopo de neutralidade, mais vale assumir as novas disposições tecnológicas em termos de *algoritarismos*, entendidos, segundo conceituado por Augusto Jobim do Amaral como o

[...] conjunto multidimensional de práticas políticas reatualizáveis por diversos agenciamentos, práticas estas dispostas tecnologicamente a sequestrar o ritmo vital que faz vibrar qualquer sentido, ou seja, modos de um dispositivo ‘dado’ a informar, planificar funções repetíveis e a conformar futuros prováveis sob lógicas de dor padronizadamente aprofundadas (Augusto Jobim do Amaral, 2020, p. 11)<sup>20</sup>.

<sup>18</sup> NOBLE, Safiya Umoja (2018). *Algorithms of oppression: how search engines reinforce racism*. New York University Press, Nova York.

<sup>19</sup> HARCOURT, Bernard (2015). *Exposed: desire and disobedience in the digital age*. Harvard University Press, Cambridge, Londres.

<sup>20</sup> AMARAL, Augusto Jobim (2020). Prólogo. *Algoritarismos*. Tirant Lo Blanch, São Paulo.

O conceito arriscado por Amaral dá a tônica das problemáticas enfrentadas contemporaneamente, notadamente os termos em que o novo regime de práticas, organizado em um dispositivo algorítmico, impinge dor ao subsumir o múltiplo a lógicas de padronização conformadas por disposições tecnológicas segundo as quais “tudo o que é relevante é definido de antemão”, conformando e condenando o horizonte de possíveis ao provável com base em procedimentos de repetição e planificação de funções (Augusto Jobim do Amaral, 2020, p. 11)<sup>21</sup>.

De modo semelhante, Antoinette Rouvroy e Thomas Berns oferecem importante contribuição ao descreverem a emergência de novas formas de controle organizadas em uma *governamentalidade algorítmica*, isto é, a racionalidade pretensamente objetiva “[...] que repousa sobre a coleta, a agregação e a análise automatizada de dados em quantidade massiva de modo a modelizar, antecipar e afetar, por antecipação, os comportamentos possíveis” (Antoinette Rouvroy, Thomas Berns, 2015, p. 42)<sup>22</sup>. Segundo os autores, diferenciando-se da racionalidade jurídico-discursiva baseada na repressão penal - que concerne a ações supostamente cometidas no passado -, a governamentalidade algorítmica projeta sua atuação normativa de modo definitivamente mais ativo, na medida em que “não apenas percebe o possível no presente, produzindo uma ‘realidade aumentada’, uma atualidade dotada de uma ‘memória do futuro’, mas também dá consistência ao sonho de um ‘acaso’ sistematizado” em que o real venha a se tornar o possível antecipadamente, em

---

<sup>21</sup> AMARAL, Augusto Jobim (2020). Prólogo. *Algoritarismos*. Tirant Lo Blanch, São Paulo.

<sup>22</sup> ROUVROY, Antoinette; BERNS Thomas (2015). “Governamentalidade algorítmica e perspectivas de emancipação: o díspar como condição de individuação pela relação?” em *Revista Eco Pós: Tecnopolíticas e Vigilância*, v. 18, n. 2, pp. 36-56. Disponível em [https://revistaecopos.eco.ufrj.br/eco\\_pos/article/view/2662](https://revistaecopos.eco.ufrj.br/eco_pos/article/view/2662).

um ambiente de “liberdade” controlada (Antoinette Rouvroy, Thomas Berns, 2015, p. 48)<sup>23</sup>.

Na mesma direção, Fernanda Bruno indica que os algoritmos configuram uma “gestão de possíveis” em que um “rastreamento proativo” visa estimar probabilidades para antecipar futuros passíveis de intervenção. Segundo a autora, esse novo modo de gestão algorítmica funciona conforme um mecanismo de triagem que pode “condenar o presente ao futuro antecipado”, colocando em jogo o próprio campo de experiência e de ação possíveis dos sujeitos. No seu entendimento, ao “restringir possibilidades de acesso a espaços, informações, produtos e experiências que podem envolver mecanismos de discriminação automatizada”, a dimensão performativa da antecipação expressa um novo tipo de punição cuja ordem é do acesso, e que tem implicações diversas ainda não definidas sobre o modo como distribui oportunidades informacionais, cognitivas, sociais, políticas, etc. (Fernanda Bruno, 2016, pp. 36-37)<sup>24</sup>.

De acordo Virginia Eubank, a experiência do mundo digitalizado e mediado por algoritmos tem sido distribuída socialmente de modo que “a maioria das pessoas costuma ser alvo do escrutínio digital como membros de grupos sociais, e não como indivíduos”. Segundo a autora, “[p]essoas de cor, migrantes, grupos religiosos impopulares, minorias sexuais, pobres e outras populações oprimidas e exploradas carregam um fardo muito maior de monitoramento e rastreamento que os grupos favorecidos”, tendo em vista que costumam ser direcionadas para “suspeita e escrutínio extra” em uma espécie de “sinalização vermelha coletiva”

---

<sup>23</sup> ROUVROY, Antoinette; BERNs Thomas (2015). “Governamentalidade algorítmica e perspectivas de emancipação: o díspar como condição de individualização pela relação?” em *Revista Eco Pós: Tecnopolíticas e Vigilância*, v. 18, n. 2, pp. 36-56. Disponível em [https://revistaecopos.eco.ufrj.br/eco\\_pos/article/view/2662](https://revistaecopos.eco.ufrj.br/eco_pos/article/view/2662).

<sup>24</sup> BRUNO, Fernanda (2016). “Rastrear, classificar e performar” em *Ciência e Cultura*, São Paulo, v. 68, n. 1, pp. 34-38. Disponível em <http://cienciaecultura.bvs.br/pdf/cic/v68n1/v68n1a12.pdf>.

que gera um “ciclo de feedback de injustiças” (Virginia Eubank, 2018, p. 11, tradução nossa)<sup>25</sup>.

No mesmo sentido, dando ênfase ao roteiro tecnológico que inscreve a racionalidade algorítmica, Cathy O’Neil demonstra que algoritmos podem funcionar como “armas de destruição matemática” que aumentam a desigualdade ao restringir e condicionar oportunidades em função de interesses privados ou mesmo governamentais. A autora enfatiza o modo como os modelos algorítmicos utilizados em processos decisórios importantes sobre a vida das pessoas favorecem os afortunados e castigam os já menos favorecidos ao codificar o passado e suas injustiças e criar “espirais viciosas” de retroalimentação de danos em larga escala (Cathy O’Neil, 2016)<sup>26</sup>.

Na definição da autora, um modelo algorítmico é uma representação abstrata e simples de certos processos, que renuncia à precisão e à compreensão detalhada das situações para se concentrar exclusivamente em critérios de eficiência vinculados a determinados interesses. Além disso, respaldados pela opacidade que os torna muitas vezes incontestáveis, bem como por uma promessa discursiva de justiça procedimental, os algoritmos “definem sua própria realidade e utilizam-na para justificar seus resultados”, dando ensejo a um “ciclo de retroalimentação” em que, ao invés de “buscar a verdade”, os resultados produzidos pelo modelo a personificam (Cathy O’Neil, 2016, p. 16, tradução nossa)<sup>27</sup>.

Para O’Neil, os algoritmos refletem objetivos, valores e desejos que acabam por definir desde os dados que são recompilados até as variáveis que informam a estrutura dos modelos, o que confere um peso determinante em sua com-

<sup>25</sup> EUBANK, Virginia (2018). *Automating Inequality*. How high-tech tools profile, police, and punish the poor. St. Martins Press, Nova York.

<sup>26</sup> O’NEIL, Cathy (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown Publishers, Nova York.

<sup>27</sup> O’NEIL, Cathy (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown Publishers, Nova York.

posição às escolhas que determinam quais dados merecem atenção e quais devem ser deixados de fora. Além disso, a própria concepção de que um algoritmo esteja funcionando adequadamente ou não também depende de certas escolhas, tendo em vista que um dos componentes chave de todo modelo é justamente o modo como é definido o seu êxito (Cathy O'Neil, 2016, pp. 19-20)<sup>28</sup>.

Nesse sentido, conforme apontado por Pasquale, um primeiro desafio na compreensão das problemáticas que tocam a questão é justamente entender que as escolhas que constituem a composição dos modelos são realizadas por seres humanos vinculados a agendas políticas, econômicas, sociais, culturais e etc. Por essa razão, embora os algoritmos muitas vezes sejam pensados segundo expectativas e pressupostos de neutralidade e objetividade, é importante ter em conta que os atores envolvidos na sua projeção imprimem em suas formulações crenças e valores de toda a sorte:

Os sistemas automatizados afirmam classificar todos os indivíduos da mesma maneira, evitando assim a discriminação. Eles podem garantir que alguns padrões não baseiem mais as decisões de contratação e tomada de decisões em palpites, impressões ou preconceitos. Mas os engenheiros de software constroem os conjuntos de dados extraídos por sistemas de pontuação; definem os parâmetros das análises de mineração de dados; criam os clusters, links e árvores de decisão aplicadas; geram os modelos preditivos aplicados. Os preconceitos e valores humanos estão embutidos em cada etapa do desenvolvimento. A informatização pode simplesmente impulsionar a discriminação rio acima (Frank Pasquale, 2015, p. 35, tradução nossa)<sup>29</sup>.

---

<sup>28</sup> O'NEIL, Cathy (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown Publishers, Nova York.

<sup>29</sup> PASQUALE, Frank (2015). *The black box society: the secret algorithms that control money and information*. Cambridge, Londres: Harvard University Press.



As escolhas que compõem a estrutura dos modelos não são neutras ou objetivas, portanto, mas baseiam seus critérios muitas vezes em crenças exclusivamente valorativas sem qualquer fundamento consistente, em geral vinculadas aos propósitos e interesses de seus desenvolvedores. Dessa forma, os modelos expressam, em alguma medida, “opiniões integradas em matemáticas” que refletem suas prioridades, mas não sem produzir efeitos (Cathy O’Neil, 2016, p. 32)<sup>30</sup>.

## 2. Práticas punitivas na era do big data

Embora na maioria das vezes os modelos não façam referências discriminatórias explícitas a caracteres como raça, classe, gênero, etc., as variáveis utilizadas para informá-los podem funcionar como equivalentes substitutas, enviesando os resultados gerados (Cathy O’Neil, 2016, p. 23)<sup>31</sup>. Danielle Citron e Frank Pasquale demonstraram o modo como os algoritmos podem produzir efeitos danosos em cascata ao transformar dados em pontuações poderosamente estigmatizantes, tendenciosas e arbitrárias. Segundo os autores, “[o] ato de designar alguém como um provável risco de crédito (ou má contratação, ou motorista imprudente) tende a elevar seu custo de financiamento futuro (ou trabalho, ou taxas de seguro), aumentando a probabilidade de eventual insolvência ou invalidez” (Danielle Keats Citron, Frank Pasquale, 2014, p. 18, tradução nossa)<sup>32</sup>. Assim, os sistemas de pontuação acabam comportando-se como “profecias autorrealizáveis” que “criam a angústia financeira que alegam

<sup>30</sup> O’NEIL, Cathy (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown Publishers, Nova York.

<sup>31</sup> O’NEIL, Cathy (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown Publishers, Nova York.

<sup>32</sup> CITRON, Danielle Keats; PASQUALE, Frank (2014). “The Scored Society: due process for automated predictions” em *Washington Law Review*, v. 89, n. 1, pp. 1-33. Disponível em [https://digitalcommons.law.umaryland.edu/fac\\_pubs/1431/](https://digitalcommons.law.umaryland.edu/fac_pubs/1431/).

apenas indicar” (Danielle Keats Citron, Frank Pasquale, 2014, p. 18, tradução nossa)<sup>33</sup>.

Para Pasquale, embora os modelos de pontuação sejam promovidos como modelos de justiça, costumam ser informados e justificados apenas com generalidades que, ao fim e ao cabo, funcionam em favor de seus próprios interesses. A maioria das pessoas submetidas ao seu escrutínio não costuma sequer estar ciente de que seus atos os mais inócuos estão sendo avaliados em sistemas que tomam decisões importantes sobre suas vidas. Essa é uma realidade alarmante em um contexto no qual os sistemas de pontuação têm escapado cada vez mais o âmbito financeiro e se estabelecido como árbitros da confiabilidade geral em outras áreas como a concessão de seguros, a contratação para vagas de emprego, entre outras aplicabilidades, funcionando como verdadeiros sistemas de reputação (Frank Pasquale, 2015, p. 23)<sup>34</sup>.

Esses sistemas funcionam de modo integrado a outros sistemas de pontuação, adquirindo potencial para produção de um impacto amplo de discriminação sistematizada. Associações entre empregos de baixa remuneração e a atribuição de baixas pontuações de crédito, por exemplo, ajudam a reforçar e ampliar ainda mais o ciclo de retroalimentação danosa que pune as classes economicamente desfavorecidas:

Imagine uma jovem que não conseguiu um emprego na faculdade. Esse fracasso reduziu sua pontuação de “empregabilidade” utilizada por empregadores em potencial para determinar sua aptidão para o trabalho. A jovem conseguiu um emprego de meio período em um restaurante de *fast food*. Sua pontuação de crédito ficou abaixo dos 600 pontos sem que ela soubesse, talvez em função de inferências asso-

<sup>33</sup> CITRON, Danielle Keats; PASQUALE, Frank (2014). “The Scored Society: due process for automated predictions” em *Washington Law Review*, v. 89, n. 1, pp. 1-33. Disponível em [https://digitalcommons.law.umaryland.edu/fac\\_pubs/1431/](https://digitalcommons.law.umaryland.edu/fac_pubs/1431/).

<sup>34</sup> PASQUALE, Frank (2015). *The black box society. the secret algorithms that control money and information*. Cambridge, Londres: Harvard University Press.

ciadas a empregos de baixa remuneração. A baixa pontuação de crédito causou-lhe consequências ruins em cascata, ao colocá-la em circunstâncias de vida cada vez mais desafiadoras. As empresas de análise de talentos a categorizaram como “não inovadora” e “desperdiçadora”. Com pontuações baixas em diversos aspectos, a jovem possivelmente não conseguirá um emprego de tempo integral (Danielle Keats Citron, Frank Pasquale, 2014, p. 32, tradução nossa)<sup>35</sup>.

De modo semelhante, no âmbito dos seguros, modelos algorítmicos têm sido cada vez mais utilizados para avaliar a concessão de apólices com base em cálculos de risco, substituindo gradativamente os tradicionais métodos atuariais. Esses modelos baseiam-se em uma variedade cada vez maior e mais diversificada de dados, desviando muitas vezes o foco de seus critérios tradicionais e produzindo disparidades discriminatórias nos preços. O’Neil exemplifica o caso dos seguros automobilísticos, que têm extraído pontuações de relatórios de crédito para informar seus modelos proprietários, de modo que dados financeiros têm sido mais relevantes na avaliação das apólices que a própria prudência estimada dos motoristas (Cathy O’Neil, 2016, p. 133)<sup>36</sup>.

A partir disso, fica novamente evidenciado o modo como a crescente dependência das pontuações de crédito em diversos domínios prejudica as classes mais baixas. Esse é apenas mais um exemplo dessa ampla tendência fundamentada em sistemas de reputação, mas é especialmente intrigante, tendo em vista que o seguro de automóveis é obrigatório para motoristas. Além disso, o uso de dados equivalentes relacionados a aspectos econômicos nesse

<sup>35</sup> CITRON, Danielle Keats; PASQUALE, Frank (2014). “The Scored Society: due process for automated predictions” em *Washington Law Review*, v. 89, n. 1, pp. 1-33. Disponível em [https://digitalcommons.law.umaryland.edu/fac\\_pubs/1431/](https://digitalcommons.law.umaryland.edu/fac_pubs/1431/).

<sup>36</sup> O’NEIL, Cathy (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown Publishers, Nova York.

caso evidencia o ânimo exclusivo de lucro da atividade dessas corporações, que ao invés de “penalizar” maus motoristas ao atribuir-lhes apólices mais caras (o que poderia ser considerado de alguma forma uma justa medida), segue punindo economicamente aqueles que já estão em evidente desvantagem (Cathy O’Neil, 2016, pp. 133-134)<sup>37</sup>.

Outrossim, modelos de predição de risco também têm sido utilizados por entidades públicas em processos decisoriais de vital importância. Ao investigar o impacto do uso de novas ferramentas para tomada de decisão automatizada em programas de serviço social nos Estados Unidos, Eubank constatou consequências gravemente danosas e nocivas especialmente para as pessoas pobres e a classe trabalhadora, demonstrando o modo como tais modelos algorítmicos “destroem a rede de seguridade social norte-americana, criminalizam os pobres e intensificam a discriminação [...]” (Virgínia Eubank, 2018, p. 15, tradução nossa)<sup>38</sup>.

Analisando um modelo preditivo de risco que promete proteger o bem-estar infantil no Condado de Allegheny, Pensilvânia, Eubank verificou danos consideráveis vinculados a estigmas associados a famílias pobres. A chamada Ferramenta de Rastreamento Familiar de Allegheny (*Allegheny Family Screening Tool* — AFST)<sup>39</sup>, lançada em 2016 pelo Gabinete das Crianças, Jovens e Famílias do Condado de Allegheny (*Allegheny County Office of Children, Youth and Families* — CYF), utiliza modelagem estatística preditiva para “aprimorar” a tomada de decisão sobre a abertura de processos de investigação de abuso e negligência infantil em casos consi-

<sup>37</sup> O’NEIL, Cathy (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown Publishers, Nova York.

<sup>38</sup> EUBANK, Virginia (2018). *Automating Inequality*. How high-tech tools profile, police, and punish the poor. St. Martins Press, Nova York.

<sup>39</sup> THE Allegheny Family Screening Tool (2021). *Allegheny County*. Disponível em <https://www.alleghenycounty.us/Human-Services/News-Events/Accomplishments/Allegheny-Family-Screening-Tool.aspx>.

derados suspeitos. A função do modelo consiste basicamente em auxiliar nesse processo de triagem gerando uma pontuação automática que deverá informar o alerta de risco (Virgínia Eubank, 2018, p. 103-139)<sup>40</sup>.

De acordo com Eubank, o sistema tem falhas de design inerentes que acabam visando injustamente os pobres para o escrutínio do bem-estar infantil, tendo em vista que quase três quartos dos casos reportados acabam sendo de negligência, e não de abuso, e que “quase todos os indicadores de negligência infantil também são indicadores de pobreza”, isto é, “falta de comida, moradia inadequada, creches não licenciadas, transporte não confiável, desligamento de serviços públicos, falta de moradia, falta de assistência médica”, etc. (Virgínia Eubank, 2018, p. 126-127, tradução nossa)<sup>41</sup>. Desse modo, as famílias desabrigadas, por exemplo, enfrentam desafios particularmente difíceis para manter seus filhos, já que a própria condição de sem-teto é considerada uma condição de negligência (Virgínia Eubank, 2018, p. 118)<sup>42</sup>.

Soma-se a isso o fato de que o modelo se baseia para realizar a triagem apenas em dados de famílias beneficiárias de serviços públicos; dados coletados por programas governamentais de tratamento de drogas, aconselhamento de saúde mental, apoio financeiro, etc. Dessa forma, a necessidade de recorrer a recursos públicos para acessar determinados tipos de serviços acaba submetendo as famílias mais precarizadas a uma ingerência maior da autoridade governamental para, eventualmente, retirar-lhes seus filhos, tendo em vista que o escopo informacional utilizado na triagem do programa

<sup>40</sup> EUBANK, Virginia (2018). *Automating Inequality*. How high-tech tools profile, police, and punish the poor. St. Martins Press, Nova York.

<sup>41</sup> EUBANK, Virginia (2018). *Automating Inequality*. How high-tech tools profile, police, and punish the poor. St. Martins Press, Nova York.

<sup>42</sup> EUBANK, Virginia (2018). *Automating Inequality*. How high-tech tools profile, police, and punish the poor. St. Martins Press, Nova York.

acaba reduzido a famílias que não dispõem de recursos para optar por serviços privados (Virgínia Eubank, 2018, p. 126)<sup>43</sup>.

Em suma, pode-se dizer que as variáveis preditivas do modelo carregam um estigma humano embutido, isto é, a interpretação do uso de recursos públicos como um sinal de fraqueza, deficiência e vilania, e a sua conseqüente vinculação à situação de pobreza (Virgínia Eubank, 2018, p.135)<sup>44</sup>. Assim, o modelo acaba confundindo paternidades e maternidades pobres com paternidades e maternidades ruins, assumindo pais e mães que buscam o auxílio de programas públicos como riscos para seus filhos. Segundo Eubank, “podemos chamar isso de definição de perfis de pobreza [...] que selecionam indivíduos como alvos para escrutínio extra não com base em seu comportamento, mas com base na situação de pobreza em que vivem” (Virgínia Eubank, 2018, p. 127-128, tradução nossa)<sup>45</sup>.

A partir de inferências desse tipo, o modelo acaba sobrerrepresentando as pessoas pobres e a classe trabalhadora nas investigações de abuso e negligência infantil, acarretando-lhes conseqüências danosas duradouras de caráter punitivo que aumentam o estigma e as adversidades da pobreza, como vigilância e requisitos de conformidade comportamen-

---

<sup>43</sup> “As famílias evitam a CYF se puderem, porque a agência mistura dois papéis distintos e contraditórios: provedor de apoio familiar e investigador de maus-tratos. Aceitar recursos significa aceitar a autoridade da agência para remover seus filhos. Essa é uma troca invasiva e assustadora que os pais com outras opções provavelmente não escolherão. Famílias pobres e da classe trabalhadora se sentem forçadas a trocar seus direitos à privacidade, proteção contra buscas irracionais e devido processo legal para ter uma chance de obter os recursos e serviços de que precisam para manter seus filhos seguros” EUBANK, Virginia (2018). *Automating Inequality*. How high-tech tools profile, police, and punish the poor. St. Martins Press, Nova York - tradução nossa).

<sup>44</sup> EUBANK, Virginia (2018). *Automating Inequality*. How high-tech tools profile, police, and punish the poor. St. Martins Press, Nova York.

<sup>45</sup> EUBANK, Virginia (2018). *Automating Inequality*. How high-tech tools profile, police, and punish the poor. St. Martins Press, Nova York.

tal mais rigorosos (Virgínia Eubank, 2018, p.129)<sup>46</sup>. Visar famílias de “alto risco” pode agravar ainda mais o problema ao leva-las a se retirarem das redes de serviços e apoio comunitários produzindo situações de isolamento social, privação material e estresse parental que se agravam quando os pais se sentem vigiados, perdem recursos de que precisam ou sentem medo de recorrer ao auxílio de programas públicos (Virgínia Eubank, 2018, p. 135)<sup>47</sup>.

É difícil dizer que um modelo preditivo funciona se ele produz o resultado que está tentando medir. Uma família classificada como de alto risco pelo AFST passará por mais escrutínio do que outras famílias. Comportamentos comuns que poderiam não levantar suspeitas antes de uma pontuação AFST elevada tornam-se motivo de confirmação para a decisão de rastreá-los e investiga-los. [...] Graças em parte a pontuações de risco mais altas, pais tendem a ser alvos de um tratamento mais punitivo, devendo atender às expectativas das agências governamentais e enfrentar julgamentos mais severos. E no caso de perder os filhos, o modelo de risco tenderá a reivindicar outra previsão bem-sucedida (Virgínia Eubank, 2018, p. 135, tradução nossa)<sup>48</sup>.

Conforme alertado por Eubank, a despeito dos efeitos destrutivos que apresentam sobretudo para as comunidades negras e de baixa renda, sistemas como esse têm sido integrados aos serviços de seguridade social nos Estados Unidos em um rito alucinante, reformulando concepções sobre a vida em comunidade com pouca ou nenhuma discussão política sobre seus impactos. Segundo a autora, esse con-

---

<sup>46</sup> EUBANK, Virginia (2018). *Automating Inequality*. How high-tech tools profile, police, and punish the poor. St. Martins Press, Nova York.

<sup>47</sup> EUBANK, Virginia (2018). *Automating Inequality*. How high-tech tools profile, police, and punish the poor. St. Martins Press, Nova York.

<sup>48</sup> EUBANK, Virginia (2018). *Automating Inequality*. How high-tech tools profile, police, and punish the poor. St. Martins Press, Nova York.

texto amplo deve ser adequadamente pensado segundo o histórico de criminalização que há muito sujeita os pobres e a classe trabalhadora a vigilância invasiva, ataques noturnos e políticas públicas estigmatizantes que agravam ainda mais as adversidades da pobreza:

Hoje, forjamos o que chamo de abrigo digital com base em bancos de dados, algoritmos e modelos de risco que prometem eclipsar o alcance e a repercussão de tudo o que veio antes [...] Como as inovações tecnológicas anteriores na gestão da pobreza, o rastreamento digital e a tomada de decisão automatizada escondem a pobreza do público profissional de classe média e dão à nação a distância ética necessária para fazer escolhas desumanas: quem deve conseguir comida e quem deve passar fome, quem deve ter moradia e quem deve permanecer desabrigado e quais famílias devem ser desmembradas pelo estado (Virgínia Eubank, 2018, p. 15, tradução nossa)<sup>49</sup>.

Na mesma linha, *softwares* de perfilamento preditivo têm sido utilizados em sistemas de justiça, em especial nos Estados Unidos, para a automatização de cálculos de reincidência, ou seja, para avaliar a probabilidade de um réu criminal voltar a cometer crimes no futuro. A pontuação produzida pelo modelo pode afetar diretamente o cálculo que define a pena de um condenado, bem como seu direito à fiança e à liberdade condicional, por exemplo, de modo que assume importância vital nas jurisdições em que seu uso tem sido implementado<sup>50</sup>. De acordo com Pamela Ugwu-dike, embora esses modelos costumem ser ostensivamente neutros com re-

<sup>49</sup> EUBANK, Virginia (2018). *Automating Inequality*. How high-tech tools profile, police, and punish the poor. St. Martins Press, Nova York.

<sup>50</sup> ANGWIG, Julia; *et al* (2016). "Machine Bias" em *ProPublica*. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>; ANGWIG, Julia; *et al* (2016). "How we analyzed the COMPAS Recidivism Algorithm" em *ProPublica*. Disponível em <https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>.



lação a critérios raciais, há evidências crescentes de que seus algoritmos superestimam o risco de reincidência representado por grupos racializados, exacerbando sua vulnerabilidade à intervenção ainda mais severa e injustificada da justiça criminal (Pamela UgwuDike, 2020)<sup>51</sup>.

Um clássico estudo realizado por Julia Angwig *et al* demonstrou que o modelo chamado COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions* — COMPAS), elaborado pela empresa norte-americana Northpointe e utilizado em centenas de jurisdições estaduais nos Estados Unidos, apresentou resultados consideravelmente enviesados em termos raciais. Os resultados obtidos na pesquisa apontaram que, entre réus com perfis criminais semelhantes, o aspecto racial apresentou-se decisivo na atribuição de uma nota maior para réus negros com relação a réus brancos, apesar de o modelo não fazer qualquer menção à raça como um critério diferenciador. O estudo demonstrou que réus negros tinham duas vezes mais chances de serem falsamente sinalizados como alto risco de reincidência, enquanto réus brancos tinham quase a mesma probabilidade de serem sinalizados como baixo risco. Essa discrepância foi constatada a propósito da margem de erro do modelo, que indica que, em cada dez réus avaliados como alto risco, quatro deles não vêm efetivamente a reincidir, de modo que, entre eles, essa margem de erro é distribuída desigualmente entre negros e brancos<sup>52</sup>.

De acordo com Wachter-Boettcher, esse resultado deve-se ao fato de que, entre os 137 indicadores utilizados pela empresa desenvolvedora para informar o modelo, isto é, os chamados “fatores de risco ou necessidades criminogênicas”,

<sup>51</sup> UGWUDIKE, Pamela (2020). “Digital prediction technologies in the justice system: the implications of a ‘race-neutral’ agenda” em *Theoretical Criminology*, v. 14, n. 3, pp. 482-501. Disponível em <https://journals.sagepub.com/doi/abs/10.1177/1362480619896006>.

<sup>52</sup> ANGWIG, Julia; *et al* (2016). “Machine Bias” em *ProPublica*, 2016. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

muitos deles funcionam como dados substitutos que acabam remetendo à raça como um critério subjacente. Segundo a autora, embora muitas das perguntas feitas se concentram no histórico criminal do réu — como quantas vezes foi preso, quantas condenações teve, quanto tempo passou na prisão no último ano, etc. —, outras dezenas de perguntas não têm relação alguma com qualquer critério jurídico-legal, concentrando-se em saber se pessoas da família ou do círculo social do réu já foram presas, por exemplo (Sara Wachter-Boettcher, 2017, pp. 130-131)<sup>53</sup>.

Segundo a autora, a inclusão de indicadores desse tipo no modelo significa que, de acordo com os desenvolvedores, esses fatores estão relacionados ao “nível de risco” de uma pessoa, ignorando absolutamente o evidente aspecto de segregação racial e encarceramento em massa que permite que pessoas negras tenham mais chances de conhecer alguém que já tenha passado pelo sistema de justiça criminal. Além disso, o *software* é projetado para que as perguntas possam ser aplicadas aos réus de maneira flexível, conferindo ampla margem de liberdade aos agentes para coletar os dados de maneiras variadas e não padronizadas, sujeitas a diferenciações que podem reforçar ainda mais o viés racial do modelo (Sara Wachter-Boettcher, 2017, pp. 131-132)<sup>54</sup>.

Similarmente, departamentos de polícia também têm recorrido a modelos algorítmicos para policiamento preditivo e alocação de recursos, apresentando comportamentos injustos com alarmante frequência. Programas como o chamado PredPol<sup>55</sup> são utilizados para distribuir o contingente

---

<sup>53</sup> WACHTER-BOETTCHER, Sara (2017). *Technically Wrong: sexist apps, biased algorithms, and other threats of toxic tech*. W.W. Norton & Company, Nova York, Londres.

<sup>54</sup> WACHTER-BOETTCHER, Sara (2017). *Technically Wrong: sexist apps, biased algorithms, and other threats of toxic tech*. W.W. Norton & Company, Nova York, Londres.

<sup>55</sup> REY, Emmy (2020). “What PredPol is and what Pred Pol is not” em *PredPol*. Disponível em: <https://www.predpol.com/whatispredpol/>.

policial no perímetro urbano, concentrando-o nos locais onde crimes ocorrem com maior frequência. Para tanto, o modelo organiza as informações sobre crimes registrados pela polícia, analisando a ocorrência de um certo tipo de crime em uma área determinada e incorporando a análise a padrões históricos para prever quando e onde aquele mesmo crime “tende” a ocorrer novamente no futuro (Cathy O’Neil, 2016, p. 16)<sup>56</sup>.

No entanto, problemáticas surgem a propósito da diferenciação feita pelo modelo quanto à de gravidade dos crimes perseguidos:

Quando a polícia configura seu sistema PredPol, ela tem uma escolha. Ela pode se concentrar exclusivamente nos chamados crimes da Parte 1, que são os crimes violentos, geralmente denunciados, como homicídios, incêndios criminosos e agressões. Mas ela também pode ampliar o foco para incluir o mapeamento dos crimes da Parte 2, que são os crimes de vadiagem, mendicância agressiva e venda ou consumo de pequenas quantidades de drogas. Muitos desses crimes “incômodos” não chegariam sequer a ser registrados se um policial não estivesse lá para vê-los (Cathy O’Neil, 2016, p. 76-77, tradução nossa)<sup>57</sup>.

Conforme lembrado por O’Neil, os crimes “incômodos” costumam ser endêmicos e numerosos em muitos bairros pobres, de modo que a sua inclusão na análise acaba distorcendo os resultados produzidos pelo modelo ao criar um ciclo de retroalimentação que conduz o contingente policial sempre e cada vez mais para esses mesmos bairros, embora os crimes mais graves estejam muitas vezes localizados fora deles. Os crimes de menor potencial ofensivo acabam povoando os da-

<sup>56</sup> O’NEIL, Cathy (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown Publishers, Nova York.

<sup>57</sup> O’NEIL, Cathy (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown Publishers, Nova York.

dos em termos quantitativos, portanto, e produzindo novos dados que justificam mais policiamento, e assim por diante. Desse modo, mesmo que o modelo seja teoricamente neutro para fatores como raça e classe, o critério geográfico utilizado para informa-lo acaba funcionando novamente como um equivalente altamente eficaz para tais aspectos, especialmente em cidades amplamente segregadas, intensificando ainda mais sua vulnerabilidade à intervenção das agências de segurança pública (Cathy O’Neil, 2016, p. 77)<sup>58</sup>.

Ademais, inovações em outros tipos de tecnologias avançadas como *softwares* para reconhecimento facial, por exemplo, também têm contribuído para esse novo contexto em que promessas preditivas ameaçam cada vez mais o horizonte de possíveis. Baseado em tecnologias de computação visual (Tarcízio Silva, 2020, p. 433)<sup>59</sup>, o reconhecimento facial é um método utilizado para verificar a identidade de um indivíduo a partir dos dados biométricos de seu rosto. Tais tecnologias têm sido empregadas por autoridades policiais para fins de identificação de suspeitos em diversos países do mundo, a despeito das alarmantes taxas de erros que costumam apresentar especialmente na identificação de rostos de pessoas negras<sup>60</sup>.

<sup>58</sup> O’NEIL, Cathy (2016). *Weapons of math destruction: how big data increases inequality and threatens democracy*. Crown Publishers, Nova York.

<sup>59</sup> “De acordo com Tarcízio Silva, “[a] computação visual (*visual computing*) é um termo que engloba áreas da computação e iniciativas interdisciplinares dedicadas a construir sistemas para o entendimento automatizado de imagens e vídeos. Ou seja, busca-se fazer *softwares* e *hardwares* computacionais capazes de entender objetos, entidades e situações a partir de informação visual, como imagens ou *frames* de vídeo. Suas implicações são amplas, a indústria e mecatrônica até o marketing e publicidade” (SILVA, Tarcízio (2020). “Visão computacional e racismo algorítmico: branquitude e opacidade no aprendizado de máquina” em *Revista da Associação Brasileira de Pesquisadores/as Negros/as (ABPN)*, v. 12, n. 31, pp. 428-448. Disponível em <http://abpnrevista.org.br/revista/index.php/revistaabpn1/article/view/744?fbclid=IwAR1uD0ab3TQcKGAsEuaVemudHgBWE-Ep5aKldDplG-9tN59Jf8b3BuHt5kQ>).

<sup>60</sup> FACE Recognition (2017). *Eletronic Frontier Foundation*. Disponível em: <https://www.eff.org/pt-br/pages/face-recognition>.

Um estudo realizado pelo Instituto Nacional de Padrões e Tecnologia dos Estados Unidos (*National Institute of Standards and Technology* — NIST) constatou que os sistemas comerciais de reconhecimento facial para identificação de suspeitos utilizados pelos departamentos de polícia norte-americanos apresentam taxas consideravelmente expressivas de erros na identificação de rostos de pessoas afro-americanas e asiáticas, bem como nos rostos de mulheres e de pessoas com idade avançada. A análise baseou-se em mais de 18 milhões de fotos de cerca de 8,5 milhões de pessoas, e testou 189 algoritmos de reconhecimento facial de 99 desenvolvedores comerciais diferentes, incluindo empresas como a Microsoft e empresas de tecnologia biométrica como a Cognitec e chinesa Megvii. Os sistemas de outras empresas como Amazon, Apple, Facebook e Google não forneceram seus algoritmos para a realização dos testes<sup>61</sup>.

De acordo com Clare Garvie e Jonathan Frankle, as disparidades de precisão constatadas podem ser introduzidas involuntariamente em vários momentos no processo de concepção e implantação de um sistema de reconhecimento facial. O desenvolvedor do algoritmo pode programá-lo para focar em características faciais que são mais facilmente distinguíveis em algumas fenótipos raciais ou de gênero que em outros, por exemplo, de modo a favorecer a precisão na identificação de certos tipos de rostos. Além disso, os algoritmos “aprendem” a calcular a semelhança entre as imagens ao praticar combinações em conjuntos de dados de treinamento pré-existentes, de modo que a identificação será mais precisa quanto mais abundante for a base de da-

---

<sup>61</sup> NIST Study Evaluates Effects of Race, Age, Sex on Face Recognition Software (2020). *NIST*. Disponível em <https://www.nist.gov/news-events/news/2019/12/nist-study-evaluates-effects-race-age-sex-face-recognition-software>. Cf. também: FACE Recognition Vendor Test (2020). *NIST*. Disponível em <https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt>.; e SINGER, Natasha; METZ, Cade (2019). “Many facial-recognition systems are biased, says U.S. study” em *The New York Times*. Disponível em <https://www.nytimes.com/2019/12/19/technology/facial-recognition-bias.html>.

dos de treinamento para diferenciação de detalhes. Assim, ainda que os critérios nos quais o algoritmo se concentra sejam “neutros” com relação a raça ou gênero, isto é, tenham sido desenvolvidos para identificar detalhes e características fenotípicas diversificadas em termos raciais e de gênero, desproporcionalidades na quantidade de imagens poderão influenciar as suas taxas de precisão (Clare Garvie, Jonathan Frankle, 2016)<sup>62</sup>.

Além disso, ao invés de retornarem um único resultado com rigoroso grau de precisão, os *softwares* costumam apresentar uma listagem de combinações possíveis ranqueadas segundo a atribuição de pontuações de semelhança entre as imagens comparadas, de modo que os sistemas acabam favorecendo a ocorrência de falsos positivos, isto é, de reconhecimentos equivocados, incentivando, quando aplicadas à justiça criminal, investigações e acusações indevidas<sup>63</sup>. Diante desse contexto, o uso dessas tecnologias torna-se particularmente preocupante em contextos de manifesta desigualdade racial, como é o caso dos Estados Unidos:

Os afro-americanos têm pelo menos duas vezes mais chances de serem presos do que pessoas de qualquer outra raça dos Estados Unidos e, segundo algumas estimativas, até 2,5 vezes mais probabilidade de ser alvo da vigilância policial. [...] Em outras palavras, não apenas os afro-americanos são mais propensos a serem identificados incorretamente por um sistema de reconhecimento facial, mas também são mais propensos a serem inscritos nesses sistemas

<sup>62</sup> GARVIE, Clare; FRANKLE, Jonathan (2016). “Facial Recognition software might have a racial bias problem” em *Atlantic*. Disponível em <https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/>. Cf. também: GARVIE, Clare; et al (2016). “The perpetual line-up: unregulated police face recognition in America” em *Georgetown Law Center on Privacy and Technology*. Disponível em <https://www.perpetuallineup.org/conclusion>.

<sup>63</sup> FACE Recognition (2017). *Electronic Frontier Foundation*. Disponível em: <https://www.eff.org/pt-br/pages/face-recognition>.

e sujeitos ao seu processamento (Clare Garvie, Jonathan Frankle, 2016, tradução nossa)<sup>64</sup>.

Tecnologia semelhante é empregada em sistemas de visão computacional para reconhecimento de imagens em dispositivos *online*. O caso emblemático em que o Google Fotos “identificou”, através de sua ferramenta de visão computacional e etiquetamento automático, fotografias de afro-americanos como “*apes*” (“gorilas”) reforça o teor alarmante da problemática: na Internet e nos usos cotidianos da tecnologia, a discriminação algorítmica também está embutida em códigos computacionais e, cada vez mais, em tecnologias de “inteligência artificial” em processos de produção de sentido que refletem uma lógica corporativa de negligência intencional ou um imperativo de lucro que ganha dinheiro com o racismo e o sexismo (Safiya Noble, 2018, p. 12)<sup>65</sup>.

## Conclusão

Diante desse amplo contexto em que modelos algorítmicos têm sido cada vez mais utilizados de modo indiscriminado em processos importantes de tomada de decisão, alerta-se para um dano generalizado e sistematizado que funciona quase que de modo exclusivamente punitivo, tendo em vista que os algoritmos acabam alimentando-se mutuamente em um sistema pernicioso.

Em suma, pode-se dizer que os algoritmos ajudam a criar o entorno que justifica suas suposições: ao determinar o perfil de um indivíduo em função de suas circunstâncias e atributos, os modelos basicamente lhe impõem castigos

---

<sup>64</sup> GARVIE, Clare; FRANKLE, Jonathan (2016). “Facial Recognition software might have a racial bias problem” em *Atlantic*. Disponível em <https://www.theatlantic.com/technology/archive/2016/04/the-underlying-bias-of-facial-recognition-systems/476991/>.

<sup>65</sup> NOBLE, Safiya Umoja (2018). *Algorithms of oppression: how search engines reinforce racism*. New York University Press, Nova York.

com base em seus caracteres (quase sempre assumidos fora de contexto), operando contra seus interesses e causando-lhe danos efetivos das mais diversas ordens. Um círculo destrutivo alimenta a si mesmo de modo a aprofundar cada vez mais as injustiças sobre cujas premissas está antecipadamente fundamentado.

Essa dinâmica funciona com base em padronizações codificadas que distribuem arbitrariamente recompensas e penalidades, e alastram suas consequências para todos os campos da vida comum, prolongando as práticas punitivas que incidem cada vez mais sobre os sujeitos reduzidos a dado. Desse modo, sistemas algorítmicos transformam cada instância de decisão humana em verdadeiros vereditos, instaurando uma dinâmica punitiva que abrange expansivamente todas as instâncias de julgamento humano em que previsões algorítmicas podem ser utilizadas.