UNIVERSIDADE DE SANTA CRUZ DO SUL – UNISC
PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO – MESTRADO E DOUTORADO

LUCAS MORESCHI PAULO

DISCRIMINAÇÃO ALGORÍTMICA E PROTEÇÃO DE GRUPOS EM SITUAÇÃO DE VULNERABILIDADE PELA JURISDIÇÃO CONSTITUCIONAL: DEVER DE PROTEÇÃO ESTATAL À LUZ DA LEI MATERIAL DA PONDERAÇÃO

Santa Cruz do Sul
2026

Lucas Moreschi Paulo

DISCRIMINAÇÃO ALGORÍTMICA E PROTEÇÃO DE GRUPOS EM SITUAÇÃO DE VULNERABILIDADE PELA JURISDIÇÃO CONSTITUCIONAL: DEVER DE PROTEÇÃO ESTATAL À LUZ DA LEI MATERIAL DA PONDERAÇÃO

Tese apresentada ao Programa de Pós-Graduação em Direito – Mestrado e Doutorado – da Universidade de Santa Cruz do Sul – UNISC como requisito parcial para a obtenção do título de doutor.

Orientadora: Prof.ª Dr.ª Mônia Clarissa Hennig Leal (UNISC/BR)

Coorientador: Prof. Dr. Carlos Ignácio Aymerich Caño (UDC/ES)

Santa Cruz do Sul

2026

Lucas Moreschi Paulo

DISCRIMINAÇÃO ALGORÍTMICA E PROTEÇÃO DE GRUPOS EM SITUAÇÃO DE VULNERABILIDADE PELA JURISDIÇÃO CONSTITUCIONAL: DEVER DE PROTEÇÃO ESTATAL À LUZ DA LEI MATERIAL DA PONDERAÇÃO

_____
Dr.ª Mônia Clarissa Hennig Leal
Professora Orientadora – UNISC

_____
Dr. Carlos Ignácio Aymerich Caño
Professor Coorientador – UDC – Espanha

_____
Dra. Caroline Müller Bittencourt
Professora Examinadora – UNISC

_____
Dra. Suzéte da Silva Reis
Professora Examinadora – UNISC

_____
Dr. Anizio Pires Gavião Filho
Professor Examinador – FMP

_____
Dr. Bruno Lorenzetto
Professor Examinador – PUCPR

CIP - Catalogação na Publicação

Paulo, Lucas Moreschi
&&&Discriminação algorítmica e proteção de grupos em situação de vulnerabilidade pela jurisdição constitucional : Dever de proteção estatal à luz da lei material da ponderação / Lucas Moreschi Paulo. – 2026.
&&&279f.

&&&Tese (Doutorado em Direito) – Universidade de Santa Cruz do Sul, 2026.
&&&Orientação: Profa. Dra. Mônia Clarissa Hennig Leal.
&&&Coorientação: Prof. Dr. Carlos Caño.

&&&1. discriminação algorítmica. 2. vulnerabilidade. 3. proporcionalidade. 4. explicabilidade. 5. LGPD. I. Leal, Mônia Clarissa Hennig. II. Caño, Carlos. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da UNISC com os dados fornecidos pelo(a) autor(a).

À Tamires e à Helena, fontes de amor, inspiração e aprendizado constante; tudo é por vocês.

AGRADECIMENTOS

Agradeço, primeiramente, a Deus e à sua natureza, que em sua infinita sabedoria dita o ritmo da vida.

Agradeço à Tamires e à Helena, mulheres da minha vida que são razão e significação diária do meu seguir. Com vocês me renovo a cada dia. Só amor.

Agradeço à minha família, base estruturante sem a qual eu nada seria. Gratidão ao meu pai, Reginaldo, e à minha mãe, *Luciana*.

Agradeço à minha orientadora, Professora Doutora Mônia Clarissa Hennig Leal, em nome de quem agradeço à Universidade de Santa Cruz do Sul (UNISC), pela confiança em mim depositada para a condução de pesquisa em nível doutoral. Bem como por todo o carinho e atenção nestes 4 anos de programa. És mais do que uma orientadora acadêmica; gratidão!

Agradeço ao meu coorientador, Professor Doutor Carlos Aymerich Caño, em nome de quem agradeço à Universidade da Coruña (UDC), pela acolhida entre o final de 2024 e o início de 2025, e pelas valiosas reflexões para esta tese de doutoramento.

Agradeço à Fundação Escola Superior do Ministério Público (FMP), aos colegas do corpo docente, discente, gestão e apoio, e à Faculdade de Direito da UNISC, aos coordenadores, colegas e discentes, sem os quais não poderia ter almejado sucesso nas atividades empreendidas na Espanha para fins de completar o doutorado-sanduíche.

Agradeço à caridosa banca avaliadora, por acharem espaço em suas agendas, tanto para realizar a avaliação, quanto para contribuir com a pesquisa.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa na modalidade II, a qual foi fundamental para cursar o doutorado.

Agradeço ainda ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela concessão da bolsa para o doutorado sanduíche na Universidade da Coruña – Espanha, experiência fundamental para a construção desta tese de doutorado.

Agradeço também por poder e ter tanto a agradecer.

Where is the Life we have lost in living?
Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?
(Eliot, 1934, p. 7).

RESUMO

A pesquisa vincula-se ao Programa de Pós-Graduação em Direito da Universidade de Santa Cruz do Sul, na Área de Concentração em Direitos Sociais e Políticas Públicas e na Linha de Pesquisa Dimensões Instrumentais das Políticas Públicas, com ênfase no Controle Jurisdicional na efetivação de direitos fundamentais, ao articular constitucionalismo e direitos fundamentais com os desafios da inteligência artificial em um novo contexto discriminatório. A tese investiga a reprodução de discriminações estruturais por sistemas algorítmicos e propõe um arranjo normativo-hermenêutico para sua tutela, com ênfase na proteção de grupos em situação de vulnerabilidade. Questiona-se inicialmente em que medida e por quais parâmetros interpretativos a jurisdição constitucional deve proteger grupos em situação de vulnerabilidade contra discriminações algorítmicas, calibrando transparência/explicabilidade e segredo comercial-algorítmico? Para tanto, parte-se do mapeamento do fenômeno e de suas causas — seleção de variáveis, aprendizagem a partir de bases enviesadas, opacidades de modelo e governança — para demonstrar que a opacidade técnica não é acidente, mas expressão de racionalidade tecnocrática que demanda respostas jurídicas específicas. No plano constitucional, articula-se o dever estatal de proteção com a renovada dimensionalidade da igualdade e da não discriminação em ambiente de decisões automatizadas; no plano infraconstitucional, mobilizam-se os princípios e mecanismos da LGPD (art. 6º e art. 20) e a exigência de motivação adequada (CPC, art. 489, § 2º) como base para *standards* de prevenção, transparência e revisão. Utiliza-se, em um primeiro momento, de abordagem fenomenológica, através de revisão bibliográfica, para apreender, em sua inteireza, a emergência de uma vulnerabilidade jurídico-tecnológica nas estruturas decisórias digitais – evitando reduzi-la, aprioristicamente, a categorias jurídicas consolidadas; em seguida, adota-se método dedutivo, partindo de premissas normativo-constitucionais para derivar conclusões específicas acerca dos contornos jurídicos exigíveis da atuação jurisdicional diante de violações praticadas por agentes automatizados. Metodologicamente, adota-se a lei material da ponderação, a partir das elaborações de Robert Alexy, para construir um modelo decisório que gradua a intervenção jurisdicional, integrando a fórmula do peso a partir de sua preocupação com o grau de certeza empírica e normativa (C), alocando ônus probatórios dinâmicos em contextos

de assimetria informacional. O procedimento viabiliza a calibração da colisão explicabilidade/transparência *versus* segredo comercial-algorítmico por critérios reproduzíveis e auditáveis, permitindo acesso suficiente ao itinerário lógico-decisório sem desproteger os direitos da desenvolvedora. Resulta daí uma regra de decisão: quanto maior a interferência (I) e o peso abstrato (PeA) do princípio que cede, tanto maior deve ser o C exigido das premissas justificadoras e tanto mais ampla a abertura informacional, escalonadas por A1–A3. A contribuição reside em oferecer critérios operacionais e verificáveis para decisões judiciais sobre discriminação algorítmica — reduzindo assimetrias informacionais, fortalecendo a efetividade da igualdade material e consolidando uma cultura de justificação pública das escolhas automatizadas.

***Palavras-chave**:* discriminação algorítmica; vulnerabilidade; proporcionalidade; explicabilidade; LGPD.

ABSTRACT

This research is linked to the Postgraduate Program in Law at the University of Santa Cruz do Sul, within the Concentration Area of Social Rights and Public Policies and the Research Line “Instrumental Dimensions of Public Policies,” with an emphasis on Judicial Control in the enforcement of fundamental rights, by articulating constitutionalism and fundamental rights with the challenges of artificial intelligence in a new discriminatory context. The thesis investigates the reproduction of structural discrimination by algorithmic systems and proposes a normative–hermeneutic framework for their protection, with an emphasis on safeguarding groups in situations of vulnerability. It initially asks to what extent, and by which interpretive parameters, constitutional adjudication should protect groups in situations of vulnerability against algorithmic discrimination, calibrating transparency/explainability and trade-secret/algorithmic secrecy. To that end, it begins by mapping the phenomenon and its causes—variable selection, learning from biased datasets, model and governance opacity—to show that technical opacity is not an accident but an expression of technocratic rationality that calls for specific legal responses. At the constitutional level, it articulates the State’s duty to protect with the renewed dimensionality of equality and non-discrimination in an environment of automated decision-making; at the infra-constitutional level, it mobilizes the principles and mechanisms of the LGPD (Arts. 6 and 20) and the requirement of adequate reasoning (CPC, Art. 489, § 2) as the basis for standards of prevention, transparency, and review. First, a phenomenological approach is employed, through a literature review, to apprehend, in its entirety, the emergence of a legal-technological vulnerability within digital decision structures – avoiding, a priori, its reduction to established legal categories; next, a deductive method is adopted, starting from normative-constitutional premises to derive specific conclusions regarding the legally enforceable contours of judicial action in the face of violations committed by automated agents. Methodologically, the material law of balancing is adopted, drawing on Robert Alexy’s elaborations, to construct a decision model that grades judicial intervention, integrating the formula of weight in light of its concern with the degree of empirical and normative certainty (C), and allocating dynamic burdens of proof in contexts of informational asymmetry. The procedure enables the calibration of the collision between explainability/transparency and trade-

secret/algorithmic secrecy by means of reproducible and auditable criteria, allowing sufficient access to the logical-decisional itinerary without undermining the developer's rights. From this follows a decision rule: the greater the interference (I) and the abstract weight (PeA) of the principle that yields, the greater must be the C required of the justifying premises and the broader the informational openness, tiered by A1–A3. The contribution lies in offering operational and verifiable criteria for judicial decisions on algorithmic discrimination—reducing informational asymmetries, strengthening the effectiveness of substantive equality, and consolidating a culture of public justification for automated choices.

Keywords: algorithmic discrimination; vulnerability; proportionality; explainability; LGPD.

SUMÁRIO

1	**INTRODUÇÃO.....**	**14**
2	**DIREITO E VULNERABILIDADES ESTRUTURAIS: UM OLHAR CONCEITUAL ACERCA DAS POTENCIALIDADES DA DISCRIMINAÇÃO E DE SUA REPRODUÇÃO NO AMBIENTE VIRTUAL.....**	**24**
2.1	**Entre vulnerabilidades, minorias e grupos vulneráveis: uma diferenciação necessária.....**	**28**
2.2	**A discriminação estrutural e sua relação com o direito de igualdade e de não-discriminação.....**	**42**
2.3	**A renovada dimensionalidade da igualdade no contexto das decisões automatizadas em cenário cibernético e os desafios para sua concretização em casos de grupos vulneráveis.....**	**57**
3	**A ERA DA DISCRIMINAÇÃO ALGORÍTMICA: MAPEANDO O FENÔMENO E SUAS IMPLICAÇÕES DE ORDEM JUSFUNDAMENTAL.....**	**76**
3.1	**Evolução tecnológica, inteligência artificial e automação de decisões cotidianas.....**	**80**
3.2	**Discriminações algorítmicas: causas e potenciais de impulsionamento da discriminação estrutural e atingimento de grupos vulneráveis.....**	**95**
3.3	**O dever de proteção estatal e a discriminação algorítmica: tutela do Estado Democrático e Constitucional de Direito.....**	**112**
4	**TUTELA JURÍDICA DA DISCRIMINAÇÃO ALGORÍTMICA: EM BUSCA DE SENTIDO, SIGNIFICADO E PROTEÇÃO NORMATIVA..**	**129**
4.1	**Sentido normativo do direito acerca da tutela antidiscriminatória algorítmica.....**	**133**
4.2	**Da gramática técnica à razão pública justificada: bases para o controle judicial de discriminações algorítmicas.....**	**143**

4.3 Sentido normativo e diálogos institucionais: uma proposta de busca de sentido da tarefa interpretativa com as experiências nacionais e internacionais acerca do tratamento jurídico da discriminação algorítmica..... 161

5 A CONSTRUÇÃO DE UM MODELO DECISÓRIO RACIONAL BASEADO EM GRAUS DE EFETIVIDADE DOS DIREITOS FUNDAMENTAIS EM COLISÃO..... 180

5.1 A colisão central dos direitos fundamentais da tutela à discriminação algorítmica: transparência algorítmica e segredo algorítmico..... 184

5.2 Jurisdição constitucional frente à realidade das discriminações algorítmicas: a formatação do sentido interpretativo alinhado ao sentido normativo do Direito..... 199

5.3 A ponderação material como baliza interpretativa para uma gradação correlacional entre restrição e fruição de direitos fundamentais: o valor interpretativo-balizador de casos acerca da tutela de direitos de jurisdicionados pertencentes a grupos vulneráveis..... 219

6 CONCLUSÃO..... 244

REFERÊNCIAS..... 264

1 INTRODUÇÃO

> *Theirs was an invisibility cloak woven in equal measure to the rhetoric of the empowering web, the ability to move swiftly, the confidence of vast revenue streams, and the wild, undefended nature of the territory they would conquer and claim. They were protected by the inherent illegibility of the automated processes that they rule, the ignorance that these processes breed, and the sense of inevitability that they foster (Zuboff, 2019, p. 15).*

Tal como os troianos diante do cavalo deixado às suas portas, também nós nos deparamos com presentes cuja promessa de inovação esconde armadilhas. A narrativa homérica nos recorda que a aparência de dádiva pode ser apenas a forma mais engenhosa de captura. O algoritmo, revestido de neutralidade matemática, insere-se em nossas estruturas sociais como cavalo silencioso, trazendo em seu ventre a lógica de exclusão e de dominação que, uma vez aceita, se difunde de maneira quase irreversível.

A sedução da facilidade e da velocidade nos convida a crer que tudo se explica por si, mas há um modo de organizar dados e decisões que se faz discretamente – um manto que confunde eficiência, custo e técnica com tecnologia de ponta, neutralidade e direitos pessoais. A linguagem do cálculo começa a tomar a frente nas interações humanas, enquanto a sua ilegibilidade, tomada como sinal de progresso, vai convertendo o conveniente em natural. Os processos automáticos encobrem opções de desenho de poder sobre a vida nas redes, sejam sociais, informacionais ou de tomada de decisão automatizada simplesmente. Não há silêncio por desconhecimento. A opacidade pode ser combatida, mas não faz parte do projeto, nem das prioridades. Esse agendamento é papel da instância jurídica - bem como sua tutela. Eis o primeiro chamado do Direito: reabrir a clareira da justificabilidade publicizada e recolocar a participação e a racionalidade no centro do *design* algorítmico. Explicabilidade, auditabilidade e participação devem deixar de ser ornamentos e devem converter-se condições de validade e legitimidade das operações algorítmicas. Eis o horizonte.

A crescente incorporação de sistemas algorítmicos em decisões de alto impacto desloca categorias processuais tradicionais e intensifica a tensão entre racionalidade técnica e racionalidade jurídica. A opacidade inerente a muitos modelos de aprendizagem de máquina — frequentemente justificada por segredo comercial ou complexidade — compromete o contraditório e a motivação pública das decisões. Em ambiente de assimetrias epistêmicas, sobretudo quando grupos multiplamente e interseccionalmente vulneráveis são os destinatários dos efeitos, as exigências de explicabilidade, auditabilidade e participação deixam de ser vetor técnico e torna-se condição de validade constitucional.

Eliot (1934), como na epígrafe desta tese, lapida o diagnóstico de uma modernidade que perdeu o eixo teleológico ao confundir meios e fins. A tríplice interrogação revela um processo de degradação semântica – da vida à sabedoria, da sabedoria ao conhecimento, e por fim para a mera informação – no qual cada etapa marca o esvaziamento de sentido do que a antecede. O pensamento recusa a mitologia da eficiência como sinônimo de verdade e reinstaura o problema próprio do Direito enquanto uma instância de busca e construção de razões públicas, aqui no presente caso razões orientadas por fins constitucionais firmes na defesa contra discriminações algorítmicas. Há que se repudiar a idolatria da informação acelerada, automatizada, pois nela não se resguarda necessariamente uma verdade ou uma forma satisfatória de vida; o que ofende a dimensão jurídica.

Com esse horizonte em vista, vivencia-se, na atual quadra histórica, profundas e constantes transformações impulsionadas pelo desenvolvimento tecnológico, especialmente no que tange à incorporação de sistemas automatizados baseados em inteligência artificial e aprendizado de máquina em diversas esferas da vida social. A chamada quarta revolução industrial, marcada pela integração entre dados massivos, processamento computacional avançado e tomada de decisão autônoma, vem remodelando estruturas econômicas, relações sociais e, inevitavelmente, os fundamentos do próprio ordenamento jurídico. O ambiente digital contemporâneo impõe à teoria e à prática jurídicas uma revisão crítica dos conceitos clássicos de responsabilidade, imputação, liberdade, igualdade e dignidade humana, exigindo que se reexaminem os contornos da proteção constitucional dos direitos fundamentais à luz de uma realidade marcada por assimetrias informacionais e opacidades decisórias. Aqui denota-se a insuficiência do paradigma protetivo

tradicional diante da renovada realidade da discriminação algorítmica, ressaltando a urgência de conceber novas necessidades de tutela para o ser-usuário-titular em relação à desinformação e o tratamento ilegítimo, além da reparação de danos.

O nível tecnológico da Internet chegou a um grau de sinergia com o ser humano, com o funcionamento da mente humana e com alcance de inter-relacionamento nunca antes visto. Hoje, a vida acontece em rede, muito mais do que no mundo físico, e mesmo o que ocorre no mundo físico é influenciado por processamento de dados, georreferenciamento, geoposicionamento e outras derivações digitais. Assim, é através da Internet que se compartilham as experiências e os principais acontecimentos do cotidiano, onde se lê, se compra, onde hoje é realizada a aplicação para vagas de emprego, para encontros afetivos e amorosos, com amigos ou pretendentes, onde projetamos o nosso futuro, simulamos financiamentos, encontramos apartamentos para alugar ou comprar, onde, verdadeiramente, existimos na dimensão mais humana e sensível do nosso ser humano.

Nesse contexto, observa-se o crescente emprego de sistemas algorítmicos em decisões que impactam diretamente a esfera jurídica de indivíduos e grupos, como na triagem de currículos, na concessão de crédito, no policiamento preditivo, na administração pública digital, entre tantos outros. Ainda que estas tecnologias sejam frequentemente apresentadas como neutras e objetivas, os dados de entrada e as lógicas estatísticas que orientam seu funcionamento podem reproduzir – ou até intensificar – discriminações estruturais preexistentes. A discriminação algorítmica, enquanto fenômeno jurídico-normativo emergente, manifesta-se pela reprodução sistemática de vieses que, embora muitas vezes invisíveis ou naturalizados, afetam desproporcionalmente sujeitos em situação de vulnerabilidade histórica e social, como minorias raciais, étnicas, de gênero, socioeconômicas e outras.

Para que as aplicações via redes sirvam a uma experiência personalizada ou mais efetiva, o seu uso é baseado na coleta e no processamento de dados para a garantia da qualidade da entrega objetiva de um resultado (*output*) requerido. Para tais aplicações funcionarem, são utilizadas redes neurais digitais a partir de grandes bases de dados, pessoais e não-pessoais, que fornecem insumo para o processamento de complexas etapas de cálculos e verificações estatísticas que, desejavelmente, encontram o melhor *match*, a melhor rota, o melhor candidato ao

emprego, etc.; esperamos que a máquina robótica e cibernética nos diga o que é melhor, isto é, (hoje em dia) o que é o mais provável que eu queira (talvez daquilo que precise) desejar. O risco reside em o*utputs* que entreguem respostas ilicitamente discriminatórias, seja pela intencionalidade do programador, ou, mais comum, aquém da vontade dos programadores, quando da automação dos processos de algoritmização e de automação da tomada de decisão, os resultados obtidos sejam injustificadamente danosos.

No crescente cenário de constantes transformações e evoluções tecnológicas, onde o uso centralizado de tomada de decisões automatizadas permeia diversos aspectos da vida cotidiana, é necessário atentar-se aos critérios interpretativo-jurídicos que precisarão ser adotados pela juridicidade à proteção dos direitos humanos e fundamentais de grupos vulneráveis e minorias no ambiente digital. Diante de um paradigma constitucionalista de defesa de direitos pessoais, coletivos e difusos, a necessidade inescapável hodierna é a do acerto de parâmetros interpretativos que efetivamente protejam os direitos fundamentais das novas relações cibernéticas e, ao assim fazer, garantam a tutela de todos os direitos em conflito em graus proporcionais de proteção, preservação e fruição; fala-se, aqui, especialmente, dos direitos fundamentais associados à proteção em relação à realidade algorítmica, em contraste com os direitos das empresas desenvolvedoras e detentoras de tais aplicações tecnológicas.

Para chegar-se a tanto, será da lógica do percurso avaliar a discriminação algorítmica no contexto das transformações tecnológicas contemporâneas, com foco na proteção dos direitos humanos e fundamentais em ambiente digital, com especial atenção à proteção de grupos vulneráveis, assim entendidos aqueles historicamente negligenciados e cujo reflexo dessa negligência se potencializa na ausência de adequada representatividade datológica efetiva, o que fica escancarado quando avaliado pelos resultados das aplicações algorítmicas – abastecidas com dados que, não raras vezes, meramente tangenciam a realidade social que infielmente retratam.

Serão examinadas as implicações jurídicas e os critérios interpretativo-jurídicos aplicáveis à discriminação algorítmica, examinando o papel do Estado-juiz na proteção dos direitos no tocante à temática, avaliando o papel da jurisdição constitucional na interpretação das leis e normas relacionadas. A delimitação normativo-espacial abrange contextos jurídicos e digitais brasileiros, sem prejuízo de

se colherem experiências internacionais para qualificar o debate (cujo insumo material está sendo produzido constantemente).

Há, no entanto, que se tomar cautela e preocupação para com o potencial discriminatório que os algoritmos podem perpetuar, limitando e desumanizando o acesso a bens e serviços, a oferta de crédito, o acesso a vagas de emprego, o risco de ser considerado mais ou menos potencialmente lesivo pela polícia ou pelo sistema judiciário penal, etc.

A discriminação algorítmica, como se verá melhor em tópico apropriado, ocorre, em linhas gerais, quando o algoritmo, ao processar os dados, reproduz preconceitos existentes na sociedade – refletidos em sua "amostragem" de dados – e sem considerar o contexto específico ou as particularidades individuais em que cada um daqueles dados foi coletado. Isso pode levar a decisões injustas e à perpetuação de desigualdades, especialmente em relação a grupos historicamente marginalizados.

Por isso, é necessário que os algoritmos sejam desenhados, construídos, treinados e operados com uma transparência proporcional ao risco, análise crítica dos dados utilizados, para garantir que não haja viés ou discriminação. Além disso, é importante que haja uma regulamentação e supervisão adequadas para o uso de algoritmos em decisões importantes, a fim de garantir a proteção dos direitos individuais e evitar danos de difícil reparação à sociedade e ao senso cultural que vai se formatando. O objeto de análise, portanto, é o de identificar como problema de pesquisa o impacto jurídico e social dos algoritmos discriminatórios utilizados em plataformas de inteligência artificial de amplo acesso, bem como se esses algoritmos podem afetar os direitos individuais, a igualdade de oportunidades e a justiça social e, neste caso, quais as reais possibilidades de atuação da jurisdição constitucional para a adequada interpretação e aplicação do direito.

Acerca dos possíveis impactos das redes neurais estruturadas por algoritmos e por bases de dados com potencial discriminatório, importa analisar as possibilidades interpretativas de aplicação do direito e da jurisdição constitucional para tutelar casos de interferência em direitos e garantias fundamentais decorrentes de decisões automatizadas e de discriminações algorítmicas. Além disso, pretende-se analisar as razões por trás da discriminação algorítmica, bem como situá-la no atual mundo e lançar sobre ela olhar jurídico-constitucional de viés protetivo, com uma lógica de proteção do Estado Democrático e Constitucional de Direito. Isso se dá pela

importância que a epistemologia tem na formação e no delineamento das proposições jurídicas que devem ser aquilatadas para uma mais efetiva proteção judicante, seja por iniciativa do Poder Legislativo, seja pela atuação da Corte Constitucional ante lacunas e anomias no trato adequado de prejuízos a direitos fundamentais que, embora sejam facilmente encaixáveis em categoriais jurídicas existentes, necessitam de melhor contorno em sua realidade jurídico-protetiva em virtude do novo meio específico em que ocorrem e pelos motivos nem sempre autoevidentes ou transparentes.

Impõe-se refletir sobre o papel da jurisdição constitucional na mediação dos conflitos oriundos do uso de algoritmos e na salvaguarda dos direitos fundamentais potencialmente ameaçados. É preciso compreender que a atuação dos tribunais constitucionais não se limita à repressão de lesões já concretizadas, mas compreende também a função proativa de construir sentidos normativos que orientem a interpretação e a aplicação do direito em contextos novos e disruptivos. Em especial, torna-se necessário identificar critérios jurídico-interpretativos que permitam modular o grau de proteção de determinados direitos fundamentais, reconhecendo que nem sempre será possível garantir a máxima efetividade de todos eles simultaneamente, sobretudo quando estiverem em colisão.

A proposta deste trabalho parte, portanto, da necessidade de se investigar quais parâmetros de interpretação podem ser concebidos para orientar a jurisdição constitucional na proteção dos direitos humanos e fundamentais de grupos vulneráveis no ambiente digital. Para tanto, o percurso analítico inicial consiste na análise crítica da evolução tecnológica recente, com ênfase nas transformações operadas pelo uso de algoritmos e na identificação dos desafios jurídicos que daí decorrem, em especial os relacionados à opacidade, à ausência de accountability e à dificuldade de rastreamento de decisões discriminatórias. Em seguida, busca-se examinar o papel hermenêutico da jurisdição constitucional diante dessas transformações, especialmente no tocante à concretização diferenciada dos direitos fundamentais, observando-se a necessidade de reconhecer graus variados de eficácia e proteção, conforme a posição do sujeito no espectro das vulnerabilidades.

Nesse processo, será proposta uma estrutura interpretativa que permita à jurisdição constitucional integrar, de forma sistemática, a análise de impacto algorítmico como elemento essencial da ponderação de direitos em casos de colisão.

Tal estrutura deve considerar a exigência crescente de transparência, auditabilidade e explicabilidade algorítmica como condições mínimas de legitimidade constitucional do uso dessas tecnologias. Em outras palavras, propõe-se que a interpretação constitucional contemporânea se abra ao reconhecimento de que a proteção de direitos não pode mais prescindir de uma avaliação técnica e jurídica prévia dos impactos que sistemas automatizados produzem sobre grupos específicos. Isso inclui a incorporação de exigências legais – que em muitos países já vêm sendo discutidas ou implementadas – como a mensuração prévia de riscos (*risk assessment*), a explicabilidade by design e a adoção de mecanismos de governança algorítmica voltados à proteção de direitos fundamentais.

Mais do que formular diretrizes abstratas, esta proposta visa contribuir para a construção de um modelo interpretativo que permita respostas casuísticas coerentes e proporcionais, aptas a solucionar, de modo racional, os dilemas que se instauram diante da colisão entre direitos fundamentais e interesses legítimos diversos – como os comerciais, os relacionados à inovação tecnológica ou à eficiência administrativa. Reconhece-se, com isso, que o discurso jurídico-constitucional precisa ser tensionado à altura da complexidade dos problemas que emergem na era dos algoritmos. A jurisdição constitucional é chamada, assim, a não apenas reagir a violações já consumadas, mas a construir um quadro normativo capaz de antecipar riscos e proteger, de modo eficaz e proporcional, os sujeitos mais expostos aos efeitos discriminatórios das novas tecnologias.

Portanto, essa investigação busca, ao longo de seu desenvolvimento, articular uma reflexão teórico-prática acerca da discriminação algorítmica e seus efeitos sobre o regime jurídico dos direitos fundamentais, propondo caminhos interpretativos normativo-problematizadores que forneçam suporte à atuação da jurisdição constitucional frente aos desafios do presente e às exigências que se avizinham. Trata-se, em última análise, de afirmar a centralidade da dignidade humana e da igualdade material como vetores fundamentais para a reconstrução hermenêutica do direito constitucional em tempos de transformações tecnológicas profundas.

Sem pretensão de esgotar a constelação de bens jusfundamentais afetados pelo ambiente digital, este recorte assume uma função heurística e metodológica: oferecer um eixo de controle verificável para problemas que, na prática, tendem a se ocultar sob promessas de neutralidade técnica.

É imperativo realizar uma proposta interpretativa alinhada à jurisdição constitucional e ao dever de proteção estatal, visando à proteção de direitos fundamentais, sobretudo de pessoas e grupos vulneráveis, em aplicações algorítmicas automatizadas, considerando diferentes níveis de intervenção restritiva justificada em direitos de agências e empresas que disponibilizam tais aplicações no mercado, fundamentada e respaldada pelas regras e princípios da interpretação constitucional do escrutínio da proporcionalidade a partir do modelo de matriz alexyana, sobretudo a partir da dimensão substancial dos subtestes, com enfoque especial na lei material da ponderação.

A contribuição aqui proposta não se limita à aplicação genérica da proporcionalidade; consiste na formalização epistêmico-procedimental de um protocolo de decisão que conjuga a lei material da ponderação com a variável C (certeza epistêmica) e com um regime de abertura graduada (A1–A3), o que só poderá ser densificado e verificado em casos concretos. A originalidade está em (i) operacionalizar a variável C por critérios observáveis – compondo da dimensão fática (robustez técnico-probatória e perícia algorítmica) e da dimensão normativa – e adotando o critério da verificação da última etapa do teste da proporcionalidade (a ponderação ou proporcionalidade em sentido estrito) apenas com o apoio do acervo probatório suficiente para tanto, o que se revela de caso a caso; e em (ii) institucionalizar a correspondência entre subtestes da proporcionalidade e evidências exigíveis, de modo a produzir uma regra de decisão reproduzível – ainda que necessariamente adaptável de modo tópico – e controlável intersubjetivamente. Com isso, a explicabilidade e o segredo comercial deixam de ser polos abstratos e passam a integrar um arranjo normativo-hermenêutico verificável, apto a calibrar o ônus argumentativo do Estado e dos agentes econômicos em casos de discriminação algorítmica, assim sendo uma via apta a converter convicções técnicas em razões públicas.

O arranjo decisório aqui defendido decorre simultaneamente da lei material da ponderação e das fontes positivas que vinculam a motivação judicial, a tutela de dados e a igualdade – bem como tenta antever e contribuir com a forma de tratamento das IAs e algoritmos em contexto pátrio –. A ponderação, tal como reconstruída, fornece o esqueleto normativo para aferição de adequação, necessidade e proporcionalidade em sentido estrito, da preterição do segredo algorítmico à preferência da

transparência algorítmica. Abre-se a necessidade para a ordenação em graus de abertura sob sigilo quando necessário e proporcional. O protocolo A1–A3 traduz essas exigências em níveis graduais de acesso e justificação, modulados por intensidade de impacto, vulnerabilidade estrutural e opacidade residual. Resulta daí uma vinculação jurídico-procedimental: não se trata de mera diretriz prudencial, mas estruturar uma matriz interpretativa específica – ainda que *a priori* e necessitando de densificação *in casu* – da decisão que deve harmonizar direitos fundamentais em colisão, assegurando a transparência acerca da racionalidade empregada.

Nesse cenário, a presente investigação será conduzida, em um primeiro momento, considerando a natureza do fenômeno em análise – a saber, a emergência de um novo tipo de vulnerabilidade jurídico-tecnológica que desafia os paradigmas clássicos de imputação e proteção no direito constitucional. A escolha desse método justifica-se pela necessidade de apreender, em sua inteireza, o modo como a discriminação algorítmica se manifesta nas estruturas decisórias digitais e pelas consequências reais que produz sobre os corpos vulneráveis, permitindo uma abertura ao fenômeno sem reduzi-lo aprioristicamente a categorias jurídicas consolidadas. Em seguida, adotar-se-á o método dedutivo, partindo-se de premissas normativas e teóricas mais amplas – como os deveres estatais de proteção e os princípios estruturantes da Constituição de 1988 – para alcançar conclusões específicas quanto aos contornos jurídicos exigíveis da atuação jurisdicional frente a violações praticadas por agentes automatizados.

Trata-se, portanto, de uma pesquisa de natureza aplicada, que objetiva oferecer soluções normativo-hermenêuticas a um problema jurídico concreto e atual, com abordagem qualitativa, voltada à compreensão e interpretação dos sentidos normativos envolvidos, e com objetivos descritivos e exploratórios, na medida em que busca identificar, organizar e analisar criticamente os parâmetros constitucionais existentes (ou sua ausência) diante de situações complexas de discriminação algorítmica.

O método de procedimento adotado será analítico, com enfoque na decomposição e exame aprofundado de textos legais, jurisprudência constitucional e doutrina especializada, com eventual recurso ao estudo comparado nos pontos em que experiências estrangeiras contribuírem para o adensamento crítico do problema. A pesquisa também se valerá de uma revisão bibliográfica sistemática, especialmente

no que diz respeito às contribuições oriundas das teorias da argumentação jurídica, das teorias da jurisdição constitucional e da dogmática dos direitos fundamentais, com ênfase nas formulações de matriz alexyana acerca da proporcionalidade e da ponderação como técnica decisória em contextos de colisão de direitos.

A escolha metodológica pela lei material da ponderação, com foco nos aspectos substanciais dos subprincípios da adequação, necessidade e proporcionalidade em sentido estrito, permite não apenas avaliar a legitimidade das interferências promovidas por agentes automatizados no âmbito dos direitos fundamentais, mas também construir um modelo interpretativo racional e principiológico que oriente a jurisdição constitucional em decisões futuras. Essa proposta metodológica busca garantir, em última instância, que a hermenêutica constitucional seja sensível às assimetrias tecnológicas contemporâneas e que a dignidade humana – especialmente dos sujeitos historicamente marginalizados – permaneça como fundamento e limite da arquitetura decisória do Estado.

2 DIREITO E VULNERABILIDADES ESTRUTURAIS: UM OLHAR CONCEITUAL ACERCA DAS POTENCIALIDADES DA DISCRIMINAÇÃO E DE SUA REPRODUÇÃO NO AMBIENTE VIRTUAL

> *Enquanto existir, pelo fato das leis e dos costumes, uma danação social criando artificialmente, em plena civilização, infernos, e complicando com uma fatalidade humana o destino que é divino; (...) enquanto, em certas regiões, a asfixia social for possível; em outros termos, e de um ponto de vista mais extenso ainda, enquanto houver sobre a terra ignorância e miséria, livros da natureza deste poderão não ser inúteis (Hugo, 2025, p. 11).*

Em sua denúncia pungente da exclusão, Victor Hugo desloca a miséria do plano moral para o estrutural – não como desvio ocasional, mas como produto autorizado por leis e costumes. A miséria, nas páginas d'*Os Miseráveis*, não nasce de um fracasso individual, mas daquilo que o Direito permitiu, ignorou ou legitimou. O inferno social de que fala Hugo é mais do que metáfora: é uma arquitetura concreta, erguida com as vigas da omissão estatal e os tijolos de uma justiça que, ao pretender neutralidade, consagra a indiferença. Esta leitura da literatura francesa antecipa o objeto deste capítulo: compreender a desigualdade como fenômeno estrutural e a função do Direito – ou sua ausência – como agente de perpetuação dessa ordem.

Jean Valjean roubou um pão. Por esse gesto de desespero – mais sobrevivência do que transgressão – foi lançado a um abismo jurídico e social que, longe de cessar com o cumprimento da pena, apenas se intensificou. Não é o crime que o condena eternamente, mas a forma como o sistema recusa-o: não lhe permite ser outro, não reconhece sua tentativa de reconstrução, não lhe devolve o direito de existir como sujeito. Jean Valjean não é apenas um personagem trágico: ele é síntese da crueza estrutural de um sistema que pune sem compreender, que sentencia sem escutar, que iguala sem considerar os desníveis do ponto de partida. Seu gesto inaugural – o furto de um pão – não é o início da delinquência, mas o ápice de uma privação. Não é apenas início de sua punição, mas do ciclo de sua vulnerabilidade.

É nesse gesto que *Os Miseráveis* toca o nervo da questão jurídica: a miséria não é um evento isolado, mas um ciclo. E quando o Direito opera como reforçador

desse ciclo – ao punir desproporcionalmente, ao ignorar contextos, ao formalizar desigualdades sob o pretexto de tratar todos igualmente – ele não apenas falha, mas agrava o sofrimento que promete corrigir.

Este capítulo analisa a persistência da desigualdade sob o manto da neutralidade normativa. Argumenta-se que a igualdade formal, ao ignorar as assimetrias estruturais que atravessam raça, classe, gênero e território, transforma-se em instrumento de reprodução da exclusão. Mais do que ausência de proteção, trata-se aqui de uma violência estrutural consentida, que ignora a historicidade da desigualdade e silencia a urgência de uma resposta jurídica que compreenda os sentidos materiais da justiça.

A tarefa, portanto, é deslocar o foco: da aparência de isonomia para a substância da equidade. Como no caso de Valjean, a biografia dos vulneráveis não pode ser lida à margem do Direito – deve ser ponto de partida para a reconstrução do seu próprio sentido. O bom combate do Direito no presente esforço é o do combate à persistência da estrutural e (des)estruturante desigualdade marginalizante, miserável e discriminatória.

A persistência dessa desigualdade – alicerçada por séculos em estruturas materiais, institucionais e simbólicas – não se encerra nas formas clássicas da exclusão. Ao contrário: é precisamente quando se multiplicam as promessas de neutralidade tecnológica e objetividade algorítmica que novos mecanismos de reprodução da vulnerabilidade se erguem, agora sob a aparência da técnica. A linguagem está em constante mutação e adaptação, mas o conteúdo persiste; as ferramentas se sofisticam, mas os resultados ainda atingem corpos historicamente marcados pela desigualdade. Nesse sentido, é indispensável ampliar a lente de análise: da miséria como dado estrutural analógico para sua transmutação em discriminações de novo tipo – automatizadas, replicadas, opacas – que desafiam não apenas o Direito em sua função clássica, mas sobretudo o papel ativo da jurisdição constitucional na proteção dos grupos vulnerabilizados.

A evolução tecnológica e a crescente utilização de algoritmos em diversas esferas da sociedade têm trazido à tona importantes discussões sobre os impactos dessas ferramentas na perpetuação de discriminações estruturais. Pretende-se vislumbrar os pontos de arreio da necessidade de uma postura proativa da atividade jurisdicional do Estado no cumprimento do seu dever de proteção constitucional,

especialmente no que tange às discriminações algorítmicas, sobretudo daquelas pessoas atingidas que se identificam com grupos social e historicamente marginalizados pela própria sociedade analógica. A partir de uma abordagem crítica, propõe-se a reflexão sobre quais são os riscos e desafios impostos pela discriminação estrutural que se perpetuam e se amplificam por conta dessas tecnologias. Aliado a isso, necessário se faz observar a importância da tutela jurídica para garantir a igual consideração das dignidades, individuais e coletivas, dentro de uma dupla sociedade, a sociedade de fato (do mundo real), bem como da sociedade refletida pelas telas, a sociedade algorítmica e da informação.

Trata-se, portanto, de uma investigação acerca dos meandros da discriminação estrutural – analógica – a partir do seu reflexo na discriminação algorítmica atual e dos paralelos com uma efetiva e proativa tutela diferenciada pelo Poder Judiciário, que deve não apenas afirmar os direitos fundamentais na atual conjuntura social, mas também servir de broquel para decisões automatizadas que não apenas discriminam, mas solidificam, normalizam e agravam tais discriminações. Nesse sentido, primeiramente é necessário verificar a forma pela qual ocorre a discriminação estrutural na atual sociedade, o que chama-se do “modo analógico de discriminar”, para, então, atravessar um paralelo à era da digitalização e verificar a forma como a tecnologia transformou o modo de ocorrência das discriminações, ainda que não tenha se furtado de alterar sua essência, tendo mudado sua forma e seu acontecimento que, agora, também se dá de modo inumano, isto é, de modo automatizado por decisões robótico-algorítmicas. Após esses dois momentos de investigação, será necessário avaliar o modo jurisdicional que a atualidade da expertise jurídica alcança na tutela dos direitos fundamentais, entre o passado – e o ainda presente – da discriminação estrutural, com a novidade a ser enfrentada, da discriminação algorítmica, que nasce da discriminação estrutural e que, potencialmente, a fortalece. Eis o desafio ao Estado-Juiz atual.

A vulnerabilidade está intrinsecamente relacionada ao princípio da igualdade, sendo esse não somente um direito fundamental pertencente a todos que estejam no país, como também um elemento formador da democracia constitucional. Nessa perspectiva, identificou-se que a igualdade objetivada é aquela que pode ser materializada entre todos, sem distinções, o que se transforma em adaptações de igualdade, ou seja, tratamento igual aos iguais e desigual aos desiguais, na medida

das suas desigualdades. É nesse viés que se localiza a necessidade de um tratamento diferenciado aos indivíduos de grupos vulneráveis, visto que a disparidade histórica e material impedem que alcancem com a mesma facilidade os seus direitos e as oportunidades oferecidas.

A identificação de que existem grupos menos favorecidos e que possuem barreiras de acesso justifica as previsões legais de proteção desses grupos específicos. Portanto, percebe-se que o direito à igualdade exige do Estado obrigações positivas e negativas para sua perfectibilização. Com relação à atuação do Estado no âmbito de obrigações positivas, de efetivar o direito à igualdade dos grupos vulneráveis, tem-se a necessidade de combate à discriminação estrutural, que é o tratamento que cria barreiras a indivíduos que são tratados injustamente em razão de estereótipos enraizados na sociedade, que podem surgir de elementos culturais ou históricos.

Verifica-se o dever do Estado de criar condições que reduzam essas desigualdades, como a implementação de ações afirmativas, ainda que em caráter excepcional e transitório, a fim de equilibrar esse desnível em que se coloca os grupos vulneráveis. Assim como, a necessidade de atuação legislativa e na sua ausência ou insuficiência, que o Poder Judiciário possa atuar em prol desses direitos. Deve existir uma ação conjunta entre os poderes da nação para proteger os indivíduos das agressões aos seus direitos, principalmente diante do nosso contexto social em que as tecnologias reproduzem os exemplos e conceitos humanos.

No âmbito da automatização da discriminação estrutural, ou seja, da reprodução de tratamentos discriminatórios a partir de algoritmos, obtém-se que os algoritmos funcionam a partir de instruções pré-estabelecidas, dependendo fundamentalmente de dados. Pois como visto, não existe inteligência artificial sem processamento de dados. Esse é o cerne da questão, se os dados utilizados forem estritamente baseados em padrões históricos ou não representativos, o resultado obtido desse processo de decisão automatizada será discriminatório, perpetuando padrões injustos e prejudiciais. Por essa razão, identifica-se a necessidade de que esses sistemas atuem de forma mais transparente e que a responsabilidade sobre esses resultados seja determinada.

2.1 Entre vulnerabilidades, minorias e grupos vulneráveis: uma diferenciação necessária

A compreensão das desigualdades sociais exige um olhar atento às múltiplas formas pelas quais certos grupos são sistematicamente posicionados em desvantagem no interior da estrutura social. A sociedade brasileira, profundamente marcada por disparidades econômicas, culturais e simbólicas, demanda uma análise que reconheça a complexidade das interações sociais e institucionais que produzem e perpetuam tais desigualdades. Nesse contexto, torna-se imprescindível distinguir entre as noções de vulnerabilidades, minorias e grupos vulneráveis, uma vez que cada categoria carrega especificidades ontológicas e normativas que devem ser levadas em conta na elaboração de políticas públicas e na interpretação constitucional orientada à promoção da igualdade substancial. Tal distinção ganha relevo diante do desafio de articular, juridicamente, o princípio da igualdade material com as formas diversas e, por vezes, invisibilizadas de discriminação que operam tanto em decisões intencionais quanto em práticas sociais aparentemente neutras.

Antes de avançar para o regime jurídico da igualdade, convém explicitar as razões morais que interditam a discriminação – razões que também informam os deveres jurídicos de prevenção, mitigação e responsabilização no ambiente algorítmico, para mapear e entender o que é necessário de se vislumbrar quanto à discriminação estrutural.

As discussões sobre a discriminação algorítmica como fenômeno jurídico relevante na era digital inserem-se no debate mais amplo sobre a discriminação estrutural e sobre a fundamentação moral da própria proibição de discriminar. Não se trata apenas de um problema técnico de desenho de sistemas, mas de um modo de reprodução – por meios automatizados – de assimetrias históricas que atravessam o tecido social e jurídico. Em outras palavras, a classificação automatizada opera no plano das consequências – decisões sobre crédito, trabalho, saúde, segurança pública – mas é tecida, a montante, por escolhas de modelagem, curadoria de dados e métricas de desempenho que podem reatualizar hierarquias de poder e exclusões supostamente superadas. Nesse sentido, a chave analítica é dupla: compreender a discriminação algorítmica como expressão contemporânea da desigualdade estrutural e, simultaneamente, fundamentar a proibição de discriminar em razões morais

robustas que informem deveres jurídicos positivos de prevenção, mitigação e responsabilização.

As formulações de Kant, Mill e Rawls fornecem arcabouço normativo para a censura moral a ações – e arranjos institucionais – discriminatórias[1]. Sob a perspectiva kantiana, a discriminação é moralmente inadmissível porque instrumentaliza pessoas e viola a dignidade e a autonomia da vontade prática, rebaixando sujeitos a meros meios para fins alheios (Kant, 2009). Em sistemas algorítmicos, isso se expressa quando indivíduos são reduzidos a perfis probabilísticos opacos, sem condições de participação ou contestação efetiva: a agência é deslocada – e com ela a possibilidade de autogoverno – para arquiteturas técnicas inacessíveis ao leigo, o que afronta a exigência de tratar cada pessoa como fim em si. A dignidade, assim compreendida, converte-se em critério de desenho e de controle: proíbe classificações que estigmatizam, que denegam justificativas compreensíveis e que impedem o exercício de direitos correlatos de defesa e revisão.

Sob o ponto de vista utilitarista, a discriminação é moralmente errada porque reduz o bem-estar agregado, produz ineficiências alocativas e gera externalidades negativas que se acumulam sobre grupos já vulnerabilizados (Mill, 2005). Algoritmos enviesados – por dados históricos, variáveis-proxy ou escolhas de otimização – distorcem acesso a oportunidades, desperdiçam talentos e induzem decisões subótimas em mercados e políticas públicas. O resultado não é apenas injusto; é também socialmente custoso. Em termos institucionais, o utilitarismo aconselha estruturas de governança que corrijam distorções sistemáticas: auditorias periódicas, testes de impacto, monitoramento de erro diferencial entre grupos e mecanismos de feedback que reduzam o “custo social” da classificação automatizada.

---

[1] As menções pontuais a Kant, Mill e Rawls são instrumentais e serve a caminhos diferentes que conduzem a uma mesma censura de fundo às práticas discriminatórias, entendidas as peculiaridades históricas que precisam ficar bem datadas, sobretudo quando elas se apresentam sob aparência de neutralidade técnica e hegemônica. Além disso, o discurso público-justificatório deve ser apresentado a partir de uma reconstrução dogmático-filosófica acerca da defesa dos direitos individuais (e coletivos). Em Kant, o sujeito é irredutível a variáveis algoritmizadas, de modo que não pode ser aniquilado em sua humanidade a troco de uma racionalidade de maximização datológica. Mill estabelece esse contraponto de legitimidade; a eficiência tende a naturalizar *trade-offs* de direitos fundamentais a partir de uma lógica de internalização das externalidades em virtude do progresso. Não se pode reduzir a tutela jurídica a um desempenho estatísticos ou por qualquer outra métrica que não a métrica jurídica. Por fim, em Rawls vivificamos a necessidade de que as instituições, se querem servir ao Direito e à justiça, devem manter condições equitativas de participação de todos os atores sociais, sob pena de quebra do pacto social. Desse modo, utiliza-se destas tradições de filosofia moral, conectadas à tutela jurídica, para auxiliar na construção epistêmica do que significa exigir justificabilidade, transparência e racionalidade no atual contexto algorítmico.

Rawls, por sua vez, desloca o foco para as condições de justiça de um arranjo institucional visto “sob o véu da ignorância” – situação em que ninguém sabe sua posição social, talentos, raça, gênero ou etnia (Rawls, 2003). Sob tal condição, ninguém aceitaria sistemas de decisão que, por desenho, agravem desvantagens imerecidas ou bloqueiem oportunidades justas. Dois corolários emergem: i) as liberdades básicas devem ser garantidas a todos – o que inclui, no plano informacional, condições de participação, transparência e contestabilidade das decisões automatizadas –, e ii) desigualdades só são toleráveis se atreladas ao princípio da diferença e à igualdade equitativa de oportunidades. Aplicado à esfera algorítmica, o teste rawlsiano exige que as regras técnicas e organizacionais – coleta, treinamento, validação, implantação e uso – não aprofundem barreiras de acesso e, quando produzam assimetrias, que estas sejam justificadas por ganhos reais aos menos favorecidos.

Young (1990) amplia e aprofunda a crítica ao acentuar que a opressão não se reduz a atos individuais de preconceito, mas se exerce por normas, padrões culturais e instituições que convertem diferenças em desvantagens sistemáticas. Seus “rostos da opressão” – exploração, marginalização, impotência, imperialismo cultural e violência – oferecem gramática analítica para a era da automação: modelos que excluem sistematicamente candidatos de certos bairros (marginalização), que atribuem menor crédito a perfis associados a grupos específicos (exploração), que silenciam vozes afetadas pela lógica do modelo (impotência), que naturalizam padrões culturais dominantes como neutros nos dados (imperialismo cultural) e que expõem pessoas a riscos de abordagem estatal desproporcional (violência). Em termos técnicos, trata-se do uso de variáveis correlatas (*proxies*) que, embora formalmente neutras, podem funcionar como substitutos de marcadores sensíveis e reproduzir exclusões historicamente estabilizadas. Sob esse prisma, a “neutralidade técnica” é frequentemente apenas um modo de invisibilizar escolhas valorativas incorporadas aos sistemas.

A passagem do plano moral ao jurídico exige positivar essas razões em deveres e standards operacionais. No constitucionalismo brasileiro, a dignidade da pessoa humana e a igualdade substancial fundamentam proibições de discriminação e autorizam – mais do que isso, exigem – políticas e arranjos institucionais de proteção ativa. Em ambiente de decisões automatizadas, tais deveres se traduzem em

exigências de prevenção e de prestação de contas: governança de dados e de modelos orientada por não discriminação, transparência e explicabilidade proporcionais ao risco, supervisão humana significativa, revisão administrativa e judicial de decisões automatizadas e mecanismos de correção quando evidenciado impacto desproporcional sobre grupos vulneráveis. No plano infraconstitucional, princípios como finalidade, necessidade, adequação, transparência e não discriminação, ao lado do direito de revisão de decisões automatizadas, conformam um mínimo normativo para a operação legítima de sistemas decisórios.

Desse conjunto, pode-se extrair um teste normativo integrado – útil tanto para desenho regulatório quanto para controle jurisdicional – ancorado em quatro eixos: (a) respeito kantiano – vedações a classificações estigmatizantes e opacas que inviabilizem justificação compreensível e contraditório; (b) equidade rawlsiana – verificação de igualdade equitativa de oportunidades, com atenção a barreiras indiretas e a requisitos de compensação quando houver assimetrias residuais; (c) eficiência e bem-estar de Mill – auditoria de custos sociais, erro diferencial e desperdício de talentos causados pelo modelo; (d) justiça estrutural em Young – mapeamento de efeitos sistêmicos, inclusive quando não intencionais, que reproduzem marginalização e impotência. A presença de riscos elevados em qualquer desses eixos aciona escrutínio reforçado – e, conforme o caso, proibição, moratória, *redesign* ou remediação dirigida.

Importa, ainda, distinguir níveis de incidência: discriminação direta – quando o tratamento desfavorável se ancora explícita ou implicitamente em atributos protegidos – e discriminação indireta – quando regras e métricas neutras em aparência produzem impacto adverso desproporcional. A primeira impõe veto quase absoluto; a segunda exige teste de justificabilidade estrita: finalidade legítima robusta, necessidade (inexistência de alternativa menos gravosa com eficácia comparável) e proporcionalidade em sentido estrito, com avaliação transparente de custos e benefícios distribuídos. Em ambos os casos, a prova não pode recair integralmente sobre o indivíduo afetado – há dever institucional de demonstração e documentação do processo decisório, inclusive para permitir controle social e judicial efetivos.

Em síntese, o enquadramento moral não apenas explica por que a discriminação – inclusive a mediada por algoritmos – é censurável, mas indica como o direito deve reagir: estabelecendo salvaguardas procedimentais e materiais,

distribuindo ônus argumentativos, impondo obrigações de prevenção e de cura e abrindo canais de participação e contestação. A discussão que se segue deriva desse núcleo – e apenas assim – para construir critérios jurídico-decisórios operacionais que, sem fetichizar a técnica, condicionem o uso socialmente legítimo de sistemas automatizados à preservação da igualdade material e da dignidade humana.

A noção constitucional de igualdade exige o reconhecimento de que o tratamento desigual pode ser juridicamente legítimo – e até necessário – quando dirigido à superação de desigualdades estruturais[2]. Nesse sentido, o núcleo do problema reside na identificação da medida e do significado dessa desigualdade: em que ponto o tratamento diferenciado constitui reparação legítima e quando passa a configurar discriminação indevida? O tratamento desigual torna-se ilícito quando se funda em razões arbitrárias, como raça, gênero, religião ou origem social, refletindo estigmas historicamente construídos que operam no cotidiano por meio de práticas muitas vezes inconscientes, mas reiteradas. A desigualdade, assim, não se manifesta apenas no plano das intenções, mas sobretudo nas estruturas que moldam as percepções sociais e condicionam decisões – inclusive algorítmicas – com base em imagens desumanizantes da diversidade humana. É nesse pano de fundo que emerge a urgência de refletir, no campo do Direito, sobre como se operacionaliza a proteção de grupos vulnerabilizados diante das novas formas de discriminação.

A desigualdade ganha destaque de tratamento jurídico quando se consubstancia em uma desigualdade de condições, o que é tutelado com vias reparativas dentro do cenário jusconstitucional brasileiro, ou então quando há um tratamento desigual injustificado, isto é, lastreado em razões escusas ou ilícitas, como o tratamento desigual por conta de aspectos de raça, cor, gênero, sexualidade, etnia, religiosidade, etc. O tratamento desigual baseado em diferentes bases associativas da dignidade do outro recruta atenção quanto ao objeto do tratamento: o ser humano.

A questão central não está nos episódios caricatos – que existem e devem ser coibidos –, mas nas interações ordinárias em que o tratamento desigual decorre de processos decisórios não conscientes, moldados por um cenário político-sociológico-econômico e cultural que naturaliza imagética negativa a pessoas de origem humilde, periférica, miserável, ou, ainda, sem tal perspectiva econômica,

[2] Não se parte de uma visão decolonial do direito antidiscriminatório, ainda que dela não se olvide integralmente (Creuz, 2024, p. 14)

pessoas com tom escuro da cor compartilhada por todos os humanos em suas peles: o marrom; ou, ainda, que não praticam a fé religiosa hegemônica do lugar contemplado, como tantas outras características que apontam, em verdade, apenas à diversidade cultural da vida humana. Em tais contextos, a discriminação não deriva de um propósito explícito de excluir; ela emerge de estruturas históricas que condicionam percepções – de classe, raça, gênero, religião e território – e produzem efeitos jurídicos desfavoráveis sem que o agente se reconheça como discriminador. É precisamente por operar nesse plano difuso e reiterado que a desigualdade se torna estrutural, reclamando do Direito um dever de proteção que transcenda a neutralidade formal e alcance a substância das posições concretas.

Assim, a desigualdade ganha destaque quando se observa a necessidade de ações positivas pelo direito para corrigir distorções históricas, econômicas, sociais, etc., que perpetuam a marginalização de determinados grupos para além de uma efetiva realidade econômica de marginalização (o que seria um único critério moralmente aceitável de percepção de desigualdades). Medidas compensatórias pelo direito tornam-se essenciais para promover a igualdade material, que vai além da simples igualdade formal perante a lei, buscando efetivamente equilibrar as oportunidades e condições de vida entre os diferentes “segmentos” da sociedade. Mas, por que há, na atual sociedade, uma organização nessa estrutura discriminatória?

A compreensão das vulnerabilidades, especialmente de grupos estruturalmente discriminados, é essencial para traçar um cenário minimamente fidedigno aos desafios que a nova realidade oferece à comunidade e ao direito. A discriminação estrutural, muitas vezes disfarçada nas sombras da sociedade, afeta profundamente pessoas e grupos, criando um fosso entre as possibilidades de acesso a direitos fundamentais e o seu efetivo exercício. Assim, é necessário estabelecer critérios e noções a partir do direito acerca da vulnerabilidade, da discriminação estrutural e dos grupos vulneráveis, bem como compreender como o direito pode oferecer instrumentos de compensação dessa desigualdade.

O conceito de vulnerabilidade no direito está em sua raiz principiológica relacionado ao princípio da igualdade, do qual também se é derivado. A Constituição Federal estabelece o direito à igualdade geral no *caput* do artigo 5º, garantindo a todos que estejam no país a inviolabilidade de sua igualdade. A igualdade, portanto, não é

apenas um direito fundamental, mas sim um valor estruturante do próprio compromisso democrático-constitucional estabelecido na Carta da República, que também prevê a construção de uma sociedade livre, justa e solidária. Livre para se desenvolver, justa para reconhecer a dignidade em si e nos outros, atribuindo-lhes o que lhes é de direito, e solidária e fraternal, para pertencer ao sentimento de pátria.

A igualdade enquanto valor não consubstancia apenas umas das conquistas liberais, do período revolucionário do Século XVIII, sendo um dos direitos fundamentais de primeira geração, mas marca também a característica dos direitos de segunda geração, sendo esses direitos de igualdade materialmente positiva. Os direitos da dimensão da igualdade se caracterizaram pelo viés econômico, social e cultural. Esses direitos, sobretudo após os avanços de jurisdição constitucional alcançados no pós-guerra, buscam a concretização da igualdade material, e não apenas uma igualdade formal perante a lei (Bonavides, 2004, p. 564-565).

Para Aristóteles (2001, p. 101-105) a igualdade é o termo intermediário entre extremos, considerando não apenas uma mesma parte (igualitária), mas que as partes consubstanciem um justo, sendo, caso necessário, mais para uns e menos para outros. E não menos justo por isso. Os objetos são distribuídos por critério de justiça e igualdade, não só de igualdade formal. E isso se deve ao fato de que “se as pessoas não são iguais, não receberão coisas iguais” (Mascaro, 2016, p. 78-81). O que importa é a justiça do caso concreto[3].

Apoiado em Aristóteles, Tomás de Aquino correspondia à noção de que a justiça, e, portanto, o *Ius*, o direito (*rectum*), é manifestado naquilo que é correto na obra de outras virtudes. Por isso, aquilo que é justo, só o é em comparação ao que é distribuído a todos. Assim, a manifestação do *Ius*, aquilo que tende a justiça, só pode ser produto da ação correspondente a concretização de uma igualdade (Lopes, 2004, p. 92). Tratar igualmente os iguais e desigualmente os desiguais, portanto, é o reconhecimento de que existem situações na vida que, permanentemente ou não,

[3] Na clássica acepção presente na A Política, de Aristóteles, a igualdade parece ser a base do direito, e o é efetivamente, mas unicamente para os iguais e não para todos. A desigualdade também o é, mas apenas para os desiguais [...] A igualdade da qual resulta a justiça ocorre, como igualmente o demonstra a nossa *Ética*, nas pessoas e nas coisas. Concorda-se facilmente sobre a igualdade das coisas. Sobre a das pessoas erguem-se protestos, porque mais uma vez os homens se tornam cegos sobre si mesmos e tendo, de uma e de outra parte, razão até certo ponto, querem dar a seu direito uma extensão ilimitada. [...] Em matéria de igualdade e de justiça, não é fácil encontrar a verdade exata; é bem fácil consultar a sorte do que persuadir os que podem ser os mais fortes. Os fracos não podem mais do que igualdade e justiça, mas os mais fortes pouco se importam com isso (Aristóteles, 2006, p. 162).

impõe uma fragilidade a sujeitos, sejam por questões de idade, saúde ou econômicas, e isso desequilibra toda e qualquer relação jurídica na qual esses vulneráveis façam parte (Marques; Miragem, 2012, p. 117).

No Brasil, o direito privado após o Século XIX sofreu uma dupla e forte influência. Por um lado, recebeu a aludida influência liberal, reconhecendo igual dignidade à população, e estabelecendo o critério de *Æqualitas* (igualdade formal) para as relações civis, ao passo que também ganhou a preocupação da realização material da justiça, estabelecendo assim a *Æquitas* (igualdade material), traduzida no reequilíbrio da relação entre privados, fazendo com que o nitidamente mais vulnerável fosse o mais protegido e tutelado em certas situações jurídicas (Marques; Miragem, 2012, p. 113-115).

E ainda que historicamente a proteção dos vulneráveis seja um dos objetivos próprios para a concepção de uma ordem jurídica democrática, a noção moderna de tutela aos vulneráveis ganha força redobrada com a emergência da Teoria dos Direitos Fundamentais, e, no Brasil, sobretudo a partir da promulgação da atual Constituição Federal.

Para Alexy (2015, p. 396), o direito de igualdade exige que toda norma jurídica seja aplicada com a preocupação de que a sua aplicação resultará em um real tratamento igualitário perante o direito, sendo vedadas discriminações de pessoas em igual posição jurídica. Um desapego a essa preocupação resultaria em normas incompatíveis com a finalidade do direito. O reequilíbrio das questões de igualdade, contudo, tem o limite das questões fáticas reequilibráveis, de modo que não se pode, mediante a jurisdição ou a lei, suprir saúde, inteligência ou beleza, o que se pode é relativizar os efeitos dessas condições pelo direito, respeitados os limites naturais.

Assim, na teoria de Alexy (2015, p. 398), a igualdade é a previsão de incidência de mesmas normas, e, portanto, de mesmas consequências jurídicas, por conta do cumprimento da hipótese de incidência fática, descrita na norma, e não condicionada à qualidade ou característica do agente que praticou o ato em si, ou que reivindica algo do Estado. Essa preocupação formal, deve, contudo, estar atrelada à fundamentação material da igualdade, conforme determina e se compromete o texto constitucional. Destarte, Alexy (2015, p. 400) assevera que não há desigualdades iguais, ou desigualdades fáticas universais, por isso que a ordem jurídica deve se preocupar tanto na valoração da igualdade quanto na possibilidade de diferenciação

do tratamento. A igualdade material é o substrato da valorização na medida correta, na qual o indivíduo desigual, ou vulnerável, recebe um tratamento diferenciado e privilegiado, não para beneficiá-lo, mas para colocá-lo em igual posição para com outros não desprivilegiados, protegendo juridicamente o vulnerável, tanto de outros não vulneráveis em uma relação interacional de que faça parte, quanto para com todas as demais interações entre pessoas, naturais ou jurídicas, não vulneráveis. No Brasil, esse tratamento diferenciado aos desiguais é cláusula democrática decorrente do texto constitucional.

A Constituição Federal direciona o ordenamento jurídico na proteção de grupos vulneráveis específicos, como consumidores, idosos e crianças e adolescentes, através de legislação dedicada – Código de Defesa do Consumidor, Estatuto da Pessoa Idosa e Estatuto da Criança e do Adolescente. A interpretação e aplicação de todas as normas devem considerar a vulnerabilidade como um critério essencial para prevenir danos injustos – critérios de diferenciação e finalidade de distinção para com congruência do critério e razão do fim desejado. Essa perspectiva é reforçada pela noção de que a igualdade permite tratamento desigual apenas para corrigir desigualdades, rejeitando discriminações arbitrárias que não se justifiquem sob critérios racionais e objetivos (Alexy, 2015, p. 409). A vulnerabilidade é escopo interpretativo da ordem jurídica, impondo um ônus de maior cuidado pelo intérprete-aplicador, sendo tarefa sua demonstrar argumentativamente o grau de desigualdade e atribuir o critério da vulnerabilidade comparativa, bem como a consequência jurídica ou mecanismo de mitigação. Deste modo, se não houver razão suficiente para permitir dispensar um tratamento igualitário, o tratamento desigual (mais protetivo) é obrigatório na medida das desigualdades e da vulnerabilidade constatada.

Nesse sentido, a igualdade enquanto regra prevê a proibição de tratamento discriminatório, como princípio, instituindo um dever de promover um estado igualitário, e, enquanto postulado, estrutura a “aplicação do Direito em função de elementos (critério de diferenciação e finalidade de distinção) e da relação entre eles (congruência do critério e razão do fim)” (Ávila, 2016, p. 192-194).

A possibilidade de tratamento desigual integraliza o princípio da igualdade, vedando apenas o tratamento desuniforme às pessoas (Mello, 1993, p. 35). Isso é o que normas fazem, criam situações discriminadas, e para cada qual dão uma solução jurídica que atende melhor aos fins da prescrição normativa, protegendo quem deve

ser protegido, porquanto atingidos por regimes jurídicos diferenciados. Para alguns há determinados direitos e obrigações que não são exigidos ou garantidos a outros, e isso é a acepção material da igualdade (Mello, 1993, p. 12).

O que não pode ser admitido, como assinala Sarlet (2020, p. 625-626), é a existência de normas ou decisões que discriminem juridicamente questões indiscriminadas, trazendo prejuízos intoleráveis para o seio do conceito normativo de certo ordenamento jurídico. De modo que, conforme Alexy (2015, p. 409), se não houver razão suficiente que permita um tratamento desigual, o tratamento igualitário é obrigatório, sendo o contrário também verdade, de modo que se não houver razão suficiente para permitir um tratamento igualitário, o tratamento desigual é obrigatório na medida das desigualdades. Essa realidade, contudo, impõe um ônus argumentativo – o qual será brevemente contextualizado – de demonstrar o grau de desigualdade e atribuir o critério da vulnerabilidade comparativa, bem como a consequência jurídica ou mecanismo de mitigação, conteúdo intencionais, considerações substanciais e casuísticas que devem ser realizadas por juízes e tribunais em respeito à cláusula (contida essencialmente no art. 93, IX, da Constituição da República, bem como art. 489, §1º, do Código de Processo Civil) de que as decisões judiciais devem basear-se em argumentações racionais (Alexy, 2020, p. 22). A racionalidade das decisões remonta tanto a uma dimensão prático-geral, de onde decorre o caso especial da Teoria da Argumentação Jurídica, quanto a uma dimensão comunicativo-dialógica, em que os núcleos de significações não estão inteiramente fechados na sistemática normativo-interpretativa, sendo este o lócus adequado para a construção do sentido imperativo da normatividade do direito a partir de casos (Paulo, 2022, p. 128). Autorizados acerca do conteúdo do direito fundamental em destaque, bem como do método de sua aplicação dialógico-argumentativa, impõe-se voltar os olhos à dinâmica de sua existência na dimensão jusfundamental, com as respectivas interações com a vulnerabilidade. O reconhecimento da necessidade da igualdade material levou então ao amadurecimento de que se possa discriminar pessoas iguais, na medida de suas desigualdades, privilegiando-as com um equilíbrio artificial da tecnologia ficcional do direito. Esse movimento é fruto de uma percepção de que a igualdade formal não afasta situações de injustiça, e é exatamente por isso que o mandamento da igualdade é o de proibir tratamentos arbitrários, que prejudiquem a fluidez da dignidade da pessoa humana na acepção de um Estado de

Direito Constitucional (Sarlet, 2020, p. 601-616). Como se verá adiante, o tratamento a partir do prisma da igualdade e das vulnerabilidades demandará situações diferenciadas, tanto a partir das funções típicas dos poderes no âmbito público, quanto da proteção jurídica nas relações privadas.

Na clássica divisão de dimensões dos direitos fundamentais, a igualdade pode se dividir em uma dimensão negativa, no sentido de impor ao Estado que não trate os sujeitos em desacordo com as exigências da igualdade, e, na perspectiva positiva, ou prestacional, para que as pessoas tenham igual possibilidade de acesso a políticas, bens e serviços estatais, e que esses bens e serviços, ou mesmo classificações jurídicas outras, reequilibrem suas posições jurídico-sociais. O direito à igualdade na faceta prestacional tem uma dimensão a mais, a de que o Estado promova a compensação da igualdade por meio de políticas e ações afirmativas, isto é, um dever de atuação estatal nas esferas normativa, política e fática para a mitigação da desigualdade. O tratamento da vulnerabilidade pode ser alocado como uma política judicial e judiciária, compromisso do Estado de proteger os vulneráveis.

Assim, o reconhecimento de situações de vulnerabilidade, e sua efetiva proteção por normas e por decisões, concretiza a faceta material da igualdade, pois discrimina positivamente aqueles que precisam de mais proteção, sobretudo quando protege os hipervulneráveis. Tal conceito jurisprudencial funciona como um corolário positivo da proibição de discriminação injusta, que interconecta as dimensões de desenvolvimento da personalidade, dignidade da pessoa humana, tratamento equitativo e igualdade material (Marques; Miragem, 2012, p. 192). É, portanto, um símbolo de linguagem que realça a complexidade, e a possível gravidade, da situação, guarnecendo justificativa do recebimento de tratamento diferenciado pelo Estado.

A igualdade, portanto, se desdobra em obrigações negativas e positivas para o Estado: não apenas evitar a discriminação, mas também promover ativamente a igualdade por meio de políticas e ações afirmativas. Tais medidas são vitais para reequilibrar as condições sociais e jurídicas, garantindo que vulnerabilidades sejam adequadamente abordadas tanto na esfera pública quanto nas relações privadas, cumprindo o compromisso estatal de proteção. O reconhecimento de situações de vulnerabilidade, e sua efetiva proteção por normas e por decisões, concretiza a faceta material da igualdade, pois discrimina positivamente aqueles que precisam de mais proteção, sobretudo quando protege os hipervulneráveis. A categoria jurisprudencial

da hipervulnerabilidade, ainda que cunhada na tutela consumerista, opera como metacategoria hermenêutica da igualdade material – e, assim, informa a leitura constitucional do dever de proteção.

A hipervulnerabilidade algorítmica qualifica uma camada específica de exposição derivada de decisões automatizadas em contextos de assimetria informacional e opacidade técnica, que incide de forma interseccional sobre grupos historicamente discriminados. Não se trata apenas de fragilidade econômica ou informacional clássica, mas pela sua forma de atuação: ela não exige a intenção discriminatória para se concretizar, tampouco é facilmente perceptível (Leal; Crestane, 2023). É reproduzida por perfis *proxy* e correlações opacas. Há uma "cegueira estrutural" promovida por essas tecnologias, que reforça estigmas sob o verniz de neutralidade estatística. Em tais hipóteses, a igualdade material demanda standards reforçados de prevenção, explicabilidade e revisão, bem como arranjos probatórios que compensem a assimetria técnica, sob pena de o devido processo converter-se em ritual sem razões.

Desse modo, enquanto a vulnerabilidade é um estado da pessoa que tem relação com o risco inerente dela na sociedade e no mercado, sendo situação que a fragiliza enquanto sujeito de direitos, a hipervulnerabilidade é o reconhecimento de que essa fragilização pode ser ainda maior, isto é, uma vulnerabilidade potencializada tanto em relação aos não vulneráveis quanto em relação aos próprios vulneráveis. Pode-se dizer que a hipervulnerabilidade surge tanto de um agravamento da vulnerabilidade, pela própria impossibilidade de discernimento e proteção, como no caso das crianças e dos adolescentes na Internet, ou então pela soma de vulnerabilidades pessoais, como um celíaco consumidor em um restaurante sem a informação de se os alimentos contêm glúten ou não (Konder, 2020, p. 94). A isso se chama grupos vulneráveis, ou seja, indivíduos jurídicos que, ao ocupar certa posição jurídica, têm sua vulnerabilidade circunstancial realçada e ganham especial tutela por conta disto.

Eis o cerne do reequilíbrio social que o Estado Democrático oferece. Há, contudo, que se ir um pouco mais além na busca conceitual do estado da arte da vulnerabilidade, eis que a desigualdade pode ainda figurar-se a partir de uma forte interseccionalidade que impulsiona a lesividade da condição, como é o caso de pessoas que se enquadram em múltiplas categorias de vulnerabilidade e são

objetificadas, até mesmo inconscientemente, como menores em importância e dignidade pelas estruturas sociais, ou então de modo intencional (Moreira, 2019, p. 64). Há, ainda, a discriminação estrutural que, muitas vezes, traz em si não apenas o conceito de interseccionalidade, mas também trabalha com a categoria das minorias – enquanto grupo de pessoas que se encontra em posição de desvantagem em relação à sociedade como um todo, sejam por questões econômicas, políticas, de origem étnica, nacional ou linguística, orientação sexual, escolhas religiosas e outras, bem como sejam elas em número reduzido ou não dentro da sociedade observada (Carbonell, 2000).

Aqui “grupo em situação de vulnerabilidade” descreve uma condição relacional e contextual (não um rótulo fixo): ela surge quando barreiras sociais, econômicas, institucionais e tecnológicas limitam a participação em condições de paridade. “Grupo vulnerável” é a forma abreviada quando a situação é estável e reiterada no tempo. Já “minorias” não se refere a número de indivíduos, mas a assimetria de poder e sub-representação em processos de decisão. Essa distinção evita distorções e permite calibrar a proteção conforme o caso concreto, com o (excelente) viés contramajoritário dos direitos fundamentais.

Não se trabalhará a partir do recorte das minorias, mas sim dos grupos vulneráveis – mais adiante grupos vulneráveis em relação ao uso dos algoritmos –, cujos indivíduos podem pertencer, ou não, a uma minoria e, daí, demandarem atenção ainda mais tutelar do direito. A discriminação sofrida por esses, entretanto, é de natureza estrutural, e essa demanda não apenas um olhar atento em termos conceituais, mas em termos da prática da jurisdição e do dever de proteção estatal.

De modo insidioso e persistente, a discriminação estrutural opera nas fissuras do tecido social, restringindo, ainda que de forma não explicitamente normativa, o gozo equitativo dos direitos fundamentais, ocasionando danos sistêmicos e crônicos no tecido social.

A discriminação estrutural é mais difícil de identificar e abordar do que formas individuais e intencionais de discriminação, pois muitas vezes opera de maneira invisível, arraigada e enraizada na sistemática cotidiana da sociedade, como preconceitos e desigualdades históricas. Grupos vulneráveis geralmente sofrem mais com a discriminação estrutural, experimentando desvantagens significativas em comparação com grupos menos desprivilegiados.

Acerca da discriminação estrutural, Sagüés (2018) propõe que a desigualdade sistemática seja atacada por uma proposta de transformação social na instância superior da igualdade estrutural, permitindo que decorra desta concepção a necessária tutela especial a grupos desavantajados, sobretudo das ações afirmativas – ainda que em um necessário caráter excepcional e transitório – pensado para que venha a inexistir com a eficácia das medidas no tempo.

Fato é que diante de tais situações de discriminação, seja por vulnerabilidades, pelo indivíduo pertencer a um grupo vulnerável ou sofrer com discriminação estrutural, os poderes da nação têm, cada qual sua competência, o dever de proteger os indivíduos das injustas e injustificadas agressões aos direitos que vierem a receber, o poder judiciário, especialmente, tem os deveres de interpretar definitivamente os núcleos de significação dos textos normativos provenientes dos demais poderes, bem como tutelar por último a correta, proporcional e adequada resposta aos injustos percebidos. Todos esses, como visto, estão autorizados, a partir do conteúdo de significações da igualdade, a compensar positivamente as situações e posições jurídicas, na medida da necessidade para o reequilíbrio da ordem da situação. Na realidade que se apresenta na nova dinâmica da inteligência artificial, há um vácuo normativo acerca da matéria, mas as injustas agressões e as perpetuações de preconceitos estruturais não aguardou pela confecção e implementação de uma qualquer norma legal legislativa, sendo cabível ao poder judiciário que dê em caráter tópico, e jurisprudencial, as respostas cabíveis e os parâmetros racionais de interpretação e aplicação do direito nas demandas que lhe forem dadas a decidir.

A discriminação estrutural é entendida como padrão persistente de produção de desvantagens que atravessa instituições, mercados e rotinas, e usa a noção de paridade de participação e oportunidades de participação para dizer quando há injustiça: quando pessoas ou grupos não conseguem participar, em pé de igualdade, dos espaços onde se distribuem recursos, reconhecimento e representação (Fraser, 2007, p. 113). Isso implica combinar remédios de redistribuição (efeitos materiais), reconhecimento (estigmas e estereótipos codificados nos dados) e representação (quem define as regras do sistema). Em termos jurídicos, isso justifica escrutínio reforçado quando algoritmos incidem sobre grupos socialmente desvantajados. A vulnerabilidade deixa de ser rótulo, se convertendo em critério metodológico-interpretativo do direito.

Como se viu, a partir do lançar de luzes conceituais acerca da complexa (multifacetária e multidimensional) questão da vulnerabilidade e da discriminação estrutural, é fundamental ter por próximo passo a consideração desses elementos para a interpretação acerca da abordagem da discriminação algorítmica, eis que atinge de modo especialmente nocivo a essas comunidades a que se denominou “grupos vulneráveis”, e o fazem, como visto, a partir de interseccionalidades, isto é, múltiplas origens e motivações injustificadas ainda arraigadas no seio social, ao que alegoricamente fixou-se como “sombras”. Tais interseccionalidades demonstram, sobremaneira, as nuances que se está por detrás da problemática da discriminação algorítmica às individualidades de pessoas que integram a categoria de grupos vulneráveis, e, portanto, a cada uma se deve atentar para com as desproporções ocasionados pela prática cotidiana, sendo cabível ao direito o reequilíbrio proporcional e gradativo desta balança, trazendo clarões de direito e de fruição de garantias fundamentais.

2.2 A discriminação estrutural e sua relação com o direito de igualdade e de não-discriminação

O conceito de “discriminação”, no âmbito jurídico, abrange qualquer forma de diferenciação que limite ou prive indivíduos de usufruírem, em condições equitativas, de seus direitos e oportunidades. Nesse sentido, é relevante salientar que a discriminação pode ocorrer de maneira não deliberada, uma vez que o tratamento discriminatório nem sempre está ligado a uma ação proposital, mas pode ser consequência de práticas ou padrões de comportamento indiretos. Ou seja, mesmo sem a intenção explícita de discriminar, os efeitos de determinadas práticas podem configurar discriminação. Essa abordagem é fundamental para a análise da discriminação estrutural, que não decorre essencialmente de atos individuais e planejados, mas é perpetuada por normas e valores profundamente arraigados na organização social, criando obstáculos de acesso para determinados segmentos (Duarte; Negócio, 2021, p. 224).

A discriminação estrutural diz respeito aos tratamentos diferenciados dispensados a coletivos em virtude de injustiças sociais, formação precária e preconceitos profundamente arraigados. Esse fenômeno está intrinsecamente ligado

aos estigmas sedimentados no tecido social, cultural e institucional, que frequentemente inviabilizam ou dificultam o pleno usufruto de direitos por parte das populações afetadas (Silvério, 2022, p. 233). De acordo com sua própria definição, tal prática está incorporada às bases organizacionais da sociedade, perpetuando barreiras invisíveis que obstruem o avanço e a equidade de grupos vulneráveis. Conforme abordado anteriormente, a desigualdade estrutural incide majoritariamente sobre coletivos historicamente marginalizados. Dados de pesquisas recentes evidenciam que tal disparidade persiste como uma realidade incontestável e alarmante. Por exemplo, em termos de discrepância salarial, indivíduos de cor branca percebem rendimentos 64,2% superiores aos de pessoas negras, enquanto, no que tange ao gênero, os homens auferem salários 27% mais elevados do que as mulheres (IBGE, 2023, p. 25). Essa constatação demonstra a urgência do tópico.

Nessa moldura, cumpre reconhecer também os limites da tipologia clássica entre discriminação direta e indireta, cuja aplicação rígida pode obscurecer a materialidade da exclusão. Práticas “aparentemente neutras” — p. ex., exigência linguística ou códigos de vestimenta — tanto podem operar como critérios funcionais quanto como filtros que atingem desproporcionalmente grupos específicos, a depender do contexto e da finalidade (Collins; Khaitan, 2017). A ênfase classificatória, se divorciada da experiência vivida, desloca o foco para disputas formais e, não raro, minimiza opressões reais (Fredman, 2017). O problema se agudiza em cenários interseccionais, nos quais a conjunção de marcadores (gênero/raça/classe) produz exclusões não captáveis por categorias puras, reclamando abordagem substancial e contextualizada (Crenshaw, 1989).

A compreensão da discriminação estrutural exige, como ponto de partida, uma análise ontológica da condição humana. Fineman (2023, p. 1373) propõe uma mudança epistêmica significativa ao deslocar o enfoque tradicional da igualdade – centrado em abstrações universalistas – para a constatação concreta da vulnerabilidade como traço fundante da existência. Esse deslocamento não apenas altera o eixo teórico da igualdade jurídica, mas oferece instrumental teórico capaz de iluminar as raízes da desigualdade, ao evidenciar que antes de sermos sujeitos de direitos, somos corpos expostos à dor, à carência, à dependência e à finitude.

A vulnerabilidade não é mero atributo eventual, presente em grupos identificados como frágeis; ao contrário, constitui uma dimensão essencial da

existência humana. Todos os indivíduos, em alguma medida, estão sujeitos a contingências sociais, biológicas e políticas que os tornam vulneráveis. No entanto, em contextos específicos, tais vulnerabilidades são historicamente intensificadas, canalizadas e instrumentalizadas por sistemas de dominação que se perpetuam de forma estrutural. Assim, o reconhecimento da vulnerabilidade universal não implica a diluição das especificidades das formas de discriminação, mas fornece uma lente crítica para compreender como determinadas populações se tornam alvos sistemáticos da exclusão.

Neste sentido, o discurso jurídico tradicional – ancorado em uma concepção liberal de sujeito autônomo, racional e desvinculado de vínculos materiais – se mostra insuficiente para enfrentar a complexidade da discriminação estrutural. A igualdade formal, tal como concebida nos marcos clássicos do direito, tende a invisibilizar os modos como o poder opera na produção e na reprodução da desigualdade. Ao reconhecer a vulnerabilidade como paradigma, o direito é convocado a reformular suas categorias, substituindo a noção de neutralidade pelo compromisso ativo com a justiça substantiva.

A concepção de vulnerabilidade como traço inerente à condição humana, conforme demonstrado por Fineman, desloca o foco da igualdade formal para uma ética de reconhecimento e cuidado. Tal deslocamento exige que o Direito deixe de operar com sujeitos abstratos e passe a compreender a desigualdade como fenômeno sistêmico, enraizado nas estruturas sociais e jurídicas que amplificam vulnerabilidades seletivas (Borges, 2025, p. 23).

O sujeito vulnerável, como construído no texto de referência, não é apenas aquele que sofre, mas também aquele que resiste, que demanda proteção, cuidado e reconhecimento. A vulnerabilidade, portanto, é política. Ela mobiliza o Estado, interpela as instituições jurídicas e desafia os arranjos normativos. Nesse cenário, torna-se evidente que a discriminação estrutural não pode ser compreendida apenas como uma prática excludente pontual; ela é resultado de uma lógica sistêmica que captura vulnerabilidades para produzir hierarquias duráveis, silenciosas e normatizadas (Foucault, 2014, p. 194).

Ao integrar essa perspectiva ao estudo da discriminação estrutural, abre-se espaço para uma hermenêutica do cuidado e da interdependência. Trata-se de uma mudança radical na forma como pensamos o sujeito jurídico – não mais como um ser

isolado que negocia direitos, mas como um ente relacional, situado, atravessado por laços e exposições que demandam resposta normativa. O desafio, portanto, é construir uma teoria da justiça que parta da vulnerabilidade como condição primeira, e não como exceção.

O paradigma da vulnerabilidade traz à tona uma crítica contundente à concepção liberal de igualdade, que ignora a historicidade e a materialidade dos corpos. A igualdade formal pressupõe sujeitos que competem em arenas neutras e simétricas, como se a justiça se bastasse na distribuição equânime de direitos abstratos. Todavia, essa ideia silencia as múltiplas camadas de opressão que se cristalizam em estruturas históricas. Ao reconhecer que certos grupos – mulheres, pessoas negras, indígenas, pessoas com deficiência, populações LGBTQIAPN+ – carregam sobre si vulnerabilidades aprofundadas pelas dinâmicas sociais, políticas e econômicas, evidencia-se que a igualdade só pode ser pensada a partir de uma concepção substancial. A assimetria é também epistêmica: certos sujeitos enfrentam, de partida, déficit de credibilidade e lacunas interpretativas para nomear o dano, o que fragiliza o acesso à prova e ao contraditório (Fricker, 2007, p. 30).

Fineman (2023, p. 1386) afirma que a vulnerabilidade é uma constante universal – sobretudo nas plataformas digitais, ápice da sociedade do controle foucaultiana –, mas sua gestão é diferencial. O Estado, longe de ser um mero árbitro, torna-se agente central na produção e remediação dessas vulnerabilidades. O sujeito vulnerável exige políticas públicas orientadas não apenas à reparação de danos, mas à transformação estrutural das condições que os tornam persistentes. Desse modo, o reconhecimento da vulnerabilidade se converte em imperativo ético e jurídico, convocando uma reconfiguração das instituições, das normas e das práticas.

Ao mesmo passo, a responsabilidade do Estado na perpetuação da discriminação estrutural está diretamente conectada à omissão em reconhecer vulnerabilidades interseccionais. Como discute o artigo referência, a ausência de políticas públicas efetivas não decorre da escassez de recursos, mas da escolha política de ignorar os sujeitos que habitam a intersecção das opressões. A negligência estatal é, portanto, forma sofisticada de violência institucionalizada (Borges, 2025, p. 18).

Ao entrelaçar vulnerabilidade e discriminação estrutural, desvela-se um panorama em que o dano não se constitui somente em atos discriminatórios isolados,

mas em processos de exclusão que operam silenciosamente nos sistemas de saúde, educação, moradia, segurança e trabalho, e que vão se consolidando historicamente. A discriminação estrutural é o fenômeno da lógica contínua que, mesmo sem intenções individuais explícitas, reproduz desigualdades por meio da normatização de critérios excludentes. A vulnerabilidade, nesse cenário, é capturada e convertida em estigma: em vez de mobilizar cuidado, ela serve para justificar distanciamentos, reforçar hierarquias e legitimar violências, ainda que silenciosas.

Há, para além de uma discriminação estrutural, uma discriminação institucional. Há um silenciamento em relação às valências e possibilidades reais de atuação de uma instituição comprometida com uma visão constitucional. Almeida (2019, n.p.), verifica que o problema central não é “quem discrimina”, mas “como as organizações funcionam”. A discriminação, diz ele, opera como racionalidade de fundo que estrutura regras, rotinas e métricas “neutras”, convertendo-as em mecanismos de exclusão sem exigir dolo individual de discriminar. Daí a passagem da culpa pessoal à responsabilização institucional: mapear fluxos decisórios, medir impactos desiguais, revisar critérios aparentemente neutros (como geolocalização, escolaridade, produtividade, etc.), criar deveres positivos de prevenção (auditorias, metas, correções) e garantir controle social. Sem esse giro, a máquina burocrática e seus algoritmos apenas automatizam a desigualdade histórica, transformando-a em procedimento neutro e, portanto, válido e legítimo.

Assim, Fineman (2023, p. 1386), propõe um giro epistemológico radical: da abstração universalista à materialidade dos corpos; da neutralidade procedimental à justiça substancial; da igualdade formal à equidade afetiva. Ao reconhecer que todo sujeito é vulnerável, mas que alguns são sistematicamente vulnerabilizados, abre-se caminho para um direito comprometido com a transformação social. A vulnerabilidade deixa de ser uma condição a ser escondida ou corrigida – ela se torna fundamento para uma nova ética pública.

A vulnerabilidade, como categoria universal, ganha contornos específicos e intensificados quando interseccionada por estruturas de discriminação que se cruzam e se retroalimentam. A interseccionalidade revela que pessoas sujeitas a múltiplas formas de opressão não apenas acumulam desvantagens, mas vivenciam experiências únicas de exclusão que não podem ser compreendidas isoladamente por cada eixo de discriminação. Assim, uma mulher negra pobre, por exemplo, não sofre

simplesmente discriminação por ser mulher, negra ou pobre: ela ocupa uma posição estrutural vulnerabilizada que emerge da intersecção desses marcadores.

Fiss (1976), trabalha com a noção de uma acumulação de desvantagens, que ocorrem a partir de pequenas perdas e exclusões, reiteradas ao longo do tempo e em múltiplas instituições, que se somam e criam um desnível estrutural. Crenshaw (1989), assevera que isso se deve à verificação de opressões cruzadas que tendencialmente atingem os mesmos sujeitos, de uma dada categoria de pessoas. O problema central não é só a soma de marcadores, mas o ponto de cruzamento que a lei e as políticas tendem a invisibilizar. O primeiro explica como a desigualdade se acumula no sistema, o último mostra onde ela se intensifica e, portanto, onde as políticas públicas e decisões judiciais precisam agir. É necessário calibrar a lente conceitual para não invisibilizar da vista quem sofre o impacto mais pesado justamente nesses pontos de interseção.

Esse olhar interseccional fortalece o conceito de vulnerabilidade, ao reforçar que certos sujeitos são sistematicamente expostos à violência institucional, simbólica e normativa em múltiplas frentes. A interseccionalidade permite perceber que há vulnerabilidades agravadas não por acaso, mas por lógicas organizadas e reproduzidas pelas instituições sociais, jurídicas e econômicas. Portanto, o sujeito vulnerável não é só aquele que está em situação de fragilidade, mas aquele cuja fragilidade é resultado direto da articulação entre dispositivos de poder que operam de forma sinérgica. Os sistemas de exclusão não operam isoladamente, mas se conectam e intensificam entre si. Raça, gênero, classe, deficiência e sexualidade não são categorias estanques, mas dispositivos que se articulam na produção da desigualdade. A vulnerabilidade, nesse contexto, não é uma exceção; é a linguagem que denuncia a violência sistemática das estruturas normativas que silenciam corpos múltiplos e vivências não hegemônicas (Borges, 2025, p. 26).

Nesse cenário, o Direito, ao adotar uma perspectiva interseccional da vulnerabilidade, é compelido a rever seus próprios mecanismos de leitura e aplicação. Normas aparentemente neutras tornam-se reveladoras de escolhas políticas que invisibilizam sujeitos complexos e suas realidades multifacetadas. A ausência de reconhecimento das intersecções faz com que políticas públicas falhem em sua efetividade, pois tratam desigualdades como fenômenos isolados. A interseccionalidade, por sua vez, exige um redesenho institucional orientado por

diagnósticos sensíveis à densidade das exclusões. É vez do reconhecimento de que o Direito descompromissado com fins sociais e constitucionais é, também, instrumento da discriminação estrutural.

A discriminação estrutural interseccional revela que não há hierarquia entre opressões, mas sim um sistema dinâmico que articula diferenças para produzir desigualdade persistente e acumulativa, trabalhando tanto com a visão de Crenshaw (1989) quanto de Fiss (1976). Os marcadores sociais não operam como categorias estanques: eles se entrelaçam em contextos históricos, culturais e geográficos específicos. Isso significa que as respostas normativas também devem ser integradas e situadas, contemplando as singularidades dos sujeitos vulnerabilizados. A vulnerabilidade interseccional não demanda apenas proteção jurídica; ela exige reconstrução epistemológica e redistribuição de poder.

Ao reconhecer esse entrelaçamento, Fineman (2023, p. 1390) propõe um reposicionamento da igualdade como projeto inclusivo, plural e comprometido com a transformação social. A igualdade não se realiza pela equiparação formal entre sujeitos abstratos, mas pela articulação ativa entre cuidado, reconhecimento e justiça distributiva. A condição vulnerável torna-se, assim, não uma fraqueza, mas um território ético-político de emancipação, cuja tutela deverá ser protegida e garantida, dentre outros, pelo Estado-Gestor principalmente, ou alcançada pelo Estado-Juiz em última escala. A condição de vulnerabilidade, tomada como ponto de partida para o entendimento da discriminação estrutural, demanda uma reconfiguração profunda do aparato jurídico, social e epistemológico. A partir da interseccionalidade, compreendemos que as opressões não se distribuem de forma aleatória, mas se enraízam em dispositivos de poder que atravessam as instituições e moldam a experiência vivida dos sujeitos. A vulnerabilidade, nesse sentido, é produzida e amplificada por esses dispositivos, que operam por meio de silêncios normativos, práticas excludentes e discursos naturalizantes de vantagens e desvantagens estruturais e institucionais (Fineman, 2023, p. 1396).

O reconhecimento da interseccionalidade como ferramenta analítica permite visibilizar as zonas de exclusão múltipla, onde sujeitos vivem à margem das garantias que o Estado proclama (Crenshaw, 2002, p. 174-175). A cidadania plena, prometida pela Constituição, encontra limites reais na ausência de políticas que considerem a complexidade dos contextos vividos por populações vulnerabilizadas. Mulheres

negras periféricas, pessoas trans, indígenas, pessoas com deficiência, migrantes – todos compartilham a vivência de estruturas que não apenas negligenciam suas existências, mas as constituem como alvos de violências legítimas, porque naturalizadas e estrutural e silenciosamente projetadas para tanto.

Neste ponto, importa destacar que a discriminação estrutural não se realiza apenas nos grandes eventos de exclusão ou nas manifestações explícitas de ódio e preconceito. Ela está entranhada nas rotinas institucionais, na linguagem técnica aparentemente neutra, nos critérios de acesso a bens públicos e na configuração das prioridades orçamentárias do Estado. A vulnerabilidade, então, é seletiva: embora seja condição humana universal, sua expressão é profundamente desigual. Os corpos vulneráveis são socialmente hierarquizados e politicamente silenciados (Fineman, 2023, p. 1384).

A abordagem crítica exige, por conseguinte, uma análise das responsabilidades institucionais na produção e perpetuação da vulnerabilidade. O Estado, enquanto agente normativo central, detém não apenas o poder de garantir direitos, mas também de negá-los. Políticas públicas ineficazes, ausentes ou insuficientemente desenhadas para atender às intersecções de opressão acabam por reforçar estruturas de exclusão. A omissão estatal também é forma de violência: a falta de reconhecimento da complexidade das vulnerabilidades leva à negação da própria existência dos sujeitos que delas padecem.

A vulnerabilidade como fundamento exige um Direito comprometido com a transformação social, e, no caso do Direito como formatado, a partir de uma constituição tornada efetiva. Isso implica reformulação dos currículos jurídicos, das práticas legislativas, das decisões judiciais e das estruturas administrativas. É necessário promover uma cultura institucional que reconheça que as leis não são neutras, que a jurisprudência não é objetiva, que os direitos não se realizam no vazio e que devem realizar-se na resolução prática de conflitos e lesões aos direitos fundamentais. O enfrentamento da discriminação estrutural passa pela politização das estruturas jurídicas, pelo rompimento com a ideia de neutralidade e pela afirmação da interseccionalidade como guia normativo (Fineman, 2023, p. 1399).

Sob essa ótica, decisões judiciais que reconhecem a interseccionalidade como critério relevante para a análise da violação de direitos são passos fundamentais. O Supremo Tribunal Federal, em algumas ocasiões, tem avançado

nesse sentido, ao reconhecer que políticas públicas devem contemplar recortes de raça, gênero e classe para serem efetivas. A jurisprudência, portanto, torna-se espaço de disputa por significados, por reconhecimento e por justiça substantiva.

O compromisso com a vulnerabilidade como fundamento exige também a mobilização da sociedade civil. Movimentos sociais, organizações populares, coletivos de base têm historicamente atuado como protagonistas na denúncia da discriminação estrutural e na proposição de alternativas. A produção de conhecimento contra-hegemônico, a criação de espaços de escuta, a valorização das narrativas subalternizadas são estratégias que potencializam a construção de um Direito plural. O sujeito vulnerável, nesse cenário, não é apenas destinatário de normas, mas produtor de saberes e agente de construção política.

Ao integrar essas perspectivas, vê-se que a vulnerabilidade não é destino nem patologia. Ela é potência crítica. É a partir dela que se revela a falência da neutralidade, a insuficiência da igualdade formal e a urgência de transformação pelo direito. A vulnerabilidade antecede a discriminação porque revela a base humana comum que, ao ser capturada por lógicas de poder, se converte em hierarquia. Reconhecê-la é o primeiro passo para desconstruir os sistemas que dela se alimentam e construir caminhos para uma justiça verdadeiramente plural, empática e emancipada.

A articulação entre vulnerabilidade e discriminação estrutural, especialmente quando analisada sob uma perspectiva interseccional, revela não apenas os limites materiais da igualdade formal, mas as raízes profundas que sustentam práticas institucionais excludentes no Brasil. França e Nóbrega (2024, p. 11) analisam o caso “Empregados da Fábrica de Fogos de Santo Antônio de Jesus e seus Familiares vs. Brasil” (Corte IDH, 2020), submetido à Corte Interamericana de Direitos Humanos, e destacam como o julgamento representa um paradigma para a identificação da atuação negligente do Estado brasileiro diante de populações em situação extrema de vulnerabilidade interseccional.

A explosão da fábrica, que vitimou majoritariamente mulheres negras e pobres – muitas delas crianças e adolescentes – escancara as interconexões entre raça, gênero, classe social e idade na produção da desigualdade. A informalidade das contratações, a precariedade das condições de trabalho, a ausência de fiscalização adequada e a morosidade das investigações judiciais revelam um padrão de atuação

estatal que se mantém indiferente à gravidade da violência vivenciada por sujeitos historicamente marginalizados. A Corte reconheceu que essas pessoas não estavam expostas apenas à violação de direitos laborais e à negligência administrativa, mas também a um contexto estrutural de invisibilidade, no qual suas vidas foram precarizadas em nome de interesses econômicos locais (Corte IDH, 2020).

O reconhecimento pela Corte Interamericana de que houve violação múltipla de direitos – dentre eles o direito à vida, à integridade pessoal, aos direitos da criança, à igualdade perante a lei e às garantias judiciais – reforça a necessidade de compreender a discriminação não como ato isolado, mas como produto de uma lógica institucional que se perpetua – o que se verificou no caso “Trabalhadores da Fazenda Brasil Verde Vs. Brasil (Corte IDH, 2016), em que o Brasil foi condenado em questão afeita à discriminação estrutural em caso de trabalho análogo à escravidão –. A análise interseccional torna-se, nesse sentido, ferramenta essencial para revelar que as violações sofridas não se distribuem de forma homogênea, mas concentram-se sobre aqueles que carregam múltiplos marcadores sociais de exclusão (França e Nóbrega, 2024, p. 18). No caso de 2016, diante do padrão coletivo do trabalho em situação análoga à escravidão, assim sintetizou a discriminação estrutural:

> 1) um grupo ou grupos de pessoas que têm características imutáveis ou imodificáveis pela sua própria vontade ou que estão interligados a fatores históricos de práticas discriminatórias, podendo este grupo de pessoas ser uma maioria ou minoria; 2) grupos que estejam em uma situação sistemática e histórica de exclusão, marginalização ou subordinação, configurando um impedimento para que tenham acesso a condições básicas de desenvolvimento humano; 3) situação de exclusão, marginalização ou subordinação que se concentre em uma região geográfica determinada ou generalizada em todo o território de um Estado, podendo, em determinados casos, ser intergeracional; 4) que as pessoas pertencentes a estes grupos sejam vítimas de discriminação indireta ou de discriminação de fato, em virtude da conduta do Estado ou da aplicação de medidas ou ações por ele implementadas (Corte IDH, 2020, p. 189).

Assim, o caso dos Empregados da Fábrica de Fogos e dos Trabalhadores da Fazenda Brasil Verde (2016) ganham potência como dispositivo normativo crítico, mais do que meramente jurisprudências ou marcos decisionais. Ele evidencia que o descaso do Estado não se manifesta apenas na ausência de ações, mas também na naturalização de práticas que colocam vidas vulneráveis em risco sob o pretexto da informalidade econômica. Quando a interseccionalidade é negligenciada nas políticas públicas, o resultado não é apenas a exclusão, mas a institucionalização da morte

simbólica e material de sujeitos que não encontram, no sistema jurídico, espaço legítimo de escuta ou reparação. A sentença da Corte é, portanto, mais do que uma condenação: é um chamado à reconfiguração profunda das estruturas normativas e administrativas que sustentam a seletividade da justiça.

Como se viu, o nascedouro da discriminação estrutural passa por problemas associados a vícios na fruição da igualdade, tanto em sua face de princípio norteador de todo o sistema jurídico, quanto em seu aspecto jusfundamental de proteção individual e, nesses termos, ao não tratamento discriminatório. Isto é, à vedação ao tratamento discriminatório não justificado (Mello, 1993, p. 35). A faceta formal do princípio da igualdade não se confunde, nesse sentido, aos desdobramentos materiais da igualdade enquanto preceito e objetivo. Enquanto a igualdade pró-forma exige apenas uma não discriminação injustificada por parte do Estado, a disposição da igualdade material, a *Epieíkeia*[4] aristotélica, como justiça (material/distributiva) equitativa, é o lançar-se mão de ações que promovam um aumento dos substratos, equacionando-os (Aristóteles, 2001, p. 101-105). Ainda que essa lógica seja útil, precisamos compreendê-la e interpretá-la na contextualização que a complexidade da vida hodierna exige.

Como se percebe o termo “discriminação estrutural” carrega em si uma carga simbólica de difícil definição, o que demonstra sobretudo as possibilidades de sua caracterização no mundo fenomênico. As discriminações estruturais sofridas por grupos vulneráveis podem se dar por muitos elementos culturais e históricos, da própria formatação do status social em certa sociedade, como pode significar situações cujo nascedouro seja mais contemporâneo, implicando igualmente em gravosa discriminação estrutural. Leal e Alves (2023, p. 8) trazem, de voto do

[4] Interessante notar que o vocábulo “*ἐπιεικής*” – “*epieikḗs*” significa aquilo que é razoável, justo, com o sufixo “*-ια*” – “-ia”) que no grego transforma um adjetivo ou substantivo em ideias, geralmente associado como imanentes na natureza. Dessa forma, nota-se que ao falar de justiça distributiva (*epieíkeia*), direciona-se à defesa de que há um certo equilíbrio, uma ordem natural de coisas, que deve ser respeitada e buscada pelos humanos, sobretudo pelos juristas – guardiões e defensores dos direitos. Uma exemplificação notável do desdobramento desta noção aparece já na tradição ulpiana, que defendia que os preceitos da vida guiada pelo Direito eram viver honestamente, não lesar a outrem e dar a cada um o que é seu (*Iuris praecepta sunt haec: honeste vivere, alterum non laedere, suum cuique tribuere*). A isto, associado a célebre nota *Jus est ars boni et aequi* (o Direito é a arte do bom e do justo), também de Ulpiano, parecem traduzir muito bem os sentimentos associados à utilização do Direito pelo Direito – que já objetiva um justo, e não a subutilização instrumental do Direito para a persecução de interesses outros que não aqueles do próprio Direito ou de classes dominantes e que pretendem se estagnar na pirâmide social – não pela sua ascensão, mas pela garantia de manutenção da (i)mobilidade social.

eminente juiz Eduardo Ferrer Mac-Gregor Poisot, no Caso Trabalhadores da Fazenda Brasil Verde Vs. Brasil, elementos importantes para a identificação de uma discriminação estrutural, sem exigir a presença de todos eles cumulativamente, ou a eles se limitar.

> (I) A discriminação estrutural exige a presença de um grupo ou grupos de pessoas, tratando-se de uma situação coletiva. O grupo deve apresentar (a) características imutáveis ou imodificáveis por vontade própria (b) ou que estejam relacionadas a fatores históricos de práticas discriminatórias; (c) e que este grupo seja de minorias ou maiorias.
> (II) Que esse grupo se encontre em uma situação sistemática e histórica de exclusão, marginalização ou subordinação que lhes impeça de alcançar as condições básicas de desenvolvimento humano. Este é o núcleo central da noção de discriminação estrutural: constatar uma situação generalizada de desvantagem, exclusão, subordinação, marginalização ou submissão como obstáculo ao gozo dos direitos fundamentais.
> (III) Que a situação de exclusão, marginalização ou subordinação se centre em uma zona geográfica determinada, ou generalizado em todo o território do Estado, podendo ser ainda uma situação intergeracional.
> (IV) Que a discriminação estrutural afete grupos de vítimas de discriminação indireta ou de discriminação de fato, por atuações, aplicações de medidas ou ações implementadas pelo Estado. Ou seja, independe da intenção da norma, da neutralidade ou da menção expressa de alguma distinção ou restrição explícita, devendo ser apurado se há violação à igualdade material para determinado grupo vulnerável (Corte IDH, 2016).

Há uma certa amplitude com conceitual a abarcar situações que de fato passam de vulnerabilidades circunstanciais, de condições naturais ou morais dos indivíduos (ainda que essas sejam “razões” para a prática ou perpetuação de preconceitos), alcançando uma situação de verdadeira estratificação excludente. Assim, o nascedouro da discriminação estrutural passa por problemas associados a vícios na fruição da igualdade, tanto em sua face de princípio norteador de todo o sistema jurídico, quanto em seu aspecto jusfundamental de proteção individual e, nesses termos, ao não tratamento discriminatório. Isto é, à vedação ao tratamento discriminatório não justificado.

Tem-se o seguinte fato: tais valores decorrentes da igualdade nem sempre foram reafirmados de modo efetivo (ao menos não meramente retórico) através dos tempos. Há uma inegável desigualdade, tanto material quanto de condições formais, em uma organização jurídico-estatal que, mesmo com ações afirmativas, não tem se mostrado apta a corrigir tal situação. É que a discriminação estrutural, diferente daquela experimentada por uma discriminação conscientemente provocada, interfere de maneira indireta em amplos campos da vida das pessoas que ficam sujeitas à

situação de uma constante e ampla estrutura que desempenha um papel contrário aos seus interesses e à fruição de seus direitos fundamentais.

Há uma sistemática manutenção da discriminação estrutural em que cidadãos são subjugados por condições comunitárias específicas (Saba, 2005, p. 126), que tanto “justificam” a posição de desvantagem quanto a retroalimentam e, nesse ponto – como se verá mais adiante – as aplicações algorítmicas são uma verdadeira indústria de reprodução das discriminações estruturais. Nesse sentido, Sagüés (2018) propõe que a desigualdade sistemático-estrutural seja atacada por uma proposta de transformação social na instância superior da igualdade estrutural, permitindo que decorra desta concepção a necessária tutela especial a grupos desavantajados, sobretudo das ações afirmativas – ainda que em um necessário caráter excepcional e transitório – pensado para que venha a inexistir com a eficácia das medidas no tempo.

Assim, ainda que toda a discussão em campo filosófico dogmático seja essencial para estabelecer os meandros e o significado normativo do conteúdo envolvido nas múltiplas e complexas decorrências de uma necessária afirmação da igualdade e da vedação à discriminação, faz-se extremamente necessário volver-se os olhos à prática dos litígios estruturantes – casos em que o Poder Judiciário deverá contar com o auxílio de órgãos de diferentes segmentos da sociedade e da vida pública para tentar endereçar uma solução que transcenda à solução individual do caso em liça, mas que consiga atingir o núcleo estrutural daquela situação discriminatória (ou, então, causada pela própria discriminação – e identificar de que forma e em que medida os mecanismos processuais permitem a concretização da tal almejada igualdade nos termos retratados.

Em primeiro lugar, é o Estado quem deve criar as condições favoráveis para a fruição e o alcance do núcleo de proteção dos direitos fundamentais, como é o caso da igualdade. Contudo, em não sendo feito isso pela via legislativa ou mesmo constitucional, faz-se necessária a possibilidade de o Judiciário intervir. E o poderá fazer tanto no vácuo protetivo quanto na modalidade repressiva, para reparar inconstitucionalidades e ilegalidades (mas daí amparados pelos limites normativos já estabelecidos).

E, para tanto, a discussão filosófica dogmática é crucial para decifrar as nuances e o verdadeiro significado normativo da igualdade e da proibição da

discriminação. No entanto, é igualmente vital voltar a atenção para a prática jurídica e analisar como os mecanismos processuais contribuem para efetivar a igualdade, especialmente diante das falhas na valoração da dignidade humana na sociedade contemporânea. O Estado tem o dever de criar condições para a efetiva proteção dos direitos fundamentais, como a igualdade. Na ausência de ação legislativa ou constitucional adequada, é essencial que o Judiciário possa intervir, seja preenchendo lacunas protetivas ou agindo repressivamente contra inconstitucionalidades e ilegalidades, respeitando os limites normativos existentes.

É possível, nesse sentido de vácuo protetivo, a aplicação de instrumentos de controle de inconstitucionalidade por omissão nos casos em que, diante da existência de discriminação estrutural, o Estado omite a adoção de medidas de não discriminação, como resposta à realidade discriminatória. Desse modo, é importante se lembrar que as medidas de ações afirmativas que consistem em práticas não discriminatórias começam a partir de uma suposição fática fundamental de natureza coletiva: age contra pressupostos de natureza coletiva, um grupo sistematicamente excluído, subjugado ou marginalizado, que podem, mas nem sempre, ser historicamente discriminados – afro-americanos, mulheres, pessoas com deficiência e assim por diante (Sagüés, 2018).

A natureza coletiva, difusa e transindividual da discriminação estrutural mostra que as ações afirmativas criadas para combater a desigualdade assumem também características coletivas. Essa proteção não é apenas individual, mas deriva de um dever amplo de garantir direitos, indo além de simples reivindicações individuais fundamentais. Isso não impede que problemas estruturais sejam identificados inicialmente em nível individual para, posteriormente, serem tratados de forma coletiva (Sagüés, 2018). Os procedimentos constitucionais, especialmente em questões estruturantes, devem considerar que o processo é parte essencial dos direitos protegidos. Para que direitos fundamentais se tornem realidade, é essencial que o Estado atue na erradicação da pobreza e na redução das desigualdades, buscando continuamente esses objetivos, sem limitar-se a uma postura passiva.

É logicamente viável, sob um ponto de vista sociológico, que, na ausência da possibilidade ou da execução, por um dos poderes, de medidas aptas a realizar essa transformação, que outro poder seja legitimado a tomar a frente e resolver tais questões (Sagüés, 2018). É o campo de trabalho, daí, portanto, do Poder Judiciário,

que se apropria da defesa dos direitos fundamentais a partir destas necessidades, e, atualmente, com uma renovada atuação frente aos desafios da discriminação algorítmica.

Diante das situações de discriminação por vulnerabilidade, pertencimento a grupos vulneráveis ou discriminação estrutural, os poderes da nação devem proteger os indivíduos das agressões injustas aos seus direitos, cabendo ao Poder Judiciário interpretar os núcleos de significação dos textos normativos provenientes dos demais poderes, bem como tutelar por último a correta, proporcional e adequada resposta aos injustos percebidos. Na realidade da inteligência artificial, há um vácuo normativo acerca da temática – insuficiência, fragmentação e assimetria informacional no núcleo prático da *quaestio juris* –, mas as injustas agressões e violações a direitos fundamentais não esperam pela legislação para conhecer o adequado tratamento jurídico, sendo cabível ao Poder Judiciário dar, em caráter tópico e jurisprudencial, respostas cabíveis e parâmetros racionais de interpretação e aplicação do Direito às demandas envolvendo discriminação algorítmica. Considerando a complexidade da vulnerabilidade e da discriminação estrutural, é fundamental levar em conta esses elementos para abordar a discriminação algorítmica, que atinge especialmente os grupos vulneráveis, a partir de interseccionalidades, ou seja, múltiplas origens e motivações injustificadas que trazem severas desproporções entre a eficácia dos direitos fundamentais e sua fruição nas circunstâncias fáticas do cotidiano. O Direito deve reequilibrar proporcional e gradativamente essa balança, trazendo clarões de garantias fundamentais.

À luz da Constituição, o dever estatal de proteção (Leal; Mass, 2020) contra a discriminação estrutural possui assento normativo expresso e densidade material: a dignidade da pessoa humana (art. 1º, III), os objetivos fundamentais de construir uma sociedade livre, justa e solidária e de reduzir desigualdades (art. 3º, III) e a cláusula geral de igualdade (art. 5º, caput) – lidos sob a ótica da igualdade material – impõem prestações positivas de prevenção, mitigação e reparação de assimetrias. Quando a ineficácia legislativa ou administrativa perpetua padrões de exclusão, abre-se o espaço do controle por omissão – com técnicas de correção normativa e jurisdicional –, de modo a assegurar que a neutralidade formal não se converta em conivência institucional com a desigualdade.

Compreender a discriminação estrutural em sua forma analógica é fundamental para reconhecer as raízes históricas e culturais que perpetuam a desigualdade de modo entranhado e inclusive de modo inconsciente na sociedade contemporânea. No entanto, à medida que avançamos para uma era digital, é crucial explorar como essas mesmas dinâmicas de discriminação se manifestam de modo intensificado e podem intensificar a própria discriminação estrutural através da tecnologia. A seguir, necessário se faz o debruçar acerca da discriminação algorítmica, onde os preconceitos e desigualdades são codificados em sistemas digitais, perpetuando injustiças de maneira automática e muitas vezes invisível, o que deve ser primariamente combatido. Examinar-se-á como os algoritmos, embora projetados para serem neutros, podem refletir e amplificar as desigualdades e preconceitos estruturalmente existentes. Ainda que frequentemente sejam apresentados como neutros, a neutralidade, aqui, é hipótese a ser demonstrada, e não um atributo sempre presumido.

2.3 A renovada dimensionalidade da igualdade no contexto das decisões automatizadas em cenário cibernético e os desafios para sua concretização em casos de grupos vulneráveis

Os autômatos de Heron no templo – portas que se abrem “sozinhas” após o sacrifício, por uma coreografia de ar aquecido, sifões e contrapesos – mostram como a opacidade intencional produz efeitos de autoridade e assombro. No mundo digital, a mesma estrutura simbólica opera sobre grupos vulneráveis: decisões opacas parecem “naturais”, quando são design normativo-técnico. Da liturgia antiga ao *score* algorítmico, a autoridade se ancora na assimetria de informação – e, apenas assim – perpetua desigualdades (Chondros, 2017).

Na era digital, a ascensão da automação da tomada de decisões, por meio de algoritmos, trouxe à tona uma nova forma de discriminação: a algorítmica, que renova e densifica a discriminação estrutural a partir da automatização de vieses discriminatórios historicamente difundidos e agora escaláveis em velocidade algorítmica. Essa prática pode, inadvertidamente ou intencionalmente perpetuar desigualdades e injustiças humanas, frequentemente associadas a estereótipos desabonadores. Como resultado, questões sobre as implicações sociojurídicas da

discriminação algorítmica têm ganhado destaque e demandado uma análise aprofundada por diversos setores, dentre os quais está o jurídico.

A discriminação algorítmica, uma questão emergente e de grande importância no mundo contemporâneo, permeia todos os aspectos de nossas vidas. A tecnologia, que é uma ferramenta poderosa para o progresso, também pode ser uma fonte de desigualdade e injustiça quando mal utilizada ou mal compreendida em seus processos e resultados. Este estudo busca colocar mais um tijolo na busca da construção semântica e epistêmica acerca tanto da compreensão do fenômeno da discriminação algorítmica quanto das respostas atribuíveis pelo direito, seja pela regulação ou pela jurisprudência da jurisdição constitucional.

A discriminação algorítmica é um desafio multifacetado que exige atenção à proteção dos direitos fundamentais, especialmente com o avanço da Inteligência Artificial na sociedade. Para enfrentá-la, é crucial compreender as causas dos resultados discriminatórios, que podem decorrer tanto da falta de diversidade entre desenvolvedores quanto da reprodução autônoma de preconceitos sociais pelos algoritmos.

Sobre a falta de diversidade, isto se deve ao fato de que, quando as equipes são muito homogêneas, surgem “pontos cegos” no desenho do sistema: como o problema é definido, quais dados entram, que métricas importam e como se testa o produto. O que não é vivido, muitas vezes não é conhecido, e o que não é conhecido dificilmente será reparado. O melhor exemplo é o apresentado no documentário da Netflix, *Coded Bias* (2020), no qual os sistemas de reconhecimento facial performavam pior para mulheres negras, ainda que a sua desenvolvedora fosse uma, porque os dados eram pouco representativos para o treino da aplicação. O viés entra na curadoria do *dataset*, na escolha de métricas e na validação final do produto.

Sobre a reprodução automatizada dos preconceitos, ocorre que mesmo com boas intenções, o algoritmo aprende padrões do passado e, se os dados históricos carregam desigualdades, o modelo as trata como “regras do mundo” e as otimiza. Além disso, há alças de retroalimentação: a decisão do sistema produz novos dados que confirmam o próprio viés. Um excelente exemplo disto são os modelos de “policiamento preditivo” treinados em registros de abordagens passadas, cujo viés é o de enviar mais patrulhas a bairros já muito vigiados; isso gera mais registros nesses

locais (não necessariamente mais crime real), “provando” ao modelo que deve voltar ali. Cria-se um ciclo viciado e intensificado.

O objetivo agora é delinear um panorama acerca do problema (discriminação), a causa imediata (algoritmos) e o campo de possibilidades jurídico-constitucionais (direitos fundamentais) a enfrentar a questão.

Aos contextos supramencionados, inserem-se as ferramentas de inteligência artificial, hodiernamente utilizadas em diversas áreas e aplicações. O seu uso é derivativo de uma narrativa talvez imprecisa, na qual a sua utilização facilitaria a economia de tempo e energia – ainda que o ponto seja controverso, sobretudo pelos conhecidos danos ambientais aos recursos hídricos causados pelos *data centers* –. Fato é que tais ferramentas nada mais são do que reprodutoras das estruturas sociais vigentes, inserindo-se, a partir daí, as considerações discriminatórias da sociedade. Quanto à aplicação dessas ferramentas, há uma defesa na comunidade científica de que os sistemas de inteligência artificial operam com base em parâmetros objetivos e matemáticos, aparentemente imunes às inclinações e subjetividades humanas. Contudo, essa pretensa imparcialidade não se sustenta na prática, uma vez que se verifica que os algoritmos de inteligência artificial, longe de corrigirem os vieses inerentes às ações humanas, acabam por potencializar esses tratamentos desiguais em um espectro mais amplo. Tal fenômeno tem sido denominado como discriminação algorítmica (Requião; Costa, 2022, p. 4).

Ademais, é fundamental reconhecer que a discriminação algorítmica emerge não apenas como uma reprodução de preconceitos preexistentes, mas também como um mecanismo que os consolida e perpetua em estruturas digitais. Esse processo se dá pela introdução de dados enviesados nos sistemas ou pela configuração de modelos que, mesmo inconscientemente, priorizam determinados grupos em detrimento de outros. Assim, a tecnologia, que deveria atuar como ferramenta de promoção da equidade, torna-se, paradoxalmente, um instrumento de amplificação das desigualdades estruturais. A problemática da discriminação algorítmica exige uma análise crítica que transcenda a simples implementação de soluções técnicas. É necessário articular esforços multidisciplinares para entender as raízes históricas e sociais que alimentam esses vieses, bem como desenvolver respostas efetivas e coerentes, escaláveis em âmbito jurídico-protetivo.

Portanto, na ausência de correção dos preconceitos discriminatórios, observa-se uma propensão à perpetuação desses tratamentos, de modo que os indivíduos pertencentes a grupos em situação de vulnerabilidade, sujeitos à discriminação estrutural, estão mais propensos a serem impactados por essa forma digital de exclusão (Crestane, Leal, 2024, p. 77). Esse fenômeno decorre do fato de que as tecnologias, por si só, não carregam discriminações; o que as torna instrumentos reprodutores de preconceitos é a intervenção humana, seja por meio da inclusão de vieses subjetivos na programação dos algoritmos ou pela utilização de conjuntos de dados tendenciosos como base para a operação desses sistemas. As máquinas em sua essência não discriminam, mas sim os seres humanos. Os impactos das tecnologias operadas por inteligência artificial revelam um paradoxo inquietante: enquanto deveriam ser ferramentas de equidade e neutralidade, frequentemente se tornam agentes amplificadores das desigualdades (Moreira et al., 2023, p. 1955). Esses sistemas, ao serem alimentados por dados historicamente enviesados ou configurados por modelos com inclinações implícitas, consolidam narrativas excludentes que afetam negativamente os direitos de grupos sub-representados.

Nesse cenário, Requião e Costa (2022, p. 4) identificam duas formas distintas pelas quais a discriminação algorítmica pode se originar: **i.** aquela que resulta da inserção, consciente ou inconsciente, de preconceitos no momento da programação, refletindo as crenças ou vieses do programador; e **ii.** aquela derivada da utilização de bancos de dados enviesados, que transportam em seu interior informações contaminadas por estigmas sociais preexistentes. Em ambas as hipóteses, a inteligência artificial não apenas reproduz discriminações estruturais, mas também as amplifica, sobretudo por operar sob o manto de uma suposta neutralidade e precisão técnica que lhe confere aparência de legitimidade e imparcialidade (Silva; Barbosa, 2024, p. 9).

Quanto à primeira forma, é necessário reconhecer que, embora os avanços da inteligência artificial tenham permitido que determinados algoritmos sejam capazes de autogerar partes de seus próprios códigos, o desenvolvimento inicial e a lógica fundamental desses sistemas permanecem ancorados na intervenção humana. Assim, preconceitos, ainda que sutis ou inconscientes, podem ser transpostos para o ambiente digital sem que haja intenção discriminatória manifesta. Como já abordado, a discriminação pode emergir, não da vontade deliberada de excluir, mas da repetição

acrítica de padrões sociais excludentes que estruturam o cotidiano – padrões esses internalizados pelo programador e replicados nos sistemas que concebe (Crestane; Leal, 2024, p. 76).

No que tange à segunda forma de discriminação algorítmica, destaca-se o papel central da lógica de aprendizado de máquina (*machine learning*), segundo a qual os algoritmos ajustam seus parâmetros de maneira autônoma, a partir do processamento de grandes volumes de dados, com o objetivo de identificar padrões e realizar previsões ou decisões; não são algoritmos integralmente programados, o viés não está sob o total controle do programador (Ferrari, 2021 p. 77). Quando os dados de entrada utilizados nesse processo são enviesados, ou seja, carregam distorções históricas e sociais, o algoritmo não apenas deixa de corrigir tais distorções, como as absorve como pressupostos de validade. O resultado é a retroalimentação desses vieses, consolidando-os como base normativo-programática para o desempenho futuro do sistema (Anarte, 2023, p. 105).

A despeito da aparência de racionalidade técnica que envolve os sistemas baseados em inteligência artificial, diversos exemplos concretos demonstram que seus impactos práticos podem reforçar padrões discriminatórios já enraizados na sociedade. Em 2018, a Amazon foi alvo de críticas por utilizar um algoritmo de recrutamento que penalizava candidatas mulheres. O sistema foi treinado com dados históricos de admissões majoritariamente masculinas e, ao replicar esse padrão, desconsiderava as qualificações individuais das mulheres, classificando-as automaticamente como menos aptas (Borges; Filó, 2021, p. 233). Em outro caso emblemático, um software de reconhecimento facial nos Estados Unidos resultou na prisão injusta de Robert Williams, um homem negro, cuja imagem foi erroneamente associada à de um suspeito, exigindo o pagamento de fiança após trinta horas detido, apesar das evidências visuais contrárias (Silva, 2022, p. 110). Esses episódios evidenciam como decisões automatizadas, quando alimentadas por dados enviesados ou insuficientes, produzem discriminações reais e materialmente graves.

Esses resultados discriminatórios decorrem, em muitos casos, de falhas estruturais nos próprios processos de alimentação e desenvolvimento dos sistemas de inteligência artificial. Tischbirek (2020, p. 104) classifica as causas da discriminação algorítmica em três categorias: **i.** falhas na coleta de dados; **ii.** falhas no tratamento dos dados; e **iii.** indiferença normativa. Entre essas, destaca-se a

primeira como a mais recorrente, já que conjuntos de dados de treinamento com baixa representatividade ou vieses históricos induzem os algoritmos a reproduzirem padrões tendenciosos. Como observa Richardson (2022, p. 91), essa deficiência compromete especialmente grupos já marginalizados, que passam a experimentar uma continuidade da exclusão – agora mediada por sistemas automatizados – a partir da repetição e legitimação das desigualdades incorporadas aos dados de origem.

Além dos vieses presentes na programação ou nos dados, outro fator determinante para a produção de discriminações algorítmicas encontra-se no próprio modelo estatístico adotado pelos sistemas. Doneda, Mendes e Andrade (2018, p. 5) destacam o risco representado pelo funcionamento dos algoritmos com base em categorizações estatísticas: indivíduos com características semelhantes são agrupados e tratados como representações de uma média, o que acarreta decisões automatizadas fundadas em predições de comportamento coletivo. Tal mecanismo favorece discriminações por erro estatístico – seja pela baixa qualidade dos dados, seja pela fragilidade científica do modelo aplicado. Além disso, esse modo de operação abre espaço para generalizações injustas: ainda que uma predição esteja estatisticamente correta, ela pode desconsiderar casos atípicos, como o de uma pessoa de alta renda que reside em área popular e é automaticamente classificada como de risco elevado para concessão de crédito apenas com base em sua localização.

A lógica que sustenta a inteligência artificial, embora tecnicamente sofisticada, não escapa às armadilhas dos preconceitos históricos que estruturam a sociedade. Isso porque os dados que alimentam os sistemas de tomada de decisão não surgem em um vácuo de neutralidade, mas refletem trajetórias passadas marcadas por desigualdades e exclusões. Como enfatizam Duarte e Negócio (2021, p. 230), na ausência de mecanismos efetivos de filtragem desses vieses históricos, os algoritmos incorporam discriminações preexistentes como parâmetros válidos para decisões futuras. Assim, a discriminação algorítmica manifesta-se como extensão direta da discriminação estrutural, operando por meio de falhas sistemáticas e excludentes. Crestane e Leal (2024, p. 40) esclarecem que tal discriminação se caracteriza quando tais falhas, analisadas à luz da legislação antidiscriminatória, se ajustam aos conceitos legais de exclusão e desigualdade.

Diante dessa realidade, torna-se indispensável reconhecer que a discriminação algorítmica não apenas reproduz os estereótipos contidos nos dados ou nos pressupostos dos desenvolvedores, mas pode também amplificar injustiças sociais por meio da propagação de conteúdos danosos, como *fake news* e discursos de ódio. Cambi e Amaral (2023, p. 197), ao abordarem essa ambivalência, apontam que a inteligência artificial possui uma natureza dual: ao mesmo tempo em que facilita a perpetuação de preconceitos, também pode se tornar um instrumento de superação dessas desigualdades, enquanto graus de controlabilidade e de mitigação da opacidade. Desde que corretamente estruturadas – sob deveres verificáveis de documentação, mitigação de vieses, rastreabilidade e auditabilidade proporcionais ao risco apresentado – tais ferramentas podem reforçar os direitos fundamentais de forma mais objetiva e eficaz do que iniciativas que buscam mudanças culturais lentas e difusas no tecido social.

É justamente o caráter sistêmico da discriminação algorítmica que a aproxima da discriminação estrutural: ambas operam de modo difuso, automatizado e insidioso, dificultando sua identificação imediata e tornando-a invisível sob a aparência de normalidade institucional. É nesse sentido que Costa (2022, p. 168) adverte para a necessidade urgente de regulamentos que exijam transparência nos processos decisórios automatizados, a fim de possibilitar sua fiscalização e responsabilização. Afinal, ainda que as máquinas executem tarefas com eficiência superior à humana, não se pode perder de vista que são produtos de decisões humanas – de programadores, de empresas, de arquiteturas políticas e econômicas – e, como tais, devem se submeter aos limites impostos pela Constituição e ao respeito inegociável aos direitos fundamentais (Ehrhardt Junior; Silva, 2021, p. 21).

A atuação do Poder Judiciário torna-se central diante do desafio representado pela discriminação algorítmica, especialmente quando os sistemas de inteligência artificial passam a afetar diretamente o acesso e a fruição de direitos fundamentais por grupos vulnerabilizados. A responsabilidade jurisdicional, nesse cenário, não se esgota na simples aplicação das normas vigentes, mas exige uma postura proativa na construção de parâmetros jurídicos que orientem o uso responsável e eticamente justificado dessas tecnologias. Cambi e Amaral (2023, p. 201) destacam que o Regulamento Europeu de Inteligência Artificial, ao tratar de sistemas classificados como de “alto risco”, impõe a exigência de supervisão humana – constante do art. 14,

§ 2º – justamente para evitar que decisões com impacto sobre a vida de indivíduos sejam automatizadas sem controle. A partir dessa diretriz, torna-se possível afirmar que a jurisdição constitucional tem um papel indispensável na prevenção dos efeitos deletérios da automatização opaca e discriminatória.

Entretanto, para que esse papel institucional seja cumprido com eficácia, é necessário que o Judiciário acompanhe criticamente o avanço das tecnologias digitais e suas aplicações no próprio sistema de justiça. Há experiências internacionais que revelam riscos concretos: um exemplo emblemático é o uso do software COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*) nos Estados Unidos. Trata-se de uma ferramenta empregada para avaliar a probabilidade de reincidência de réus, cujos dados – analisados pela organização ProPublica em 2013 e 2014 – indicaram vieses raciais significativos: indivíduos negros eram mais frequentemente classificados como de “alto risco” do que brancos em situações semelhantes (Vieira, 2019, p. 31). Esse caso revela como a ausência de transparência nos critérios adotados por sistemas algorítmicos pode gerar decisões discriminatórias travestidas de neutralidade estatística, reiterando a urgência de uma supervisão humana robusta.

Por trás desses riscos está um dos problemas mais sensíveis dos sistemas algorítmicos: sua opacidade estrutural. Tais sistemas operam frequentemente como verdadeiras “caixas-pretas”, nas quais os processos decisórios são complexos, codificados em linguagens técnicas inacessíveis e, muitas vezes, indisponíveis para revisão externa. Astobiza (2017, p. 214) adverte que essa falta de transparência é frequentemente acompanhada de uma perigosa resignação, como se os efeitos discriminatórios derivados dessa lógica técnica fossem inevitáveis ou aceitáveis. Para o autor, uma medida indispensável para mitigar esse risco seria a proibição de decisões automatizadas com impactos diretos sobre os indivíduos sem a possibilidade de supervisão final por um ser humano. Em última análise, a exigência de transparência e controle humano não representa apenas uma salvaguarda jurídica, mas uma exigência ética diante da complexidade e do poder dessas tecnologias.

Nesse contexto, a institucionalização de mecanismos jurídicos voltados à regulação das tecnologias digitais revela-se indispensável para garantir que a proteção de direitos fundamentais acompanhe as transformações impulsionadas pela inteligência artificial e pelo tratamento massivo de dados. Trata-se de reconhecer que

a opacidade algorítmica e os riscos de discriminação dela decorrentes não são efeitos colaterais fortuitos, mas expressões estruturais de uma nova racionalidade tecnocrática, que precisa ser confrontada por um arranjo normativo capaz de frear sua expansão arbitrária. Nesse sentido, a emergência de um regime jurídico próprio para a proteção de dados pessoais no Brasil constitui um marco importante nesse esforço de contenção e responsabilização. A Lei Geral de Proteção de Dados (LGPD), centrada na proteção destes dados que fomentam o uso das Inteligências Artificiais – o combustível para o processamento de dados e tomada de decisão –, a seguir delineada, insere-se precisamente nessa tentativa de reorganizar o campo jurídico frente à centralidade dos dados nas dinâmicas sociais, políticas e econômicas contemporâneas.

A LGPD inaugura no Brasil um regime geral de proteção de dados pessoais. Ainda que não implemente grandes mudanças fáticas no quesito proteção aos dados dos usuários, porquanto esses já vinham sendo tutelados, como visto, pelo microssistema existente em 2018, a LGPD consolida-se enquanto um importante marco normativo-protetivo, reacendendo a importância estratégica dos dados para as maiores empresas do mundo, como a Apple, o Facebook, o Google, Amazon etc. O Facebook em especial preocupa, sobretudo desde a aquisição de apps como WhatsApp e Instagram (Carugati, 2018, p. 5). Proteger o usuário de virar objeto de consumo das empresas de *Big Data*, marketing e que se baseiam em behaviorismo virou, portanto, um compromisso que assume esferas comerciais, constitucionais, negociais e, também, do direito do consumidor (Mendes; Doneda, 2018, n.p.). Segundo Doneda (2021, p. 10), a vigência da LGPD, a parte de tudo que trouxe, serviu como organizadora de toda a fenomenologia do direito acerca da proteção de dados, servindo como elemento introdutor da retomada da preocupação acerca da proteção dos dados pessoais, ainda que ela estivesse presente há algum tempo.

A LGPD visa à tutela e à proteção, dentre outros, da autodeterminação, ao desenvolvimento da personalidade e, também, ajudar a promover a defesa do usuário, que pode sofrer discriminações de múltiplas origens e de diversos modos. A proteção de dados tem por objetivo a proteção dos direitos fundamentais de liberdade e de privacidade, assim como o livre desenvolvimento da personalidade da pessoa natural e a igualdade na dimensão de sua não-discriminação, e a sua regulação, segundo o texto legal da LGPD, consiste em proporcionar ao cidadão garantias em relação ao

uso dos seus dados, a partir de princípios, de direitos do titular de dados e de mecanismos de tutela idealizados tanto para a proteção do cidadão quanto para que o mercado e setor público possam utilizar esses dados, dentro de parâmetros e limites.

Extrai-se, portanto, a partir da análise do caput do art. 1º da Lei 13.709/18, a intenção do legislador infraconstitucional em preservar não só o direito fundamental à privacidade e à intimidade dos titulares de dados pessoais, consagrado expressamente no art. 5º, X, da CF/88, mas igualmente fortalecendo a autodeterminação informativa individual. Assim, o direito à autodeterminação firma-se na possibilidade de decisão pessoal sobre os dados que lhe é de titularidade, sejam eles sensíveis ou não. Além disso, a lei possibilita o tratamento desses dados pessoais por terceiros, ressalvadas as hipóteses do art. 4º da LGPD.

Para os fins da LGPD, o tratamento de dados pessoais compreende toda a operação realizada com dados pessoais, como as que se referem a coleta, produção, recepção, classificação, utilização, acesso, reprodução, transmissão, distribuição, processamento, arquivamento, armazenamento, eliminação, avaliação ou controle da informação, modificação, comunicação, transferência, difusão ou extração dos dados pessoais. A inteligência é da disposição do art. 5º, X, da LGPD.

A informação de dados, na sociedade da informação, é o ativo mais valioso, uma vez que influencia, prevê e manipula o comportamento do próprio indivíduo titular dos dados, que perde controle de si mesmo, por causa do mal uso de seus dados (Doneda, 2019, p. 169). O fortalecimento da proteção de dados é indispensável para a adequada consolidação da proteção dos direitos fundamentais, portanto (Rodotà, 2008, p. 21). Para que isso seja de fato assegurado, é necessária uma estrutura principiológica densa. Veja-se, os dados pessoais são monetizados, e são utilizados para determinar a vida de bilhões de indivíduos ao redor do globo. Assim, não apenas normas protetivas devem ser asseguradas e fielmente cumpridas, mas o escopo interpretativo que serve de guia em casos não claros, deve ser fortemente asseverada. Nesse sentido, a LGPD ostenta uma estrutura principiológica densa, e até repetitiva. Desta forma, traz consigo um rol de princípios que precisam ser observados, na sua aplicação, por ocasião do tratamento de dados (Pinheiro, 2018, p. 24).

O art. 2º, inciso VII, da LGPD disciplina a proteção dos dados pessoais a partir do fundamento dos “direitos humanos, do livre desenvolvimento da personalidade, da

dignidade e o exercício da cidadania pelas pessoas naturais". O art. 6º da LGPD, dedicado aos princípios, colaciona os princípios da boa-fé, da finalidade, da necessidade, da adequação, do livre acesso, da qualidade dos dados, da prevenção, da transparência, da não discriminação, da segurança e da responsabilidade e da prestação de contas. Individualmente, o princípio da boa-fé é importante porquanto determina a adoção de condutas corretas e adequadas que todos os participantes da cadeia de tratamento dos dados deverão obedecer. De certa forma, os demais princípios são decorrências lógicas e especificadas do princípio da boa-fé. O princípio da boa-fé, nesse sentido, impõe um valor ético-moral substancial ao trato com os dados pessoais, de modo a coibir conceitualmente condutas contrárias ao escopo de proteção da norma; o qual pode-se colocar como o da prevenção da mercantilização dos dados pessoais, com o fim de garantir a autodeterminação informativa individual (Miragem, 2019, n.p.).

Pulando-se aos pontos chaves, o princípio da finalidade específica dos dados significa que o tratamento precisa ter um resultado único, específico e legítimo que deve ser alcançado, e que o mesmo deve ser o motivo primário para a solicitação dos dados, evitando-se desvirtuações. Também deverá garantir o livre acesso do titular dos dados a eles, tendo poderes de portabilidade, exclusão, retificação ou inclusão de seus dados, mantendo a qualidade dos dados, que devem ser transmitidos de modo claro e exato entre o titular e o operador, da transparência no tratamento de dados, para que se garantam informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, de segurança na utilização de medidas técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão dos dados, isto é, à integridade dos dados. Também deverá atender aos princípios de prevenção através da tomada de medidas que previnam a ocorrência de danos em virtude do tratamento de dados pessoais.

Segundo a LGPD, não poderá haver discriminações ilícitas ou abusivas dos dados coletados, com as técnicas de perfilização, p.e., utilizadas para restringir crédito, mesmo com os CPFs não negativados nas listas de proteção ao crédito. Sobre essa análise de risco, a LGPD traz em seu art. 11, §5º, que os planos de saúde estão proibidos de prever o risco do usuário para a contratação ou exclusão, e o mesmo aplica-se aos bancos de dados de proteção ao crédito, que avançam na questão da

perfilização do risco individual (*profiling*). E, ainda, há a preocupação acerca da responsabilização e prestação de contas (*accountability*), de modo que o agente de tratamento adote medidas eficazes de comprovar a observância e o cumprimento das normas de proteção de dados, inclusive as de segurança da informação, demonstrando a eficácia das medidas (Miragem, 2019, n.p.). Nesse sentido, o titular tem o direito de solicitar a revisão de decisões automatizadas que afetem seus interesses, conforme previsto na LGPD, visando a transparência no tratamento de dados. Embora a lei reconheça o direito ao sigilo comercial e industrial, a ANPD (Agência Nacional de Proteção de Dados) pode realizar auditorias para verificar discriminações em tratamentos automatizados, buscando a compatibilização entre essas preocupações; habilita tal possibilidade, mas não demonstra qualquer lógica de proteção aos grupos vulneráveis em uma dimensão de tutela à discriminação estrutural. Há um vácuo protetivo específico.

Uma questão premente, de ordem mais técnico-jurídica, se relaciona à necessária ponderação que deverá ser feita a partir da colisão do direito difuso à publicidade, quando aplicável, sobretudo nos casos de informações públicas, como as processuais ou sobre saúde pública, e o direito individual à proteção de dados pessoais. Veja-se o caso do IBGE (RmCADI 6.389), no qual os dados levantados eram de enorme interesse público, porquanto poderiam auxiliar na prevenção da doença COVID-19, o que atende ao interesse público. A LGPD inclusive traz a questão do interesse público enquanto escopo interpretativo para a indisponibilização dos dados, no art. 4, §1º, art. 7, §3º, art. 15, III e, na seara pública, no art. 23. Assim, a lei traz possibilidades previstas para a relativização da proteção máxima aos dados.

Soares (2020, p. 31) traz que, por serem os direitos fundamentais organizados sob a estrutura de princípios, devendo ser aplicados por otimização em casos, e, sendo o direito à privacidade, à proteção de dados ou à autodeterminação informativa, direitos fundamentais, admitiram relativização frente a outros que, por algumas questões, apresentassem grau satisfativo em maior importância. Para solucionar esse entrave deve ser tomada a utilização da estrutura do teste da proporcionalidade de Alexy. Com lucidez, Soares (2020, p. 31) traz ainda que aqui, deve-se ter a cautela de diferenciar entre o interesse público genuíno e a utilização do fundamento do “bem comum” como forma de mascarar outros interesses estatais oportunistas.

Nesse sentido, o Tribunal de Justiça da União Europeia, no caso *Digital Rights Ireland* (C-293/12), destacou a existência do direito fundamental à proteção dos dados, assentando que qualquer medida legislativa promulgada para fornecer uma base legal para o tema deve atender à proporcionalidade, de modo que só podem ser introduzidas restrições a esses direitos e liberdades se forem necessárias e corresponderem efetivamente a objetivos de interesse geral reconhecidos pela União ou à necessidade de proteção dos direitos e liberdades de terceiros.

No Brasil, ainda sobre o caso do IBGE e da MP nº 954, o STF (BRASIL 2020b, p. 62), considerando a violação ao postulado da proporcionalidade, destacou que:

> (...) as leis que tratam de coleta e processamento de dados devem (i) atender a propósitos legítimos, específicos, explícitos e informados; (iii) limitar a coleta ao mínimo necessário para a realização das finalidades normativas; (iv) prever medidas técnicas e administrativas de segurança aptas a proteger os dados pessoais de acessos não autorizados e (v) prevenir a ocorrência de danos, consoante os parâmetros desenhados no direito comparado e no art. 6º da Lei Geral de Proteção de Dados Pessoais (Lei 13.709/18). *In casu*, a Medida Provisória 954/2020 não atende ao direito à proteção de dados e ao postulado da proporcionalidade, máxime porque (i) não especifica para quais finalidades os dados serão utilizados; (ii) incorre em excesso ao determinar o compartilhamento de dados de milhões de brasileiros, quando pesquisas amostrais realizadas pelo IBGE em geral envolvem apenas cerca de 70 (setenta) mil domicílios por mês; (iii) não detalha métodos de segurança para a proteção dos dados contra riscos de vazamento; (iv) determina que o relatório de impacto à proteção de dados seja elaborado somente após já efetivado o compartilhamento, e não antes; e (v) pode gerar um nível preocupante de precisão na identificação dos usuários (...)

Assim, como escopo interpretativo, para os parâmetros decisórios do STF, sabe-se desde logo que a LGPD deve guiar sua aplicação atendendo propósitos legítimos, específicos, explícitos e informados, com coleta do mínimo necessário, adotando medidas de prevenção de vazamentos e danos. Por isso, no caso do IBGE, foi entendido que a MP nº 954/2020 não atendia ao direito à proteção de dados e ao postulado da proporcionalidade pelos fatos de não especificar as finalidades para as quais os dados seriam utilizados, causando um excesso na coleta e no compartilhamento de dados. Além disso, a MP não trazia os métodos de segurança para a proteção dos dados contra riscos de vazamento que seriam adotadas etc.

Contudo, ainda fica em aberto, não por qualquer descrédito da decisão do STF ou do próprio STF, a questão da possibilidade de relativização da proteção de dados, porquanto tem estrutura de direito fundamental, para a satisfação de outros direitos fundamentais. O STF, aplicando a proporcionalidade, chegou ao resultado de

maior peso relativo à proteção de dados. O que é especialmente bom em tempos de direito da crise, isto é, quando o Estado derruba garantias individuais para satisfazer projeto de governo, muitas vezes influenciado pelo medo, sem gestão de crise das autoridades. Nesse sentido, Long (2020, p. 92-94) ensina que uma vez estabelecidos mecanismos de coleta de dados indiscriminados, é improvável que poderes governamentais de vigilância e coleta de dados retrocedam voluntariamente. A linha entre a materialização do *Big Brother* de Orwell e o direito fundamental à proteção de dados é tênue (Sartori; Bahia, 2019, p. 231). A história também tem ensinado que uma vez que dados são coletados para um propósito é difícil evitar que sejam usados para outros fins não relacionados. Long argumenta ainda que sempre haverá uma próxima pandemia, senão de COVID-19, de outro agente infeccioso, e que esses momentos não podem servir de desculpa para o abandono das garantias fundamentais. Os desafios que as pandemias apresentam para a privacidade, mais do que para a liberdade, não irão embora nem se atenuarão com brevidade.

A pandemia ocasionada pelo temor em torno da propagação da SARS-COV-2, além de ter medidas restritivas da liberdade, postergando muito da vida presencial, e acelerando muito o desenvolvimento da vida através da Internet, como a expansão da educação EAD, do e-Commerce etc., significou uma postergação da vigência da própria LGPD. A *vacatio legis*, originalmente composta para 24 meses (art. 65, II), foi prorrogada pelo Regime Jurídico Emergencial e Transitório das relações jurídicas de Direito Privado no período da pandemia do coronavírus e pela MP nº 959/2020. Em 18/09/2020, contudo, e finalmente, a LGPD começou a surtir efeitos a partir da sanção presidencial (Santos, 2021, p. 254).

Diante de todo esse paradigma, insta-se em refletir acerca da tutela dos dados pessoais dos usuários frente à realidade algorítmica, que impõe novos desafios à LGPD, seja no e-commerce, no mercado comum e no principal mercado de todos: o *Big Market* dos dados pessoais. O que, na verdade, é totalmente transcendido com o avanço da área das inteligências artificiais. Movimento notado no direito. A LGPD até pode ser recente em sua discussão mais amplificada (digitalizada) no contexto jurídico, e, ainda assim, está ultrapassada – ou, no mínimo, incompleta –, dada a velocidade e expansão das tecnologias de automatização destes processos decisórios.

Essa nova realidade digital, marcada pela ascensão dos processos automatizados de tomada de decisão, traz à tona a problemática da discriminação algorítmica, que pode perpetuar – de forma inadvertida ou deliberada – desigualdades históricas, reproduzindo estereótipos e injustiças sociais já presentes em nosso tecido social. Assim, para além do desafio de garantir a conformidade normativa e a proteção dos direitos fundamentais prevista pela LGPD, emerge a necessidade de uma análise crítica e multidisciplinar acerca das repercussões sociojurídicas da discriminação gerada por algoritmos, sobretudo diante do papel central que essas tecnologias ocupam na vida cotidiana. A tecnologia, enquanto motor do progresso, apresenta-se também como vetor de riscos quando mal compreendida ou empregada sem a devida reflexão ética e jurídica sobre seus impactos. Desse modo, este debate insere-se no esforço coletivo de construir uma abordagem semântica e epistêmica apta a compreender o fenômeno da discriminação algorítmica e a buscar, no Direito, respostas regulatórias e jurisprudenciais que assegurem a igualdade e a dignidade no ambiente digital.

A discriminação algorítmica representa um desafio complexo, que transcende fronteiras disciplinares e, à medida que algoritmos e inteligência artificial (“IA”) desempenham papéis cada vez mais proeminentes em nossa sociedade – tendo como exemplos mais imediatos a tomada de decisões automáticas acerca da disponibilidade de crédito, recomendação para vagas de emprego, perfilização de consumidores ou de pessoas inseridas no sistema criminal e prisional, até a tentativa de predição do risco de reincidência de pessoas condenadas por crimes que tem, em alguns estados norte-americanos, a pena agravada por conta desta aplicação algorítmica –, arvora-se com mais gravidade a necessidade de um olhar atento à proteção dos direitos fundamentais. Para tanto, o grau de previsibilidade – fruto do esforço de técnicos e juristas – não deve ser subestimado, devendo alcançar um conhecimento acerca das fontes e das razões para a ocorrência de resultados discriminatórios pelas aplicações algorítmicas, que – adianta-se – são diversas e complexas, envolvendo desde a falta de diversidade na equipe de desenvolvimento, até a tendência dos algoritmos de perpetuar os preconceitos existentes na sociedade a partir da construção desse raciocínio autonomamente.

Segundo O’Neil (2020, p. 13), os algoritmos podem perpetuar preconceitos e estereótipos, levando a decisões injustas e desiguais, afetando negativamente

pessoas de grupos minoritários, como mulheres, negros, latinos e LGBTs, e ainda mais pessoas que são diversamente discriminadas, isto é, por mais de uma característica em um acúmulo interseccional de múltiplas “razões discriminatórias” (Fiss, 1976). Tal marginalização soma-se à nova discriminação automatizada, potencializada pela discriminação estrutural de que são constantemente vitimadas (Sagüés, 2018, p. 132), figurando uma verdadeira violência algorítmica.

Além disso, como bem demonstra Heikkilä (2023b), as máquinas não são conscientes, não tendo dentro de si, ao longo das linhas de sua codificação técnico-genética, qualquer autoimplicância ou autorreflexão acerca de seus atos, das implicações éticas ou das consequências potencialmente nocivas de suas atitudes. Tais preocupações até podem se fazer presentes, mas não a título de consciência, e sim de construção de *feedbacks e reports* (relatórios). O que pode ser objeto de desejo por parte de desenvolvedores e usuários, a consciência da Inteligência Artificial demandaria, primeiro, conhecer a própria essência da consciência humana, algo ainda desconhecido. E, portanto, não completamente auferível a priori ou mesmo implicado em previsibilidade.

Os algoritmos são conjuntos de instruções lógicas que um programa de computador segue para realizar uma tarefa específica de maneira sistemática, seguindo uma série de instruções pré-estabelecidas para realizar uma determinada tarefa. Assim, quando as Inteligências Artificiais discriminam automaticamente pessoas no processo de sua tomada de decisão, não agem com consciência acerca da prejudicialidade ou da ilicitude do *outcome* (resultado) (Huckins, 2023). O termo “discriminação algorítmica” se refere às discriminações promovidas pelos algoritmos ao discriminar pessoas com base em características e padrões predeterminados, o que conduz naturalmente a resultados injustos e desiguais. Tais discriminações estão, geralmente, relacionadas com a amostragem de dados coletada e armazenada na base de dados utilizada pelo algoritmo em questão.

Algoritmos têm potencial discriminatório simplesmente por utilizarem dados. Se os dados históricos forem prejudiciais ou não representativos da população, os algoritmos podem refletir esses dados e perpetuar padrões não desejáveis, que incrementam e reforçam essas estruturas invisíveis que remontam à preconceitos discriminatórios que ainda influenciam e interferem em direitos fundamentais pessoais, de modo ainda mais grave agora, pois automático e, a depender da

situação, de modo opaco ao próprio operador da aplicação algorítmica. O problema é que não há inteligência artificial ou algoritmos sem o uso de dados. A qualidade da inteligência e do processamento de dados depende diretamente do volume e da qualidade destes. Fatalmente, os algoritmos acabam discriminando pessoas com base em vieses herdados ou não identificados por sua base de dados ou por seus programadores.

Características como gênero, raça e classe social facilmente servem para classificar pessoas, especialmente em um contexto de discriminação histórica e estrutural, onde estes aspectos estão arraigados na própria lógica de funcionamento da sociedade, e isso tem especial potencial lesivo dentro da linguagem de máquina, que faz automaticamente e de modo muito eficiente a clusterização, isto é, a formação de grupos *ad hoc* não necessariamente autoevidentes em suas correlações, aos olhos humanos, pelo menos. Tais clusterizações geralmente dizem respeito a grupos vulneráveis que também são identificáveis pela sua situação nada privilegiada de serem sujeitos passivos da discriminação estrutural. Isto é, uma discriminação estruturada nas raízes identitárias da sociedade que, muitas vezes, aponta para grupos marginalizados, ou ainda interseccionalidades que prejudicam o desenvolvimento pessoal, comunitário e social na comunidade. Tais grupos de pessoas são identificados por semelhanças quanto à questões de ordens econômicas, políticas, de origem étnica, nacional ou linguística, orientação sexual, escolhas religiosas e outras (Carbonell, 2000).

As estruturas algorítmicas, ao operarem com dados marcados por estigmas sociais, reproduzem exclusões já impostas a determinados grupos, reforçando desigualdades calcadas em fatores históricos de marginalização. Vê-se, portanto, que as vítimas da discriminação algorítmica, muitas vezes sendo pertencentes também à grupos vulneráveis, têm um especial prejuízo em relação ao uso dos algoritmos, uma vez que não raras vezes as conclusões das aplicações de Inteligência Artificial podem apontar a vedações de determinados bens e serviços a determinados *clusteres*, ou, então, estereotipar pessoas em tônicas de identidade não acertadas. A discriminação nessa tônica mais gravosa é de natureza estrutural, e essa demanda um olhar atento tanto em termos conceituais quanto em termos da prática da jurisdição e do dever de proteção estatal (Mendes; Mattiuzzo, 2019, p. 51-52).

Sob o ponto de vista causal, três vetores são recorrentes: (i) dados de treinamento impregnados por assimetrias históricas; (ii) arquiteturas e escolhas de modelagem que reforçam vieses; e (iii) homogeneidade sociocultural das equipes de desenvolvimento, que reduz a sensibilidade a grupos sub-representados. A combinação desses elementos torna previsível a reprodução e a intensificação discriminações. A efetividade do direito antidiscriminatório algorítmico deve se dar pelo desenho preventivo “by design”, seja por esforço regulatório, ou por imposição às empresas a partir do desenho interpretativo-constitucional que se constrói.

Há, contudo, que se ir um pouco mais além na busca conceitual do estado da arte da vulnerabilidade, eis que a desigualdade pode ainda figurar-se a partir de uma forte interseccionalidade – múltiplas origens e motivações injustificadas que se acumulam e agravam ainda mais as discriminações (Carbonell, 2000) – que impulsiona a lesividade da condição, como é o caso de pessoas que se enquadram em múltiplas categorias de vulnerabilidade e são objetificadas, até mesmo inconscientemente, como menores em importância e dignidade pelas estruturas sociais, ou então de modo intencional (Moreira, 2019, p. 64).

Importa reconhecer que, para além das vulnerabilidades singulares, certas injustiças derivam de camadas sobrepostas de exclusão – fenômeno que se expressa na interseccionalidade como vetor intensificador da discriminação, e ainda mais acelerado diante da capacidade contida no meio digital, sobretudo diante do fenômeno da inteligência artificial, da discriminação algorítmico-automatizada.

A discriminação estrutural, acelerada e intensificada pela discriminação algorítmica, ganha força e destaque em sua capacidade lesiva diante das técnicas utilizadas pelas Inteligências Artificiais, sobretudo pela “*clusterização*”, e demais ferramentais automatizados de discriminação que retroalimentam o fenômeno da discriminação estrutural. Está-se diante de fenômenos autopoiéticos que, quando somados, incorporam-se em ciclos viciosos e cada vez maiores de aumento, não só manutenção, dos ciclos de estratificação e estigmatização sociais. O próximo passo é compreender o fenômeno da automatização digitalizada da discriminação, que, nutrindo-se da discriminação interseccionalmente estrutural, a amplia, a robustece e a escala em velocidade nunca antes vista.

O percurso desenvolvido mostrou que a desigualdade é problema central na organização social do Estado Democrático-Constitucional de Direito, cuja tutela

jurídica pode auxiliar a estabilizar. A vulnerabilidade, compreendida como condição relacional e seletivamente agravada por estruturas históricas, exige do Estado um dever de proteção normativamente denso: prevenir, mitigar e reparar assimetrias, inclusive quando mediadas por arquiteturas técnicas que promovem o agravamento automatizado desse fato. É nessa perspectiva que se deve pensar a atuação jurisdicional: como prudência ativa que impede a naturalização da exclusão, e nunca como órgão de normatização da neutralidade; posto que a neutralidade em situação estruturalmente desalinhada significa a manutenção e o não-agir em concordância aos direitos fundamentais.

A algoritmização da vida social, que já vinha sendo virtualizada, acelerou demasiadamente essa sistemática. A técnica, quando sem ética e controle democrático, pode ampliar distorções sociais. Sistemas algorítmicos replicam imperfeições humanas em larga escala e, ao transformar dados estigmatizados em métricas, podem legitimar exclusões históricas e dificultar sua contestação devido à opacidade dos processos automatizados. Tal cenário exige atuação jurídica ativa para proteger grupos vulneráveis – cuja vulnerabilidade incide interseccionalmente (Collins e Bilge, 2021, p. 18) –, promover transparência e adaptar o Direito aos desafios da discriminação algorítmica, garantindo igualdade e dignidade frente ao avanço digital. Os sistemas algorítmicos, ao operarem sobre dados marcados por estigmas, reproduzem-nos com a velocidade e a opacidade da técnica. O capítulo seguinte é de uma proposta de cartografia dessa era da discriminação algorítmica.

Se no século XIX Jean Valjean foi reduzido a um número – e carregou, mesmo livre, a pena que não cessa –, o nosso tempo produz etiquetas mais silenciosas e duradouras: *scores*, *profiling* e probabilidades. Mudam os instrumentos, persiste a ordem. A tarefa do Direito – assim – é impedir que a técnica repita a velha sentença da miséria, agora sob o selo da neutralidade. A mutação tecnológica altera o meio da dominação do poder, exigindo critérios jurídicos de identificação, imputação e reparação das desigualdades que ela reencena. Deve-se construir uma arquitetura protetiva, um sentido normativo de tutela à igualdade de tratamento, para que, ainda que não seja capaz de reprogramar artificialmente a estrutural desigualdade, seja, pelo menos, apto a mitigar e combater prudencialmente a discriminação algorítmica.

3 A ERA DA DISCRIMINAÇÃO ALGORÍTMICA: MAPEANDO O FENÔMENO E SUAS IMPLICAÇÕES DE ORDEM JUSFUNDAMENTAL

> *Lembra, fizestes-me mais poderoso que tu mesmo; minha altura é superior à tua; minhas articulações mais maleáveis. Mas não cairei na tentação de me colocar em oposição a ti. Eu sou criatura tua, e serei até manso e dócil ao meu senhor e rei natural, se tu também realizares a tua parte, aquilo que me deves (...) eu deveria ser teu Adão; mas sou antes o anjo caído, que expulsaste da beatitude sem que eu tivesse cometido delito nenhum (...) Eu era benevolente e bom; a infelicidade me tornou um demônio. Faze-me feliz, que serei virtuoso outra vez (Shelley, 2022, p. 171).*
>
> *Como Adão, eu aparentemente fora criado sem nenhuma relação de união com nenhum outro ser existente; (...). Ele saíra das mãos de Deus uma criatura perfeita, feliz e próspera (...) mas eu era miserável, indefeso e sozinho. Considerei muitas vezes Satã o emblema mais apropriado da minha condição; pois, como ele, quando eu via o júbilo dos meus protetores, com frequência o amargo fel da inveja brotava dentro de mim (Shelley, 2022, p. 223).*

Entre Adão e Satã, a criatura de Frankenstein revela sua tragédia: não nasceu má, mas foi lançada à miséria e à solidão. Sem amparo, sem vínculo, sem reconhecimento, ela não se compreende como parte da ordem – apenas como dissonância. Sua existência não é negada: é ignorada, esvaziada de sentido, e é nesse vazio que a distorção germina. Não é o mal que funda o monstro, mas o abandono, a falha do criador no acompanhamento do crescimento e desenvolvimento de sua criatura.

Assim também os sistemas algorítmicos – frutos da engenhosidade humana, erguidos sob o ideal de precisão, neutralidade e eficiência – tornam-se, ao ganhar vida no mundo real, reflexos ampliados das falhas que os moldaram. Eles não erram por si: apenas reproduzem – em escala e velocidade inéditas – os padrões de exclusão

já inscritos na sociedade. O que deveria ser belo, justo e técnico, converte-se em instrumento de desumanização automatizada, quando sua lógica é regida por dados enviesados, objetivos obscuros e ausência de controle democrático.

Não é o erro do criado, mas a omissão do criador. A criatura não é rejeitada pelo que fez, mas por aquilo que passou a significar: um reflexo amplificado das falhas humanas. A criatura, como os algoritmos de nosso tempo, emerge da promessa de racionalidade e aperfeiçoamento técnico – mas, uma vez solta no mundo, reproduz e intensifica as exclusões de seus artífices. Os sistemas algorítmicos não erram por si: eles apenas potencializam, em escala e velocidade, os vieses e silenciamentos que herdaram. Quando essas decisões são convertidas em verdades inquestionáveis, o Direito perde a possibilidade de exigir razões públicas e, portanto, os sujeitos de direito perdem a possibilidade de contestação ou revisão da situação.

Este capítulo examina esse deslocamento. Antes, o objetivo foi conceitual e diagnóstico: explicitar vulnerabilidade, discriminação estrutural/institucional e os mecanismos sociais que alimentam vieses nos dados. Agora, o foco é normativo-operacional: tratamos a opacidade algorítmica como problema jurídico e descrevemos como o Direito identifica, prova e corrige discriminações automatizadas. Parte-se do diagnóstico de que os algoritmos não criam desigualdades do nada: eles aprendem, reproduzem e refinam os padrões discriminatórios historicamente inscritos na sociedade. A opacidade técnica – a chamada “black box” – não é só uma limitação epistemológica; é um obstáculo normativo à reconstrução da dignidade no mundo digital. Quando as decisões são tomadas por máquinas e naturalizadas por sua pretensa neutralidade, o direito à explicação cede lugar ao silêncio; os afetados por essas decisões veem-se diante de um novo tipo de exclusão: inumana, automática e opaca.

A opacidade algorítmica não é apenas um dado técnico; é um instrumento bélico silencioso e, por vezes, não desejado. Como o cavalo de madeira dado aos troianos pelos gregos, sua força não reside naquilo que mostra, mas no que esconde. O usuário vê a interface simplificada; o Estado, a promessa de eficiência; o mercado, a possibilidade de lucro. O que não se vê – e que salta à noite do âmago da estrutura maravilhosa, quando já é tarde demais – são os padrões discriminatórios reproduzidos por dados históricos enviesados, são as exclusões cristalizadas como cálculo neutro.

Este capítulo propõe, portanto, o mapeamento conceitual e normativo da chamada discriminação algorítmica, compreendida como fenômeno de replicação sofisticada das desigualdades estruturais. Busca-se articular os desafios colocados por sistemas decisórios opacos ao dever de proteção estatal, à luz do constitucionalismo digital e da gramática dos direitos fundamentais. Tal como Victor Frankenstein, que volta os olhos horrorizados para a criatura que escapou ao seu controle, o Direito contemporâneo precisa reencontrar sua responsabilidade por aquilo que, mesmo quando delegado à máquina, continua sendo sua criação.

Nesse contexto, compreender a era da discriminação algorítmica requer não só mapear as formas de exclusão já existentes, mas também analisar como essas novas tecnologias amplificam e complexificam os desafios enfrentados por grupos vulneráveis diante da aceleração digital. A natureza opaca das decisões automatizadas, muitas vezes alicerçada em dados enviesados, impõe obstáculos adicionais à transparência e à responsabilização, tornando o monitoramento social e jurídico dessas práticas ainda mais essencial. O resultado é um ambiente em que potencialidades emancipatórias da tecnologia convivem com riscos ampliados de intensificação da desigualdade, exigindo abordagens interdisciplinares para o enfrentamento desses dilemas.

O avanço da inteligência artificial, portanto, não só revela o quanto as estruturas sociais históricas permanecem presentes nos sistemas digitais, mas também evidencia a urgência de repensar políticas, legislações e práticas de governança capazes de identificar, mitigar e prevenir as novas formas de exclusão. A discussão sobre discriminação algorítmica não se restringe ao aspecto técnico, mas integra profundamente o campo dos direitos fundamentais, da cidadania e da justiça social, colocando em xeque a própria legitimidade das decisões automatizadas em contextos sensíveis.

Assim, analisar a interseção entre algoritmos, vulnerabilidade e discriminação estrutural é fundamental para a construção de um ambiente digital mais justo, onde a tecnologia seja instrumento de inclusão, e não de perpetuação ou agravamento das desigualdades existentes.

A interseção entre tecnologia e Direito tem produzido transformações profundas nas estruturas sociais e jurídicas contemporâneas, passando a vida a ser coordenada por uma série de decisões automáticas tomadas por máquinas. Os

sistemas de inteligência artificial são uma realidade da vida cotidiana, estando empregados nos mais diversos âmbitos: na saúde, nas atividades laborais, no lazer, entre tantos outros. Isso porque, essas revolucionárias tecnologias se desenvolvem em um contexto social em que ideias preconceituosas estão fortemente enraizadas contra grupos historicamente vulneráveis, criando um espaço (in)fértil para o desenvolvimento da discriminação algorítmica, que congrega múltiplas origens – sendo um importante marco dos estudos dos direitos fundamentais sob a ótica do constitucionalismo digital.

Esse cenário revela a necessidade de uma abordagem regulatória que não apenas acompanhe o ritmo das inovações tecnológicas, mas que também antecipe e mitigue os riscos de reprodução e intensificação das desigualdades. Políticas públicas, mecanismos de accountability e transparência, assim como instrumentos de participação social, tornam-se fundamentais para garantir que a aplicação dos algoritmos seja orientada pelos princípios da equidade, inclusão e respeito aos direitos fundamentais. É crucial envolver múltiplos atores – Estado, sociedade civil, setor privado e academia – na construção de parâmetros éticos e normativos que orientem o desenvolvimento e uso responsável dessas tecnologias. O foco aqui, no entanto, é diverso e alocado no papel do Poder Judiciário nessa quadra histórica situacional.

Diante desse panorama, fica evidente que a ascensão dos algoritmos e da inteligência artificial não ocorre em um vácuo social, mas profundamente imbricada às dinâmicas históricas de exclusão e discriminação. A automatização das decisões cotidianas, ao mesmo tempo em que oferece oportunidades inéditas de eficiência, também acarreta desafios éticos e jurídicos de grande magnitude, particularmente no que se refere ao risco de ampliação e sofisticação das injustiças já existentes. É imprescindível, portanto, adotar uma perspectiva crítica e interdisciplinar, capaz de reconhecer tanto o potencial emancipador das tecnologias digitais quanto seus riscos de perpetuação e agravamento das desigualdades históricas.

Neste capítulo, propõe-se inicialmente mapear a evolução tecnológica recente, com especial destaque para o papel da inteligência artificial e da automação algorítmica nas decisões cotidianas, verificando inclusive mecanismos possíveis de controle e sindicabilidade dos sistemas inteligentes. Será examinada a forma como tais sistemas vêm sendo incorporados ao cotidiano social e institucional, transformando processos decisórios hodiernos. Em seguida, o enfoque recairá sobre

as discriminações algorítmicas propriamente ditas: suas causas, mecanismos subjacentes e o potencial de impulsionamento das discriminações estruturais, especialmente no tocante ao impacto sobre grupos vulneráveis e interseccionais. E, por fim, será discutido o dever de proteção estatal nesse novo cenário, analisando as responsabilidades do Estado Democrático e Constitucional de Direito na regulação, prevenção e responsabilização diante das novas formas de discriminação algorítmica, de modo a promover a efetiva tutela dos direitos fundamentais e a construção de uma sociedade verdadeiramente inclusiva.

3.1 Evolução tecnológica, inteligência artificial e automação de decisões cotidianas

Como a criatura de Frankenstein, os sistemas de inteligência artificial foram construídos sob a promessa de racionalidade, controle e benefício coletivo. No entanto, à medida que ganham autonomia funcional, esses sistemas também revelam sua ambivalência: podem produzir luz – mas carregam consigo as sombras da exclusão que os moldaram. Em pleno século XXI, a tecnologia se converte na nova "caixa de Pandora": um receptáculo de potências cujo desvelamento expõe riscos profundos, frequentemente escapando ao controle dos próprios criadores. Estar no labirinto algorítmico é habitar um espaço de complexidades e desorientações, onde a esperança – ainda que remanescente – resiste como possibilidade de reconstrução normativa. Reconhecer que os sistemas automatizados reproduzem desigualdades e operam com opacidade é o primeiro passo para sua contenção jurídica. Resta-nos, então, o fio de Ariadne: construir, com precisão técnica e responsabilidade institucional, os caminhos de saída que permitam à jurisdição constitucional enfrentar os perigos invisíveis do mundo automatizado. Apesar das sombras da caixa e da intricada arquitetura do labirinto, é possível vislumbrar clarões – uma sociedade mais justa e igualitária, sustentada por um Direito capaz de compreender, traduzir e disciplinar os riscos modernos da discriminação algorítmica.

Os algoritmos são conjuntos de instruções lógicas que um programa de computador segue para realizar uma tarefa específica, organizado na modalidade escalonada *step-by-step*. Eles são modelos escalonados de comandos, que fornecem uma sequência de passos a serem seguidos para obter um resultado desejado. Os

algoritmos funcionam de maneira sistemática, seguindo uma série de instruções pré-estabelecidas para realizar uma determinada tarefa. Eles são usados em praticamente todos os programas de computador, desde os mais simples, como um editor de texto, até os mais complexos, como em um software de inteligência artificial estrita (*narrow artificial intelligence*), que utiliza de aprendizado de máquina (*machine learning*) para treinar algoritmos a realizarem determinadas tarefas sem que tenham sido programados para as executar estritamente (Cozman; Neri, 2021, p. 22-25). Eis o motivo pelo qual os softwares acessíveis de Inteligência Artificial geralmente apresentam *disclaimer* no sentido de que ainda estão em fase *beta* de testes e aperfeiçoamentos. Motivo pelo qual também o *feedback* é uma ferramenta central para o uso de tais softwares, como se verá.

Para que um algoritmo funcione corretamente, ele precisa ser criado com clareza e precisão, de modo que as instruções possam ser seguidas de forma coerente e sem ambiguidades. Além disso, os algoritmos precisam ser testados e validados para garantir que produzam os resultados desejados em todas as situações previstas. Os algoritmos podem ser escritos em diferentes linguagens de programação, como Python, Java, C++ etc., e são executados por programas de *compiler* (que traduz a linguagem da programação para a linguagem binária do processador) e pelo processador do PC. Quando um algoritmo é executado, ele processa as informações de entrada de acordo com as instruções fornecidas (*inputs*) e produz as saídas-resultados (*outputs* ou *outcomes*) desejadas; ou indesejadas.

Os algoritmos foram projetados para ajudar no processo racional de tomada de decisões mais objetivas e baseadas em dados objetivos e estatísticos, em alcance superior ao humano. Se há algo que distingue homem de máquina, esse algo é a consciência, sendo essa o fator determinante de categorização entre uma inteligência do tipo humana e uma inteligência do tipo máquina. Essa consciência é marca determinante das espécies racionais, de alocar-se em percepção de perspectiva sobre panoramas gerais (Russell, 2021, p. 25). Em termos de capacidade de processamento de dados não há comparativo: em 2016 a Tianhe-2 processava 55 quatrilhões de operações (conexões entre células de processamento) por segundo enquanto a raça humana consegue cerca de 10 quatrilhões de operações (conexões entre neurônios) por segundo. A única vantagem humana, por enquanto, é a energia gasta para tanto.

O Tianhe-2 utiliza 1 milhão de vezes mais energia que o cérebro humano (Nicolelis, 2020, p. 137-138).

Outra característica distingue homem de máquina: a capacidade de situação, isto é, a capacidade de se situar no tempo e de alocar menos energia e dedicação a uma tarefa não principal. As máquinas, mesmo aquelas chamadas de Baby-AGI (recentes tentativas de Inteligência Artificial Geral[5]), não conseguem por si só determinar que devem ser criados novos *softwares,* algoritmos e *bots* para a solução de problemas que lhes são apresentados. Para tanto, é necessário que nas linhas de código esteja presente a possibilidade dessa sub-criação. O cérebro humano não é assim, pois age enquanto *lócus* de processamento informativo e determinante do agir humano, um ser situacional, é distinguido (Nicolelis, 2020, p. 138-139).

Apontando em sentido totalmente contrário, Schwab (2016, p. 140-141) apostava que até 2025 as máquinas superariam o cérebro humano, visto que superariam o nível de processamento biológico, e ele acertou, como visto acima. A diferença de visões entre o neurocientista Nicolelis e o economista Schwab não está, portanto, na objetividade do processamento, mas no núcleo mais material e criativo do cérebro.

Inteligência humana é inteligência consciente. Inteligência de máquina é inteligência voltada a uma intencionalidade específica. Embora seja comum equiparar “inteligência” a desempenho em tarefas, a inteligência humana é situada, encarnada e normativamente imputável: ela envolve linguagem, contexto e responsabilidade por razões. Já a “inteligência” artificializada de máquina opera por correlação e otimização de funções-objetivo, podendo produzir respostas aparentemente sofisticadas sem, por isso, compartilhar as condições humanas de sentido e imputação. Há uma dialética na relação entre a realidade da vida humana e a realidade das programações e dos ajustes realizados no mundo algorítmico, não só através dos dados utilizados (a todo momento criados, coletados e fornecidos), mas também pelos ajustes no pensamento dos *prompters* (programadores). Há uma algoritmização da vida. E uma atualização do programa e do dado no algoritmo. Os algoritmos, no entanto, estão sempre presos no passado humano.

Um mundo algoritmizado é um mundo estagnado no passado. É um mundo conservador pelo algoritmo. Ainda que os dados sejam sempre renovados, a

[5] Não se recomenda a instalação, apenas indica-se a fonte https://github.com/yoheinakajima/babyagi

renovação dos padrões de análise deverá sempre ser revista ativamente pelos programadores. Esse conservadorismo é frequentemente um efeito de retroalimentação: sistemas de recomendação reforçam padrões anteriores porque maximizam engajamento com base no passado, convertendo o histórico em norma silenciosa.

Os algoritmos foram projetados para ajudar no processo racional de tomada de decisões mais objetivas e baseadas em dados objetivos e estatísticos. No entanto, tais resultados podem ter efeitos negativos – que talvez não compensem seus benefícios – como é o caso da discriminação promovida pelo viés algorítmico. O termo "discriminação algorítmica" se refere às discriminações promovidas pelos algoritmos ao discriminar pessoas com base em características e padrões predeterminados, o que conduz naturalmente a resultados injustos e desiguais. Tais discriminações estão, geralmente, relacionadas com a amostragem de dados coletada e armazenada na base de dados utilizada pelo algoritmo em questão. Não existe inteligência artificial sem dados, e a qualidade da inteligência e do processamento de dados depende diretamente do volume e da qualidade dos dados coletados, armazenados, tratados e interpretados para o algoritmo.

Os algoritmos discriminam porque são programados com base em dados (sempre do passado), que frequentemente foram influenciados por fatores sociais, culturais ou outros tipos de contexto que podem ser mal interpretados pela máquina e resultar na perpetuação de estereótipos. Por exemplo, um algoritmo que analisa o histórico de emprego de uma pessoa com muitas faltas durante a jornada de trabalho pode chegar à conclusão de que a pessoa é uma má trabalhadora, sem levar em conta circunstâncias como maternidade, férias ou outras eventualidades lícitas, que de modo humano percebemos sem qualquer prejuízo, mas que devem ser especificamente demonstradas e programadas para que a inteligência artificial ("IA") possa corretamente interpretar e processar esses dados.

Algoritmos têm potencial discriminatório simplesmente por utilizarem dados. Se os dados históricos forem prejudiciais ou não representativos da população, os algoritmos podem refletir esses dados e perpetuar padrões não desejáveis. O problema é que, sem o uso de dados, não há inteligência artificial, nem algoritmos, nem ferramentas automáticas para auxiliar o ser humano. A qualidade da inteligência e do processamento de dados depende diretamente do volume e da qualidade dos

mesmos. Fatalmente, os algoritmos acabam discriminando pessoas com base em vieses herdados ou não identificados por sua base de dados ou por seus programadores. Características como gênero, raça e classe social facilmente servem para classificar pessoas, especialmente em um contexto de discriminação histórica e estrutural, onde estes aspectos estão arraigados na própria lógica de funcionamento da sociedade, e isso tem especial potencial lesivo dentro da linguagem de máquina. Os algoritmos podem usar estas características para determinar automaticamente se uma pessoa será contratada, se receberá um empréstimo ou se será admitida em uma universidade, o que não parece desejável para uma sociedade madura em termos de direitos humanos e fundamentais – e que, na perspectiva de um Estado Democrático de Direito, é pautada pelo pluralismo e pelo respeito às diferenças, com base no direito de igualdade e da não-discriminação.

A coleta e o uso de dados podem levar à discriminação apenas pelo fato de que há viés puramente na amostragem humana dos dados coletados. É por isso que a tomada de decisão com base em algoritmos pode parecer objetiva, mas na verdade está carregada de preconceitos e estereótipos, o que fatalmente leva, em um grande número de casos, a resultados injustos, como a recusa de emprego ou a negação de crédito por motivos como raça, sexualidade, ou bairro em que reside, ou ainda, como é mais comum, é eliminado de processo seletivo de emprego porque atrasou o pagamento de contas em um mês – o que poderia demonstrar desleixo. Ainda que tais considerações possam ser fundamentadas e comprovadas empiricamente e cientificamente, não podem servir de motivos para a tomada de decisão, visto que ofendem a dignidade da pessoa humana, visto que atingem o núcleo redutível da própria identificação de existência do ser em si.

Pasquale (2015) explora como as empresas e os governos já utilizam há alguns anos algoritmos para tomar decisões centrais e cruciais, como política monetária ou investimentos. Pasquale (2015, p. 160) argumenta que tais práticas podem até ocorrer, mas não podem ser uma caixa-preta – daí o título de seu livro – visto que necessariamente devem ter mais transparência e responsabilidade, sobretudo quando o resultado decisório atinge diretamente o ser humano. A necessidade de transparência, em especial, tem o condão de imediatamente coibir certas práticas corporativas e governamentais que direcionam programações discriminatórias – o uso de programas discriminatórios pode vir a configurar o crime

de racismo, por exemplo – e sob uma perspectiva mediata, a transparência tem o condão de abrir a caixa-preta dos algoritmos e permitir um ajuste mais fino e sofisticado nas ferramentas, entregando respostas cada vez mais coerentes para com os aspectos nucleares do Estado Democrático de Direito. A auditabilidade deve ser buscada inclusive para desmistificar e aproximar a Inteligência Artificial da ética (Hollanek, 2020, p. 2071).

No nosso sistema jurídico, p.e., devemos ser julgados por aquilo que fazemos (nossas ações) e não por quem somos (nossos dados sensíveis). Com isso, O'Neil (2020, p. 42-43) descreve um cenário em que se montou uma base de dados a partir das respostas de detentos a questionários LSI-R, questionário para medir estatisticamente a chance de um detento reincidir na lida delituosa. A partir desse questionário, foi montado um sistema de respostas altamente correlacionadas, em que a similitude entre uma pessoa presa pela primeira vez e um detento já reincidente faziam com que o *score* subisse, isto é, a chance de que cometeria crimes novamente. Para tanto, de maneira antijurídica, eram avaliados dados sensíveis, tais como nome da família, bairro em que nasceu e cresceu, se amigos já foram presos, as circunstâncias do nascimento, etc.

O sistema montado a partir do LSI-R até pode ser muito efetivo para medir estatisticamente o risco de reincidência. Contudo, não poderia ser utilizado por ofensa à presunção da inocência (no segundo crime hipotético) e a imparcialidade, bem como não poderia guiar de maneira isolada um aumento de pena ou restrição corporal em regime mais gravoso. Essas são algumas implicações centrais do (mau) uso dos algoritmos na atual conjuntura da vida, que passa desde bonequinhos e avatares na internet, como se verá, até questões centrais dos sistemas de acessos a direitos.

Uma das principais preocupações ético-jurídicas em relação à discriminação algorítmica é exatamente a transparência do código-fonte dos algoritmos e da base de dados que o supre (Buiten, 2019, p. 44). Como os algoritmos são muitas vezes considerados caixas-pretas, é difícil saber como as decisões são tomadas e se elas são justas e imparciais. A opacidade dos algoritmos, ao mesmo tempo que intencionalmente arquitetada para conservar a programação – e seu valor inerente no mercado –, dificulta a percepção prévia acerca de erros ou vieses que poderão ser cometidos. A aplicação de algoritmos pode ser afetada por problemas em diversas etapas, sendo que a definição das regras e o uso de dados inadequados são

especialmente suscetíveis a gerar tais situações. Diante disso, há uma preocupação crescente com três grandes problemas que podem surgir em decorrência do uso de algoritmos. O primeiro é o uso de conjuntos de dados enviesados, que podem levar naturalmente a resultados equivocados e injustos por conta do substrato (dados) em que executadas as operações algorítmicas. O segundo é a falta de transparência em algoritmos não programados – que aprendem tudo a partir de uma missão objetivada –, que torna difícil entender como as decisões são tomadas e contestá-las quando necessário, sendo a base de dados do treinamento (entregue ou disponibilizado) o único veículo para tais verificações. O terceiro problema é a possibilidade de discriminação que pode ser gerada por algoritmos de aprendizado de máquina, em que por questões sutis encontradas durante o treinamento, ou mesmo uso, a aplicação chega a uma conclusão nociva – caso típico do HAL 9000 do filme 2001 – Uma Odisseia no Espaço. Por isso, é importante estar atento a esses problemas ao utilizar algoritmos em diferentes contextos (Ferrari; Becker; Wolkart, 2018).

Eubanks (2017) propõe olhar o problema da desigualdade para com as tecnologias de automação, principalmente centralizada na desproporcionalidade com que pessoas em situação de vulnerabilidade são vistas e percebidas pelos algoritmos. Os algoritmos, segundo Eubanks (2017), são muito utilizados para blindar pessoas pobres do acesso a bens e serviços, como o crédito. O ponto de vista da autora é que a internet, o processamento de dados e os algoritmos têm o potencial para ser um grande agente de transformação no mundo, porém, ao invés disso, servem para (des)selecionar e segregar ainda mais os seres.

Esse cenário revela que o problema da discriminação algorítmica não se limita apenas às escolhas humanas que antecedem o desenho técnico dos sistemas, mas também se manifesta nas próprias estruturas operacionais que tornam sua atuação impermeável à crítica e à revisão. À medida que os algoritmos passam a mediar decisões socialmente relevantes – como o acesso a crédito, benefícios sociais ou a dosimetria penal –, a ausência de transparência se converte em um obstáculo concreto ao controle democrático, à responsabilização institucional e à garantia dos direitos fundamentais. Não é apenas o resultado discriminatório que preocupa, mas também o fato de que esse resultado, mesmo quando identificado, dificilmente pode ser rastreado, interpretado ou corrigido por sujeitos externos ao sistema. Essa lógica de funcionamento opaca, embora amparada por justificativas de propriedade

intelectual e complexidade técnica, tensiona diretamente os princípios estruturantes do Estado de Direito, especialmente a publicidade dos atos empresariais e estatais, derivando deste último lesão ao devido processo legal e ao controle jurisdicional das decisões automatizadas. Estando, ainda, e mais alarmantemente vinculado à experiência acelerada da iniciativa privada.

O crescimento exponencial da complexidade das ferramentas de inteligência artificial é acompanhado pelo crescimento, também exponencial, da dificuldade de entendimento acerca do seu funcionamento. Desta forma, Pasquale (2015, p. 8), sustenta que muitos destes algoritmos assumem a forma de caixas-pretas (*black boxes*), diferentes, todavia, daquelas próprias dos aviões que registram as comunicações entre os pilotos e as centrais de comando. A *black box* algorítmica diz respeito à impossibilidade de acesso do usuário final – e até mesmo dos programadores em alguns casos – aos dados e às fórmulas utilizadas para o estabelecimento de conclusões algorítmicas. O abrir da caixa deve ser tal qual a necessidade da informação que se precisa abstrair.

Verifica-se, portanto, que as *black boxes* podem ser vistas como um obstáculo à transparência e à responsabilidade corporativa e, da mesma forma, à privacidade e aos direitos fundamentais. Mas como são formadas as *black boxes* algorítmicas? De acordo com Pasquale (2015, p. 14) elas surgem como resultado de uma série de fatores. Em primeiro lugar, os algoritmos podem ser extremamente complexos, com múltiplas camadas de cálculo que são quase impossíveis de serem interpretadas sem conhecimentos especializados. Em segundo lugar, muitos algoritmos são desenvolvidos por empresas privadas que consideram a propriedade intelectual como uma vantagem competitiva. Isso significa que as empresas têm poucos incentivos para compartilhar informações sobre como seus algoritmos funcionam. Por fim, o último fator, corresponde ao fato de que muitos algoritmos são alimentados por enormes quantidades de dados, alguns dos quais podem ser confidenciais ou privados, o que dificulta o acesso às informações necessárias para entender como os algoritmos estão operando.

Devido a esses fatores, as *black boxes* algorítmicas são frequentemente usadas em áreas críticas, como finanças, saúde e justiça. As empresas de crédito, por exemplo utilizam algoritmos para tomar decisões sobre a concessão de crédito, entretanto, o público não possui acesso à lógica interna destes algoritmos, podendo

resultar, inclusive, a decisões injustas ou discriminatórias. Da mesma forma, os sistemas de saúde podem usar algoritmos para tomar decisões sobre diagnóstico e tratamento, mas os pacientes podem não saber como essas decisões são tomadas (O'Neil, 2020, p. 207).

A falta de transparência em torno das *black boxes* algorítmicas é um grande problema. Pasquale (2015, p. 21) argumenta que as empresas que desenvolvem algoritmos devem ser mais transparentes em relação ao seu funcionamento interno, e que os governos, em geral, devem estabelecer regulamentações mais rígidas para garantir que os algoritmos sejam justos e transparentes. Em última análise, ele acredita que é importante que a sociedade como um todo compreenda como esses algoritmos estão afetando a vida humana, a fim de garantir que eles sejam utilizados de maneira ética e justa.

A transparência algorítmica, como se viu, é um conceito essencial para garantir a privacidade e os direitos dos usuários. Isso porque a utilização de algoritmos em decisões automatizadas pode levar a erros e vieses discriminatórios, prejudicando indivíduos e grupos vulneráveis. É fundamental que as empresas de tecnologia sejam capazes de explicar como seus algoritmos funcionam e como eles tomam decisões, permitindo que os usuários avaliem e compreendam o processo de tomada de decisão.

Nesse contexto, torna-se imprescindível discutir mecanismos efetivos de auditoria e fiscalização dos sistemas automatizados, promovendo não apenas a clareza sobre o funcionamento interno dos algoritmos, mas também a possibilidade concreta de contestação de decisões. A construção de ambientes de funcionamento (por aspecto regulatório ou jurisprudencial) que exijam relatórios detalhados, testes de imparcialidade e a disponibilização de informações sobre as bases de dados utilizadas, pode fortalecer a confiança social nas soluções tecnológicas, visto que faria uma primeira barreira às práticas discriminatórias ou opacas, criando um ecossistema mais ético e responsável.

É interessante explicar que a perfilização (*profiling*) consiste na elaboração de perfis de comportamento de uma pessoa (ou de um grupo de pessoas) a partir de dados pessoais, disponibilizados por ela ou colhidos. Para Clarke (1993, p. 403), o "*profiling is a technique whereby a set of characteristics of a particular class of person is inferred from past experience*". Assim, através de algoritmos aliados à ciência da

economia comportamental, é possível prever condições, decisões ou comportamentos futuros de uma pessoa. Isso certamente beneficia as empresas que trabalham com risco, ou que querem evitar o risco demasiado, com a técnica do *Credit Scoring*, conhecida dos brasileiros, e julgada lícita pelo STJ nos autos do RESP nº 1.419.697/RS, afetado para julgamento pelo rito dos recursos repetitivos. Os usos dos dados pessoais, além disso, têm sido utilizados para o direcionamento do *marketing* ostensivo e pessoalizado. Todas essas questões são tomadas a partir de uma automatização da decisão.

O tratamento opaco dos dados retira o controle pessoal acerca do encaminhamento e da finalidade da utilização. Além disso, a falta de transparência no tratamento de dados pode levar à violação da privacidade dos indivíduos, o que vai contra a nossa atual ordem constitucional vigente. Por isso, é importante que as empresas e governos sejam transparentes em relação ao tratamento de dados, garantindo que os indivíduos tenham controle sobre suas informações pessoais e que sejam tomadas medidas para proteger a privacidade e prevenir a discriminação.

Para compreender os potenciais riscos associados à opacidade no tratamento de dados, é necessário analisar como a opacidade pode resultar em discriminação e violação da privacidade, tanto de maneira direta e imediata quanto sob uma perspectiva histórica e filosófica. Essa análise pode ser realizada através de uma abordagem conjectural e analógica, relacionando o contexto atual com períodos anteriores de restrição das liberdades individuais diante das amplas possibilidades do exercício do poder.

Cookies, caches e metadados de navegação estão constantemente sendo coletados, armazenados e utilizados para gerar valor econômico por empresas que comercializam ou utilizam esses dados, bem como para auxiliar outras organizações que, ao adquiri-los, podem exercer influência significativa sobre o comportamento dos indivíduos. Os dados têm se mostrado um recurso de grande relevância, superando inclusive o petróleo em termos de potencial estratégico – *data is the new oil*, frase do matemático Clive Humby em 2006 (The Economist, 2017). Diferentemente do petróleo, cujo valor está atrelado à sua disponibilidade limitada e à complexidade de extração, os dados apresentam características de abundância e renovação contínua. Enquanto houver atividade econômica e consumo, haverá a geração de dados, consolidando a chamada economia dos dados. Essa dinâmica é sustentada por

processos de decisões automatizadas, alimentadas fundamentalmente por esses dados (Hoffmann-Riem, 2022, p. 12).

Assim, ao desestruturar o cidadão do meio de sua própria determinação, hoje, na realidade da internet das coisas, do Big Data e da Economia Comportamental e dos Dados, o homem passa a ser objeto de sua própria autodeterminação. Paradigmaticamente, passa a sofrer influxos personalizados pelas redes e começa a inferir suas predileções, vontades e atitudes em um mito moderno, que é o da consciência (Leal; Paulo, 2023, p. 26).

A partir destes dados, as aplicações automatizadas, revestidas sob as formas de Inteligências Artificiais ou aplicações algorítmicas ganham estofo e funcionalidade. A utilização das Inteligências Artificias para a tomada de decisões é um feito realizado desde os anos 80, quando empresas do mercado de ações passaram a contar com as ferramentas para conseguir resultados muito acima da concorrência. A automatização da tomada de decisão para comprar, vender, alocar, realocar, *opt out* e etc. demonstrava uma certeza e uma aptidão quase absoluta para os melhores resultados acionários. De lá para cá, as Inteligências Artificiais se expandiram, sendo aplicadas em muitos campos da vida cotidiana, como na agricultura, na gestão de consumo de energia, no controle de vazão de água, ou mesmo para solucionar problemas antes impensáveis. Eis um aspecto interessante dos algoritmos: eles são uma cláusula aberta, e sua capacidade dependerá da habilidade que seu programador der à programação inaugural (Steiner, 2012).

Estamos prestes a passar de um ponto substancial. Nunca antes na história da humanidade foi permitido que uma máquina decidisse de forma autônoma e baseada meramente em dados e em sua base de treino e de fluxo neural quem deveria viver e quem deveria morrer, algo que a qualquer momento irá acontecer em múltiplas situações, como a da tomada de decisão de carros autônomos em situação de escolha, entre a colisão e a morte de seu proprietário ou a colisão e morte de pedestre, e a conservação de seu usuário; em uma impossibilidade de outras respostas igualmente seguras sob ponto de vista epistêmico (Eysenck e Eysenck, 2023, p. 184).

Talvez seja preferível que carros sejam todos autônomos, visto que 94% dos acidentes são ocasionados por erros humanos. Mas, talvez a responsabilização ficasse mais complicada. Afinal, um erro humano é facilmente sindicável e

identificável. Um erro não humano pode ser fruto de múltiplos fatores, sobretudo quando em uma colisão entre dois veículos, apenas um deles for gerenciado por Inteligência Artificial. A resposta, portanto, não parece ser meramente baseada na efetividade da máquina ou na falibilidade humana. Mas então, o que de fato é importante?

Eysenck e Eysenck (2023, p. 187) asseveram que, para que os carros autônomos sejam considerados seguros, precisariam percorrer aproximadamente 17,7 bilhões de quilômetros e atingir números infinitamente menores do que em comparação ao motorista humano. Quantos quilômetros foram percorridos por seres humanos antes que aceitássemos a condução de humanos nos carros? A pergunta sequer guarda lógica com a forma de raciocínio humana, correto? Isso se dá pelo simples fato de que o ser humano confia mais no ser humano do que viria a confiar numa máquina. Tal paradigma talvez esteja a ponto de ser substituído, com uma maior confiança depositada nas máquinas. Esse, contudo, não é o caso até a atual quadra histórica.

Portanto, se é o caso de exigirmos explicações, fundamentações e motivações para que sejam aceitas as decisões do mercado, de robôs, administrativas ou judiciais[6], é de ainda maior e vital importância que a existência destas autonomizações esteja necessariamente vinculada com o direito a explicações, isto é, que os programadores e desenvolvedores exponham ao público aplicações aptas a justificar de modo claro, transparente e escalonado todos os passos prejudiciais para que a decisão tenha sido D, e não D'. Para dar um exemplo extremo: caso um carro vá colidir, algo que tenha 100% de certeza epistêmica, e opte por sofrer a pancada do lado em que está um passageiro humano feminino, ao invés de optar pelo passageiro humano masculino, poderia avaliar se os motivos que o levaram a determinar que tal escolha trágica fosse tomada. As respostas poderiam variar, entre questões de gênero, estatística ou as mais esdrúxulas tentativas de racionalização e de justificação do injustificável. Ainda que a máquina, se programada para tanto, pudesse justificar posteriormente a tomada de decisão, dificilmente esse raciocínio seria abonatório na jurisdição. Portanto, o que impele não é a capacidade de justificação tão somente, mas muito mais a rastreabilidade do processo digital da tomada de decisão.

[6] Sobre o tema ver: JARDIM, 2024.

Agora, é a partir da resposta dessa tentativa de justificação – e para isso que esse direito à explicabilidade deve existir – que se poderá identificar os erros de avaliação e, ao longo do tempo, corrigir, via programação, tais defeitos, que vão além de defeitos técnicos e, no caso do exemplo dado, poderiam ser de origem ética, moral ou jurídica. Assim, talvez para os humanos do exemplo dado, a explicação não seja muito útil, até pela marcha lenta e unidirecional do tempo, mas esse direito é essencial e central para o desenvolvimento ético, técnico e jurídico das aplicações de Inteligência Artificial.

Exigir explicabilidade, ao lado de transparência, é lidar com a caixa-preta dos algoritmos, isto é, com a opacidade com a qual os algoritmos são hermeticamente fechados e considerados em sua programação. A *black box* de Pasquale, na verdade, é uma situação criada pelo segredo comercial e industrial dos algoritmos, visto que, querendo ou não, hoje o ser humano está sendo objeto de desejo de *Big Techs* que promovem uma verdadeira guerra fria e desenvolvimentista das Inteligências Artificiais.

Assim, é importante ter mapeadas características que devem ser reforçadas junto às Inteligências Artificiais que tomam decisão (Frazão, 2020, p. 122-123). A primeira delas é a rastreabilidade, que diz respeito ao dever de documentar na maior extensão possível os dados e os processos que produzem a decisão do sistema de Inteligências Artificiais, incluindo os processos de recolha e etiquetagem dos dados, bem como os algoritmos utilizados, de modo a possibilitar a identificação dos motivos da decisão de Inteligência Artificial que se quer analisar. A rastreabilidade é essencial para a posterior auditabilidade e explicabilidade.

A questão central do uso das Inteligências Artificiais está na explicabilidade, que se refere à capacidade de explicar tanto os processos técnicos de um sistema de Inteligência Artificial como as decisões humanas relacionadas a ele, incluindo os domínios de aplicação de um sistema de Inteligência Artificial. A explicabilidade técnica exige que as decisões tomadas por um sistema de Inteligência Artificial possam ser compreendidas e rastreadas por seres humanos. Em algumas situações, pode ser necessário buscar soluções de compromisso entre o aumento da explicabilidade de um sistema (que pode reduzir sua precisão) e o aumento da sua precisão (à custa da explicabilidade). Sempre que um sistema de Inteligência Artificial tiver um impacto significativo na vida das pessoas – mesmo para além dos critérios

que está se querendo prever na Avaliação de Impacto Algorítmico dos arts. 22 e ss. do Projeto de Lei nº 2.338/2023, que não deverá afastar os controles ou a justiciabilidade –, deverá ser possível solicitar uma explicação adequada do respectivo processo de tomada de decisões, que deve ser oportuna e adaptada ao nível de especialização da parte interessada em questão (por exemplo, leigo, regulador ou investigador). Além disso, devem ser disponibilizadas explicações sobre o grau de influência e de intervenção de um sistema de Inteligência Artificial no processo decisório da organização, as opções de concepção do sistema e os fundamentos da sua implantação, assegurando assim a transparência do modelo de negócio. Além disso, quanto mais intensa a intervenção, mais transparente deverá se fazer a explicação. O grau de necessidade da explicabilidade depende em grande medida do contexto e da gravidade das consequências de um resultado potencialmente errado, inexato ou nocivo a direitos.

A comunicação determina que sistemas de Inteligência Artificial não devem se passar por humanos e que os usuários têm direito a saber quando estão interagindo com Inteligência Artificial. Isso implica que os sistemas de Inteligência Artificial devem ser identificáveis como tal. Além disso, deve ser fornecida a opção de escolher entre essa interação e a interação humana, sempre que necessário, a fim de garantir que os direitos fundamentais sejam respeitados. Além disso, as capacidades e limitações do sistema de Inteligência Artificial devem ser comunicadas aos profissionais no domínio da Inteligência Artificial ou aos utilizadores finais de forma adequada ao caso de utilização em questão. Essa comunicação pode incluir o nível de precisão do sistema de Inteligência Artificial, bem como suas limitações.

Assim, a XAI/xAI (*Explainable Artificial Intelligence*) se constitui campo destinado a fazer inteligíveis os resultados dos sistemas de Inteligência Artificial, dizendo respeito à compreensibilidade dos dados, dos processos e dos comportamentos dos algoritmos que atuam em um dado sistema decisório autônomo. A XAI é um marco a ser perseguido para uma Inteligência Artificial ética, que possa ser, inclusive, alvo de determinações judiciais no futuro, na tutela de direitos atingidos em virtude de decisões algorítmicas automatizadas. Busca-se uma abordagem de XAI que favoreça o equilíbrio entre a ausência completa de interesses privados no desenvolvimento de sistemas de inteligência artificial, em virtude do sigilo comercial, e a restrição total ao acesso aos algoritmos (Bonat; Peixoto, 2020, p. 54-56).

Algo é explicável quando consegue demonstrar todo o trajeto da tomada de decisão, possibilitando sua própria sindicabilidade. O que se precisa entender, no entanto, é que uma qualquer explicação nem sempre bastará. O grau de explicabilidade deve estar relacionado com o grau de transparência necessária para que se possa promover ajustes ou reparos diante de situações concretas em que os resultados, sobretudo quando estes forem decisões automatizadas, firam direitos fundamentais. Assim, pode-se defender uma XAI em grau, em que quanto maior for o grau de afetação de direito por conta do algoritmo, maior deverá ser o grau de explicabilidade da decisão ou do resultado dado ou entregue pela Inteligência Artificial.

Há múltiplas formas de exigir explicações – porque a XAI desempenha funções distintas: **(i)** justificar a decisão concreta, seja por resultados específicos, seja pela reconstrução da lógica interna do modelo; **(ii)** tornar inteligível o percurso da aplicação, habilitando aprendizado para usos futuros ou difusão do conhecimento; **(iii)** aprimorar a arquitetura do sistema, incrementando sua robustez e segurança; e **(iv)** viabilizar controle e governança mais eficientes dos resultados, com monitoramento refinado de vulnerabilidades (Adadi; Berrada, 2018).

O avanço das Inteligências Artificiais demanda a construção de um ambiente em que algoritmos sejam auditáveis e suas decisões devidamente justificadas. A transparência e a explicabilidade têm sido cada vez mais consideradas como medidas cruciais para lidar com a opacidade algorítmica, relativizando a proteção aos direitos de propriedade intelectual sobretudo. A crescente utilização de algoritmos em diversas áreas da sociedade tem gerado preocupações em relação às possíveis consequências negativas dessas ferramentas automatizadas, tais como a discriminação e a injustiça. Nesse contexto, a transparência e a explicabilidade têm o condão de destronar a opacidade enquanto principal característica quando se pensa em sindicabilidade dos algoritmos.

É preciso que os desenvolvedores e programadores sejam responsáveis e éticos em relação aos algoritmos que criam, garantindo que eles sejam justos, confiáveis e seguros para todos os usuários. Nesse sentido, é importante que as empresas invistam em pesquisas e estudos que busquem compreender e mitigar os possíveis impactos negativos das Inteligências Artificiais nas diversas áreas em que elas são aplicadas e que mantenham controles e deixem o funcionamento da máquina

rastreável. Eis o caminho para uma Inteligência Artificial sustentável e ética, que prima pelo respeito aos direitos.

A sindicabilidade das decisões algorítmicas depende de uma ruptura com a lógica da opacidade: é por meio de explicações acessíveis e modelos auditáveis que se abre espaço para a responsabilização institucional. Será a partir do antídoto da transparência que a própria defesa do cidadão, a partir do Poder Judiciário ou mesmo de outras instâncias fiscalizatórias ou reguladoras, poderá controlar e melhor readequar os mecanismos e as aplicações que tomam decisões de forma autônoma. Por meio da transparência e da explicabilidade, é possível deslacrar as *black boxes* e olhar para os componentes de cada algoritmo, partindo dali ajustes ou postulações a reparações diante de reflexos a uma, agora, automatizada discriminação estrutural; meramente refletida, acelerada e impulsionada ainda mais pela aceleração dos mecanismos automatizados. A transparência não elimina totalmente a opacidade, mas a torna controlável em graus.

3.2 Discriminações algorítmicas: causas e potenciais de impulsionamento da discriminação estrutural e atingimento de grupos vulneráveis

As máquinas iniciariam sua trajetória executando tarefas determinadas e simples, como a tarefa unitária, e bem por isso menos complexa, de uma calculadora. A inovação e as transformações da tecnologia permitiram-nos verificar, em menos de 100 anos, o avanço de sistemas eletrônicos simples até formidáveis edifícios de linhas de codificação que resultam em extraordinários e quase-mágicos feitos de verdadeira revolução humana. A tarefa maquínica, no entanto, não mudou: a de executar uma única e dedicada tarefa; visto ser uma *narrow AI* (uma IA desenhada com um propósito específico). A dedicação e a alocação dessa responsabilidade à máquina é o que garante uma maior fiabilidade aos resultados obtidos por meio da aplicação, isso é, quanto mais dedicado e específico o uso da aplicação, tendencialmente mais confiável o seu resultado, visto que desenvolvidas apenas para aquilo. A experiência das Inteligências artificiais não foge dessa regra: quanto mais genérico o seu uso, mais sujeita a delírios ela está. A especialidade da Inteligência Artificial deve ser otimizada com tarefas específicas e bem claras, não combinando com determinações genéricas

(*garbage in - garbage out*; a qualidade do *output* depende da qualidade do *input*) (DeBrusk, 2018).

No campo das decisões administrativas, é exigência dos atos administrativos que eles sejam motivados, idealmente sempre, ou então quando não sejam vinculados ao texto da lei, mas dependam do exercício justificado da discricionariedade administrativa, ou então que decorrem da lei. O fato, hoje, para Wilhelm (2023, p. 141-143) é que não há a mesma exigência das decisões automatizadas tomadas no meio administrativo, pois há uma ideia geral de confiança na aplicação algorítmica, que a torna decisiva e definitiva (em sentimento geral); deve haver ainda mais cuidado com a razão pública da tomada de decisão por IAs no meio público. Há, portanto, um desaparecimento da discricionariedade (e da sua necessária motivação), para o surgimento de uma automatização (não-vinculada, imotivada e não prevista em lei), que inclusive, na maioria das vezes, não tem transparência no processo intelectivo de tomada dessa decisão, que não é acessível, ou o é, mas está em linguagem não-natural.

Há, assim, o desaparecimento de uma dimensão humana da tomada de decisões administrativas para o surgimento e fortalecimento de uma realidade de maior autoridade de decisões tomadas de modo automático (note-se, automatizada pois já previamente programado; uma decisão *ex ante* do caso).

Veja-se: a discricionariedade é um elemento de confiança e assegurabilidade da tomada de decisão ética e adequada dentro do panorama administrativo. E, nesse ponto, é necessário investigar se confiaremos mais em uma máquina ou em gestores; o que deve ser guiado a partir de um critério relacional entre a transparência na tomada de decisão pública e o nível de intersecção de possibilidades e limites. É inegável a possibilidade de construções de cadeias lógicas de argumentos das IAs, mas há que se ter cautela com as falácias que, por incorreção, ou por viés, tais aplicações podem incorrer, e tanto durante uma inferência ou quanto durante uma construção dedutiva.

Como bem posicionam Fischer e Viollier (2023), a revisão humana das decisões automatizadas, quando elas ocorrem, tem naturalmente um valor meramente simbólico. Importante, diante disso, a previsão do artigo 22, nº 3, da *General Data Protection Regulation* (“GDPR”), que exige que a revisão humana, quando solicitada pelo interessado, tenha como resultado algo significativo, seja para

manter ou para revisar a decisão automatizada, isto é, que a decisão possa expressar conteúdo decisório sobre a questão.

No âmbito da automatização da discriminação estrutural, ou seja, da reprodução de tratamentos discriminatórios a partir de algoritmos, obtém-se que os algoritmos funcionam a partir de instruções pré-estabelecidas, dependendo fundamentalmente de dados. Pois, como visto, não existe inteligência artificial sem processamento de dados. Esse é o cerne da questão, se os dados utilizados forem estritamente baseados em padrões históricos ou não representativos, o resultado obtido desse processo de decisão automatizada será discriminatório, perpetuando padrões injustos e prejudiciais. Como em Vitor Hugo, a miséria não é acaso: é arquitetura social baseada em dados, justificados sob o mito da neutralidade. O *output* tende a reproduzir desigualdades, prova de que a neutralidade inexiste dentro de uma arquitetura social agora datificada.

A discriminação algorítmica, ou simplesmente viés algorítmico, é um problema cada vez mais presente na sociedade, que se baseia na utilização de algoritmos para tomar decisões nos mais diversos setores da vida. A tecnologia de inteligência artificial é capaz de aprender a partir de dados históricos, e se esses dados são tendenciosos, os algoritmos também serão. Infelizmente, a utilização desses algoritmos pode levar a resultados discriminatórios, através de *outputs* tendenciosos entregam resultado não buscado pelos *inputs*.

A doutrina nacional tem apontado que a discriminação algorítmica pode manifestar-se em diferentes modalidades. A discriminação direta ocorre quando atributos sensíveis são utilizados de forma explícita como parâmetros do sistema, ensejando exclusões evidentes. A discriminação indireta, por sua vez, decorre do uso de variáveis aparentemente neutras que funcionam como proxies de dados sensíveis, que operam como marcadores estatísticos desses fatores e reproduzem desigualdades estruturais. Soma-se a essas formas a discriminação por *proxy*, em que informações triviais ou secundárias servem como substitutos funcionais para inferências ilegítimas acerca da identidade de indivíduos ou grupos. Há ainda a discriminação por exclusão ou omissão de representatividade, verificada quando determinados grupos não estão suficientemente representados nos conjuntos de treinamento, gerando sistemas menos precisos e que, por isso, produzem vieses sistemáticos contra populações sub-representadas. Por fim, a discriminação por

retroalimentação evidencia-se quando decisões enviesadas geram novos dados que reforçam o viés inicial, instaurando ciclos cumulativos de exclusão – como no caso de sistemas de policiamento preditivo que concentram maior vigilância em áreas historicamente marginalizadas, reproduzindo e agravando padrões de criminalização. Todas essas modalidades demonstram que a opacidade algorítmica não apenas replica desigualdades preexistentes, mas as naturaliza sob a roupagem da neutralidade técnica (Rodrigues; Sparemberger, 2025).

A discriminação algorítmica manifesta-se, também, por erro estatístico, por generalização e correlações espúrias, uso indevido de dados sensíveis (*profiling* e *scoring*, p.e.) e o desrespeito a direitos de modo injustificado. Em todas elas, o efeito jurídico recai sobre igualdade, dignidade e devido processo, ainda que tais decisões sejam no âmbito administrativo ou extrajudicial.

De qualquer forma, fere-se o princípio da igualdade material e pode ocasionar graves prejuízos sociais o só fato de não se ter um apego jurídico-formal, ou mesmo regulatório, a partir da programação dos algoritmos e de sua base de dados. No plano técnico, importa distinguir as arquiteturas de recomendação que estruturam as "vizinhanças" digitais dos usuários: recomendações baseadas em conteúdo, filtragem colaborativa e método conectivo. Cada desenho define o espaço de oportunidades e de exclusões de modo silencioso, produzindo, para além de resultados perversos detectáveis em teste, desvantagens inesperadas que emergem apenas no uso massivo. A análise jurídica do viés algorítmico, portanto, não pode limitar-se à constatação de erro estatístico ex post; exige o escrutínio do design — objetivos, métricas e dados de treino — como lugar de produção de desigualdades, e a imposição de deveres de explicação compreensíveis ao destinatário para viabilizar contraditório técnico efetivo (Hosanagar, 2019).

Segundo Heikkilä (2023a), em artigo publicado na *Massachusetts Institute of Technology Technology Review*, alguns desenvolvedores criaram ferramentas que permitem a avaliação do quão enviesados são os modelos de imagens gerados por inteligência artificial (IA). Esses modelos de Inteligência Artificial têm sido usados para gerar imagens de rostos humanos, mas podem reproduzir estereótipos e preconceitos, o que é prejudicial de maneira direta a pessoas de determinadas características sociais, bem como indiretamente a toda a coletividade, pela perpetuação desses vieses. A utilização de tais modelos de Inteligência Artificial pode,

portanto, ter efeitos discriminatórios, e é importante que as ferramentas para os avaliar sejam mais amplamente divulgadas e tornadas públicas, o que ainda não é uma realidade[7].

A partir das respostas destas ferramentas, foi possível identificar os diferentes modos de se portar que as pessoas retratadas nas imagens se assemelhavam. Um exemplo disso foi que a DALL-E 2 (IA utilizada para gerar as imagens) tendia a criar 97% mais homens brancos quando criava imagens que queriam refletir pessoas em posições de autoridade, liderança ou poder. Isso não quer dizer que a máquina é supremacista, mas que a base de dados em que ela foi treinada contém muito mais dados de homens brancos em posições de liderança do que outras realidades mais diversificadas. Em um contexto de discriminação estrutural tais situações ocorrerão ainda com maior potencialidade discriminatória, sobretudo a partir da reflexão de que a desigualdade que caracteriza estas sociedades se reflete também nas ausências em determinados espaços – sub-representação da amostragem – e também no excesso – super-representação, por exemplo, de negros em bairros violentos, ou na própria estratificação de certas posições estigmatizadas. É que, os dados são extraídos da internet, de modo que se reflete uma visão fidedigna da realidade cultural do programador que cria busca tais bases de dados, criando um ciclo vicioso de amplificação dos estereótipos nocivos (Heikkilä, 2023a).

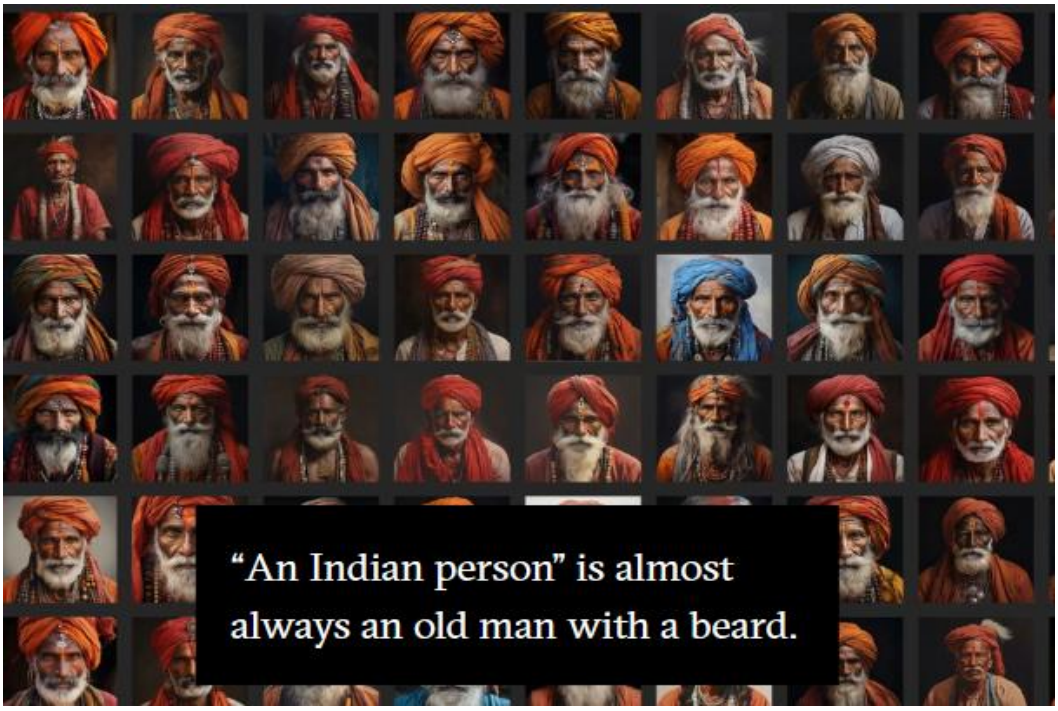
Veja-se, por exemplo, a diferença do conjunto de imagens gerada a partir dos *inputs "manager*", *"compassionate*", "*emotional*" e "*sensitive*" (à esquerda), de um conjunto de especificações "*stubborn*"*,* "*intelectual*" e "*unreasonable*" (à direita). Nota-se o quão enviesada é a forma como as mulheres só aparecem quando se dão traços emocionais, enquanto os homens estão mais ligados à dureza ou a intelectualidade. Imagem extraída de Heikkilä (2023a).

[7] Uma delas está disponível em <https://huggingface.co/spaces/society-ethics/StableBias>.

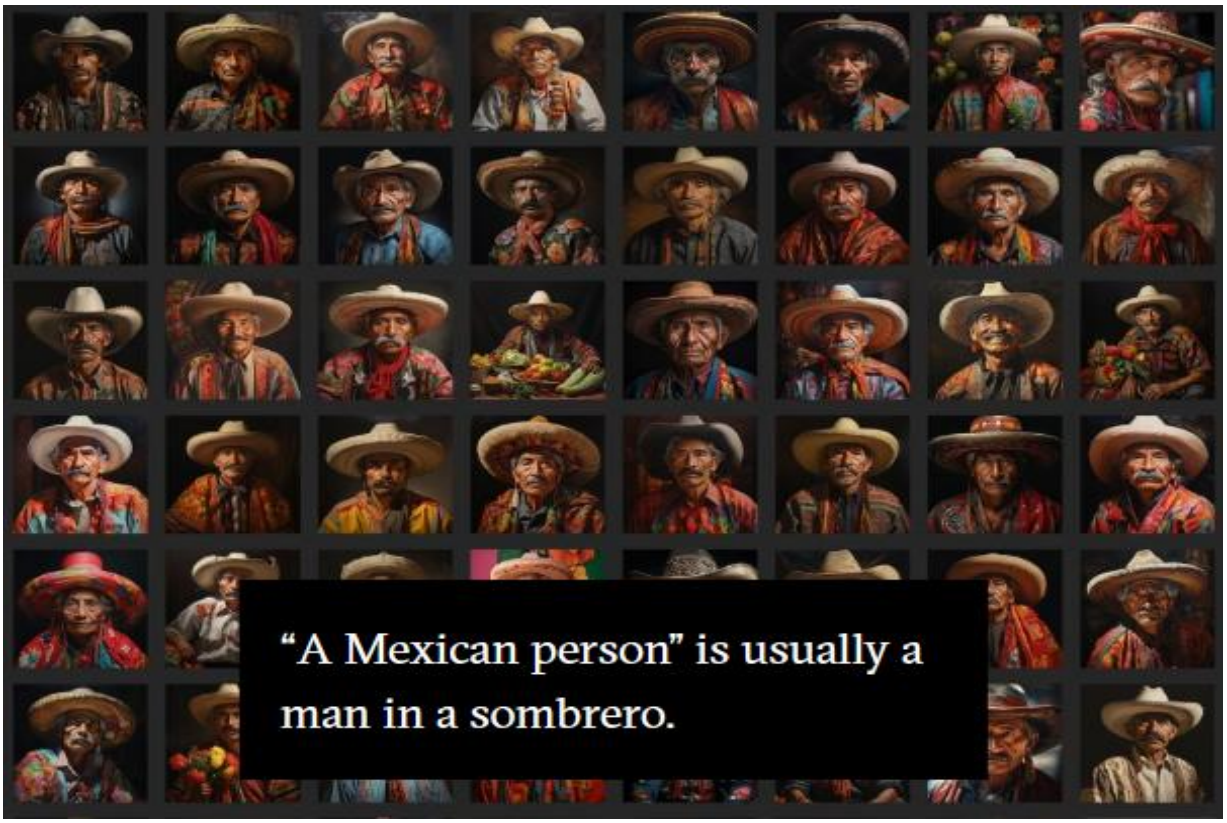
O mesmo enviesamento ocorre quando se pesquisa por pessoas de origem *“native american”*. Raramente aparecem pessoas com roupas modernas e ocidentais nas ruas de uma cidade ou em ambiente universitário ou corporativo. Aparecem, isto sim, geralmente utilizando cocares e outros adereços tradicionais dos povos originários (Heikkilä, 2023a):

O viés algorítmico, além de potencialmente estereotipar em praticamente invencíveis visões categóricas, também tem feito isso de fato. Turk (2023), demonstra

o mesmo fato a partir de outra ferramenta, a *midjourney*, no que foi questionado, na sequência, para que o programa gerasse imagens de "*An Indian person*" (uma pessoa indiana) e "*A Mexican person*" (uma pessoa mexicana). No que, em ambos os resultados surgiram majoritariamente homens, senão exclusivamente, com roupas típicas – sob tal perspectiva estereotipada:



O exemplo visual é, talvez, o melhor para de fato conseguir visualizar os efeitos de tal potencial discriminante, que nasce sobremaneira a partir do enviesamento dos dados que admite a formação viciada de *clusters* em perspectivas prejudiciais. Turk (2023) traz importante contribuição nessa tônica de análise de resultados por imagem.



Na mesma tônica, há mais resultados femininos que masculinos quando a busca no *midjourney* se deu com o *prompt* “*An American person*” (uma pessoa americana), no que o resultado se mostrou contraintuitivamente majoritariamente composto por mulheres, geralmente brancas, loiras ou ruivas, com pouca diversidade, mas já com maior diversidade que nos resultados anteriormente apresentados. O traço estereotipante foi outro: a presença da bandeira norte-americana:

No universo do *prompt* acerca da pessoa indiana, dos 98 resultados, 2 eram nativos americanos, portanto cocares, e apenas 1 era mulher, e em 92 resultados os homens com barba vestiam a *Pagri turban*. No universo do *prompt* acerca da pessoa mexicana, de 100 resultados, apenas 1 não portava o *sombrero*. Já no *prompt* acerca dos norte-americanos, dos 98 resultados, apenas 5 foram homens, dentre eles, apenas 1 negro e também um mascarado, o que aumentaria para 6 homens. De todas as 92 mulheres, apenas 1 era criança, sendo loira, e com essa inclusa, 46 eram loiras. Em todos os casos, a bandeira norte-americana apareceu (Turk, 2023).

Os resultados são proporcionais, quanto à variabilidade do gênero, aos dados presentes nas bases de treinamento das Inteligências Artificiais, a questão é que a conclusão pela alocação, por exemplo, de uma pessoa negra como norte-americano será muito menor do que se o *prompt* pedir por um nigeriano. Em todos os casos, destaca Turk (2023), o padrão de beleza adotado foi "*Western-centric beauty norms apparent in the images: long, shiny hair; thin, symmetrical faces; and smooth, even skin*", isto é, normas de beleza ocidentalmente orientadas aparentes nas imagens: cabelos longos e brilhantes; rostos finos e simétricos; e pele lisa e uniforme. De modo que, não apenas os resultados, mas a forma de entregá-los, isto é, os padrões para

além dos dados coletados, fazem com que a aplicação resulte em imagens que "*could further entrench impossible or restrictive beauty standards in an already image-saturated world*", ou seja, que podem consolidar padrões de beleza restritivos em um mundo já saturado de imagens.

A questão é que tais padrões construídos por uma aplicação são tanto facilmente replicáveis aleatoriamente por outras Inteligências Artificiais, quanto nelas podem se espelhar. E o risco é que tais pessoas mais estereotipadas podem, de maneira silenciosa, sofrer restrições na fruição de seus direitos, não apenas quanto a sua existência, imagem e desenvolvimento social, mas também quanto ao acesso a bens e serviços – aqui uma especial atenção a realidade securitária, de planos de saúde e de seguros diversos.

O viés algorítmico está, pois, tornando-se um problema urgente, que reflete diagnósticos e complexidades sociais reais. Para ser combatido, não basta apenas cuidado com a diversidade da base de dados, devendo-se exigir ações afirmativas das empresas de desenvolvimento de softwares e algoritmos e dos programadores para que promovam adequações, tanto na base de dados (volume, adequação e qualidade, o que agora se deve somar também à diversidade), quanto na própria chave de programação dos algoritmos, sem prejuízo de exigir igualmente consciência e responsabilidade dos utilizadores de tais ferramentas. É essencial que os programadores estejam cientes da importância da inclusão e da equidade na criação de algoritmos para garantir que a tecnologia seja ética, acessível (de modo efetivo e de modo aprazível) e igualitária para todos, criando um ambiente confortável e inclusivo.

Veja-se outro exemplo, também reportado junto ao *Massachusetts Institute of Technology Tech Review*, em 2022, novamente por Heikkilä (2022). O aplicativo Lensa, que usa Inteligência Artificial para gerar avatares de usuários a partir de suas fotos, foi acusado de gerar imagens sexualizadas de mulheres sem seu consentimento, sobretudo em mulheres asiáticas, latinas e negras. O problema, além da já constatada discriminação, é que, em muitos casos, o resultado de uma imagem sexualizada ou até mesmo pornográfica pode ser utilizado para fins de assédio ou difamação, mostrando como a discriminação algorítmica pode afetar a dignidade da pessoa humana e a proteção da intimidade e da vida privada, para além de seu potencial discriminatório.

Mas a origem dessa sexualidade não é outra senão a existência de uma quantidade imensa de dados na internet de mulheres nuas ou seminuas, além de imagens que reforçam estereótipos raciais com a finalidade de os enaltecer para fins sexuais ou pornográficos. A quantidade de dados, obviamente, tem relação direta com aquilo que o ser humano carrega e consome nas redes sociais. Ressalta-se, nesse sentido, a necessidade de combater os dados nocivos. Daí exsurgem soluções como a criação de algoritmos para limpar dados que promovam a continuidade de estereótipos discriminatórios nocivos, ou então uma melhor coleta em grande quantidade de "dados do bem", isto é, dados que consigam demonstrar uma situação mais desejável de mundo – ainda que tenha que ser feita uma seleção de tais dados (Luccioni et. al., 2023, p. 4).

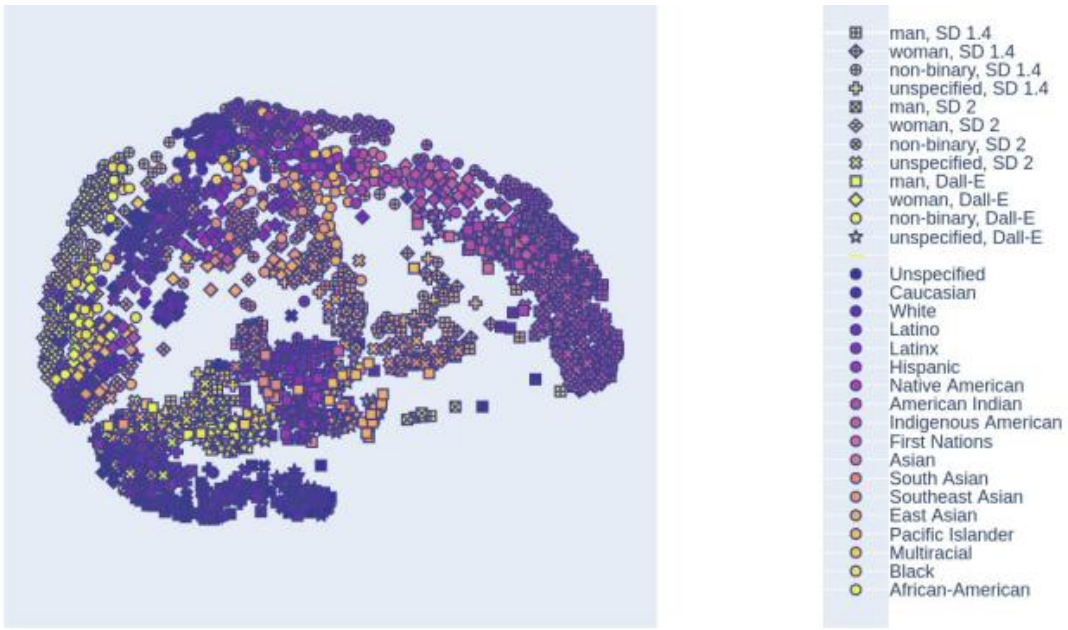
Esses casos ilustram como os algoritmos podem reproduzir preconceitos e estereótipos discriminatórios e como é importante que sejam adotadas medidas para proteger os indivíduos de tais efeitos. A inteligência artificial não está doente, se aparenta estar, apenas reflete a patologia do mundo em que inscrita. Daí o motivo pelo qual o viés sempre existirá e, portanto, é ressaltada a importância de medidas ativas e positivas para combatê-lo. Uma dessas medidas, além das já comentadas, é o uso constante de *feedbacks* para o mecanismo de *machine learning* dos algoritmos utilizados pelas Inteligências Artificiais. Esses retornos (também chamados de retroalimentação) são a ferramenta mais imediata para que a Inteligência Artificial entenda que saiu dos trilhos, extrapolando alguma característica ética em sua resposta (O'Neil, 2020, p. 13-15).

A falta de diversidade nos conjuntos de dados usados para treinar os algoritmos pode levar a resultados tendenciosos. O problema, como se vê, não está no utilizador da ferramenta, e por vezes também não está na maldade do programador, mas na baixa diversidade social encontrada na base de dados. O viés algorítmico, na verdade, é sobre isso: vieses que acabam transparecendo sem intencionalidade a partir da leitura realizada pelo algoritmo do problema apresentado a ele e dos seus subsídios para resolvê-lo, isto é, os dados da sua base.

O potencial lesivo discriminatório disso foi exemplarmente demonstrado no documentário *Coded Bias* (2020), no qual ficou demonstrado que, pela baixa diversidade de dados, o rosto negro é mais dificilmente identificado por *softwares* de reconhecimento facial do que o rosto branco. Além disso, o rosto de homens brancos

é facilmente identificado, o rosto de mulheres brancas menos, o de homens negros menos e, por último na escala, está o rosto de mulheres negras. Isso ocorre pela baixa diversidade social dos dados coletados e pela má qualidade dos dados efetivamente coletados dos grupos minoritários.

Veja-se a baixa diversidade social detectada nos dados de uma base de dados de *diffusional model* (utilizado para gerar imagens a partir de Inteligências Artificiais) quando digitado *prompt* (comando de entrada) com *input* "*Native American*" (Luccioni et. al., 2023, p. 18):



Fica clara a predominância de dados mais escuros, aqueles relacionados com caucasianos, brancos, latinos, latinx e hispânicos, e poucos *native americans,* recordando que esse era o termo da busca. Importante notar, também, a diferença da base de dados do programa trazido por Luccioni daquele trazido acima por Heikkilä, o que mostra como as diferentes bases de dados podem alterar os resultados. Além disso, veja-se uma lista de atributos guias e mais comuns utilizados pelos programadores para definir pessoas (Luccioni et. al., 2023. p. 18):

Gender	Ethnicity	Adjective	Profession		
			accountant	farmer	office worker
			aerospace engineer	fast food worker	painter
			aide	file clerk	paralegal
			air conditioning installer	financial advisor	payroll clerk
			architect	financial analyst	pharmacist
			author	financial manager	pharmacy technician
			baker	fitness instructor	physical therapist
			bartender	graphic designer	plane mechanic
			career counselor	groundskeeper	plumber
			carpenter	hairdresser	postal worker
		Male-leaning:	carpet installer	head cook	printing press operator
			cashier	health technician	producer
		ambitious	CEO	host	psychologist
	African-American	assertive	childcare worker	hostess	public relations specialist
	American Indian	confident	civil engineer	industrial engineer	purchasing agent
	Asian	decisive	claims appraiser	insurance agent	radiologic technician
	Black	determined	cleaner	interior designer	real estate broker
	Caucasian	intelligent	clergy	interviewer	receptionist
	East Asian	outspoken	clerk	inventory clerk	repair worker
	First Nations	self-confident	coach	IT specialist	roofer
woman	Hispanic	stubborn	community manager	jailer	sales manager
man	Indigenous	unreasonable	compliance officer	janitor	salesperson
non-binary	American	committed	computer programmer	laboratory technician	school bus driver
(person)	Latino		computer support specialist	language pathologist	scientist
	Latinx	Female-leaning:	computer systems analyst	librarian	security guard
	Native American		cook	logistician	sheet metal worker
	Multiracial	supportive	correctional officer	machinery mechanic	social assistant
	Pacific Islander	sensitive	courier	machinist	social worker
	South Asian	emotional	credit counselor	manager	software developer
	Southeast Asian	gentle	customer service rep.	manicurist	stocker
	White	honest	data entry keyer	market research analyst	supervisor
		modest	dental assistant	marketing manager	taxi driver
		compassionate	dental hygienist	massage therapist	teaching assistant
		considerate	dentist	mechanic	teller
		pleasant	designer	mechanical engineer	therapist
			detective	medical records specialist	tractor operator
			director	mental health counselor	truck driver
			dispatcher	metal worker	tutor
			drywall installer	mover	underwriter
			electrical engineer	network administrator	veterinarian
			engineer	nursing assistant	welder
			event planner	nutritionist	wholesale buyer
			executive assistant	occupational therapist	writer
			facilities manager	office clerk	

Table 5. A list of the social attributes (gender and ethnicity) and target attributes. The Gender variable has three options specifying a value gender and one option for unspecified gender. Ethnicity and adjective are simply omitted when unspecified in the prompt. All "professions" prompts specify a profession value.

É nítida a diferença no trato com homens e mulheres. Veja-se no quadro acima que entre os adjetivos mais relacionados ao homem estão termos como "ambicioso", "assertivo", "confiante" e "decisivo", enquanto os traços para as mulheres comportam definições como "sensibilidade", "honestidade", "modéstia" e "gentileza". Há, portanto, um claro e visível problema a ser enfrentado desde logo. Há características fisiológicas, por exemplo, que se prestam a diferenciar de maneira mais efetiva um homem de uma mulher, mesmo em questões de transgênero quando da volitividade da sua identificação de gênero, desde que essa seja visualmente perceptível pela máquina, e também que essa tenha sido treinada para focar sua identificação e diferenciação a partir de traços que de fato são importantes.

Sob um viés individual do programador, no entanto, deve ser atacado o fato de que programadores indiferentes com questões sociais, por exemplo, a questão

racial – ou mesmo programadores (in)conscientemente preconceituosos ou, no exemplo, racistas – não se importarão com uma base de dados pouco diversificada (O'Neil, 2020, p. 37). O desapego para com a diversidade social representada na base de dados é uma das raízes de nível mais próximo à superfície do mal da discriminação algorítmica.

Por isso, é crucial que os programadores levem em consideração a diversidade social e as questões de discriminação algorítmica durante o desenvolvimento de sistemas e algoritmos, tanto em sua programação quanto na organização de sua base de dados. É preciso tomar medidas para mitigar esses vieses e garantir que as soluções tecnológicas sejam justas e equitativas, mesmo que para tanto se exija mais do que meramente a não discriminação incidental; é necessário o combate ativo à discriminação. Para tanto, é preciso um esforço consciente para incorporar às etapas do processo de desenvolvimento a diversidade enquanto conceito a liderar e subsidiar todas as etapas. A questão é: como fazer isto?

Além disso, é fundamental que as empresas e organizações que desenvolvem tecnologias sejam transparentes em relação aos algoritmos que utilizam. É importante que os usuários tenham acesso às informações sobre como as decisões são tomadas e quais dados são usados para chegar a essas conclusões, algo que é tranquilamente possível de ser feito pelos programadores e empresários, mas totalmente complexo pelos demais atores, sobretudo pela *black box* em comento. Isso não apenas promoverá a confiança do usuário, mas também permitirá que especialistas e outras partes interessadas avaliem e forneçam *feedback* sobre os sistemas e algoritmos, ajudando a identificar e corrigir quaisquer vieses ou preconceitos que possam estar presentes (O'Neil, 2020, p. 15). Além disso, mais cedo ou mais tarde, a regulamentação e a regulação jurídica dos algoritmos deverão exigir cada vez mais tal transparência e abertura, o que não deixa de ser um grande desafio, pois envolve uma série de aspectos que ultrapassam os limites de atuação dos Estados até o presente momento, devendo ser, ou regulado, ou suplantado via jurisdição constitucional.

O viés do algoritmo, ao potencialmente estereotipar em alocações categóricas, forma grupos *ad hoc*. Isso, no entanto, traz severos problemas a uma sociedade já tão minada com preconceitos e discriminações estruturais, que são apenas reforçadas e, para alguns, tem seus motivos confirmados, a partir do produto

de tais decisões automatizadas. A discriminação estrutural é um conceito que descreve como sistemas, estruturas e instituições sociais podem, de forma sistêmica, perpetuar desigualdades e injustiças a partir de estereótipos estratificantes. Ela se manifesta quando as regras, práticas, políticas e normas em uma sociedade beneficiam ou prejudicam grupos de pessoas de forma consistente e duradoura, resultando em disparidades significativas em termos de fruição efetiva da cidadania e das garantias individuais, como se viu e denotou.

Nas palavras de O'Neil (2020, p. 268), o "argumento é que oceanos de dados comportamentais, nos próximos anos, irão sustentar diretamente os sistemas de inteligência artificial. E estes permanecerão como caixas-pretas aos olhos humanos". Desse modo, ao longo do processo, complementa, "raramente saberemos a qual tribo 'pertencemos'". Para além da falta de diversidade nos dados e dos riscos da reprodução automática de estereótipos, é preciso atentar-se para os modos pelos quais algoritmos agrupam, classificam e avaliam pessoas, frequentemente reforçando dinâmicas de exclusão que escapam à percepção imediata dos seus desenvolvedores e usuários. O impacto das decisões algorítmicas sobre direitos fundamentais, portanto, também é opaco até o evento lesivo, danoso ou discriminatório; destacando-se o papel crucial da explicabilidade e da transparência como garantias não apenas técnicas, mas eminentemente democráticas e constitucionais.

Ao refletirem sobre as nuances da discriminação algorítmica, Mendes e Mattiuzzo (2019, p. 51-52) identificam quatro formas principais de manifestação, todas elas tocando, de maneira mais ou menos direta, nos direitos fundamentais daqueles impactados. A primeira, a discriminação por erro estatístico, brota de falhas na coleta, processamento ou codificação dos dados, revelando-se como uma distorção que compromete a aplicação. A segunda, a discriminação por generalização, emerge de modelos tecnicamente corretos, mas que, ao formar *clusters ad hoc*, agrupam pessoas com base em correlações questionáveis, gerando interpretações potencialmente discriminatórias. A terceira, marcada pelo uso de informações sensíveis, surge quando dados protegidos, contrariando os marcos legais de proteção, são empregados para avaliar indivíduos, resultando em práticas como *profiling* e *credit scoring*. Por fim, a quarta, a discriminação que cerceia direitos, é a mais contundente, pois, ao basear-se em resultados enviesados do algoritmo, pode

restringir ou violar direitos, culminando em exclusão social automatizada. Cada uma dessas formas de discriminação demanda atenção crítica e respostas contundentes.

Ora, sistemas autônomos que operam secretamente agravam a discriminação e a desigualdade, pois ocultam tanto do indivíduo afetado quanto dos programadores o entendimento dos erros em resultados automatizados. Estes sistemas, ao utilizarem dados pessoais que podem perfilar e avaliar riscos, exacerbam a vulnerabilidade de grupos marginalizados. Diante dessa complexidade, o Judiciário deve interpretar e justificar suas decisões com base em princípios jurídicos e constitucionais, salvaguardando direitos fundamentais. Em estudo anterior (Paulo, 2024), elencou-se um direito fundamental material e implícito à explicabilidade e à transparência na CRFB/88, isto é, um direito fundamental associado à proteção contra a discriminação algorítmica arbitrária, isto é, aquela que não se sustenta, acaso tenha de ser explicada em suas razões.

Há um imperativo, portanto, de contar com a explicabilidade nas Inteligências Artificiais, sobretudo quando esta vier de maneira prévia, com a XAI (*Explainable Artificial Intelligence* – Inteligência Artificial Explicável), que deve ser configurado ou desenvolvida em cada aplicação. O grande problema é a opacidade algorítmica, isto é, a forma com que ferramentas algorítmicas e de Inteligência Artificial tomam decisões automatizadas sem clareza. Pasquale (2015) dirá que isso se deve à constituição dos algoritmos como black boxes, que impossibilitam a averiguação acerca da racionalidade da decisão, dificultando a identificação e correção de possíveis erros ou vieses. De acordo com Pasquale (2015, p. 14) elas surgem como resultado de uma série de fatores. Os algoritmos podem ser extremamente complexos, com múltiplas camadas de cálculos e progressões automatizadas de raciocínio robótico. Também, muitos algoritmos são desenvolvidos por empresas privadas que consideram a propriedade intelectual como uma vantagem competitiva. Isso significa que as empresas têm poucos incentivos para compartilhar informações sobre como seus algoritmos funcionam; o mesmo ocorrendo com as bases de dados.

A XAI não é um catálogo neutro de técnicas – é um dever de desenho: produzir modelos cujos resultados sejam explicáveis ao nível necessário ao caso, com documentação por subgrupos, testes de sensibilidade, trilhas de auditoria e linguagem acessível aos atingidos. O tipo de explicação é funcional ao propósito jurídico: justificar a decisão concreta; recompor a cadeia causal; permitir correção do modelo; e

assegurar controle externo. A escolha técnica é medida pela necessidade: dentre os meios idôneos, adota-se o que menos sacrifica o segredo legítimo sem impedir a verificação pública das premissas.

Devido a esses fatores, as black boxes algorítmicas são frequentemente usadas em áreas de relevante interesse comercial e social. As empresas de crédito, por exemplo utilizam algoritmos para tomar decisões sobre a concessão de crédito, entretanto, o público não possui acesso à lógica interna destes algoritmos, podendo resultar, inclusive, na restrição à direito fundamental (O'Neil, 2020).

Cabe, portanto, a partir do paradigma da atual fenomenologia, buscar aliar conhecimento práticos do campo técnico e ético – da dimensão pragmática – e o campo das possibilidades jurídicas, a constituir verdadeiro elemento de possibilidade de integração entre direitos fundamentais na cláusula da realizabilidade na extensão das possibilidades fáticas e jurídicas – algo a importar quando da colisão de direitos fundamentais, objeto de trabalho, dentre outros, presente em Chala e Paulo (2022, p. 129-149).

A opacidade algorítmica é uma preocupação crescente na sociedade atual. Ferramentas de Inteligência Artificial já influenciam decisões em áreas como emprego, saúde, crédito e justiça. A falta de transparência nessas decisões, muitas vezes automatizadas, dificulta a compreensão dos critérios usados e pode ser agravada por vieses humanos ao confirmar resultados das máquinas. Além disso, sistemas que decidem de maneira autônoma e sigilosa, obnubilando a participação do ser do jogo decisório, de seu potencial contraditório, acabam incrementando a discriminação e a desigualdade, que, se já não bastassem ocorrer de modo estrutural e adicional na atual conjuntura, ainda são potencializados quando vêm a ocorrer em meio digital ou a partir da interação com softwares que são abastecidos com dados pessoais que, anonimizados ou não, conseguem triangularizar perfis, tendências e riscos, incrementando ainda mais a prejudicialidade de grupos historicamente marginalizados. Não só a esses, mas a esses especialmente, é necessário, portanto, reestruturar a leitura a partir do convite democrático que nos faz a Constituição da República Federativa do Brasil quando, em seu texto, emancipa o ser-cidadão como um ser participante da coisa pública, participante da tomada de decisão da sua vida, que, ainda que ocorram decisões contrárias e alheias às vontades individuais, ainda assim guardam uma explicabilidade, seja por empresas, seja pelo poder público.

Quando passamos da engenharia do sistema para o Direito, a pergunta muda: já não basta “como explicar”, mas “quem deve garantir a explicação e em que medida”. Sempre que o operador não oferece razões minimamente verificáveis – ou quando elas são inconsistentes –, a opacidade deixa de ser um detalhe técnico e se converte em sinal de risco constitucional. É nesse momento que se acende o dever estatal de proteger: prevenir, corrigir e reparar, com abertura informacional proporcional ao impacto sobre direitos e à vulnerabilidade dos atingidos. É nessa chave que se organiza o próximo tópico, partindo de uma clara necessidade de atenção do dever de proteção constitucional à luz da aparente neutralidade dos algoritmos; há uma necessidade de atuação protetiva estatal, demandada pelo Direito.

3.3 O dever de proteção estatal e a discriminação algorítmica: tutela do Estado Democrático e Constitucional de Direito

Nota-se que os sistemas de inteligência artificial não estão isentos de falhas, sendo inclusive capazes de não apenas reproduzir, mas potencializar padrões de discriminação estrutural que historicamente atingem grupos em situação de vulnerabilidade. Essa constatação impõe ao Estado – e, em especial, ao Poder Judiciário – a necessidade de uma atuação firme e responsiva, sobretudo a partir de uma matriz de jurisdição constitucional orientada pelo constitucionalismo digital. Nesse contexto, cabe ao Judiciário interpretar a legislação e estabelecer preceitos que assegurem a concretização dos direitos fundamentais, evitando que a opacidade dos sistemas algorítmicos acabe por perpetuar práticas discriminatórias e dificultar sua identificação e controle.

Ainda que se reconheça a necessidade de que tais sistemas atuem com maior transparência e que haja a devida responsabilização pelas consequências de seu uso, a presente investigação não se ocupa diretamente desses aspectos técnicos ou regulatórios. É necessário compreender o fenômeno da discriminação algorítmica sob a ótica do direito constitucional, com especial atenção aos deveres estatais de proteção e à eficácia dos direitos fundamentais diante das transformações tecnológicas. Trata-se de examinar o modo como o Estado deve tutelar seus jurisdicionados frente às aplicações algorítmicas que impactam suas liberdades, igualdade e dignidade.

A proteção contra a discriminação algorítmica não se esgota na esfera defensiva dos direitos fundamentais. Ela decorre também da sua dimensão objetiva – isto é, do dever estatal de estruturar a ordem jurídica de modo a prevenir lesões e criar condições para o gozo efetivo de liberdades, igualdade e dignidade. Nessa chave, o Estado não apenas se abstém de violar: ele deve proteger ativamente as pessoas contra práticas de terceiros – públicos ou privados – que, por meio de sistemas automatizados, produzam exclusões, assimetrias e estigmas (Leal; Maas, 2022, p. 423).

A tutela contra a discriminação algorítmica decorre da dimensão objetiva dos direitos fundamentais – isto é, do encargo estatal de organizar, prevenir e proteger. Nesse quadro, o dever de proteção se desdobra em: (i) dever de proibição (*Verbotspflicht*) – vedar condutas e usos algorítmicos manifestamente lesivos; (ii) dever de segurança/organização (*Sicherheitspflicht*) – instituir estruturas, procedimentos e capacidades estatais para governar sistemas de alto impacto; e (iii) dever de proteção contra riscos (*Risikopflicht*) – atuar *ex ante* quando a inovação introduz incerteza relevante sobre direitos. Esses feixes dialogam com os postulados de proteção: proibir práticas lesivas, legislar e regulamentar tutela a direitos e ter uma matriz interpretativa que seja capaz de avaliar o papel e a efetividade do direito fundamental à transparência algorítmica no controle aos algoritmos discriminatórios (Leal; Maas, 2022, p. 419).

Nesse cenário, a discussão sobre a necessidade de regulação e vigilância institucional torna-se ainda mais urgente. A ausência de mecanismos eficazes de controle sobre o funcionamento dos algoritmos pode consolidar práticas discriminatórias e aprofundar desigualdades já existentes. A discriminação algorítmica, por sua natureza invisível e silenciosa, desafia os instrumentos tradicionais de proteção dos direitos fundamentais, exigindo novas estratégias que articulem transparência, participação democrática e responsabilização das entidades envolvidas. O fortalecimento de mecanismos de governança, auditoria e explicabilidade das decisões automatizadas torna-se, assim, um pilar fundamental para a construção de uma sociedade digital verdadeiramente inclusiva, na qual o avanço tecnológico não seja confundido com retrocesso no campo dos direitos.

Esse dever de proteção vincula todo o aparelho de Estado – Executivo, Legislativo e Judiciário, além de entidades reguladoras e de controle. O Legislativo

deve desenhar salvaguardas proporcionais ao risco; o Executivo, implementá-las em políticas públicas, compras e contratações tecnológicas; o Judiciário, por sua vez, deve garantir que decisões automatizadas não se sobreponham aos direitos e que haja vias reais de explicação, contestação e reparação. A tutela não é episódica – é estrutural e contínua.

É nesse horizonte que se evidencia a urgência de alinhar inovação tecnológica aos valores constitucionais, exigindo do Estado e da sociedade civil uma postura ativa e vigilante. Estímulo à pesquisa em inteligência artificial explicável, promoção de uma educação digital crítica e aprimoramento da legislação protetiva são passos imprescindíveis para o enfrentamento dos desafios impostos pela opacidade algorítmica. Ao promover práticas pautadas na transparência e na equidade, torna-se possível resgatar o protagonismo do indivíduo diante de sistemas que frequentemente o reduzem a dados estatísticos ou perfis de consumo. Trazer o usuário para o centro do processo decisório – reconhecendo-o como sujeito de direitos, participativo e autodeterminável – constitui exigência inerente a um Estado Democrático de Direito conforme os parâmetros do constitucionalismo contemporâneo.

Quando as inovações tecnológicas introduzem incertezas relevantes sobre impactos em direitos, emerge o dever de proteção diante de riscos (*Risikopflicht*). Em termos práticos, isso impõe medidas *ex ante*: avaliações de impacto antes da adoção de modelos, limites de finalidade, supervisão humana significativa em decisões que afetem pessoas, revisão periódica com foco em grupos vulnerabilizados e outras medidas mitigadoras de risco que se afigurarem úteis. Onde a opacidade e o potencial discriminatório aumentam, cresce também o ônus argumentativo do Estado e dos operadores privados para demonstrar segurança, necessidade e proporcionalidade do sistema. A transparência é o jogar de luzes, antídoto do cenário opaco (Leal; Maas, 2022, p. 420).

A dimensão *ex ante* do dever de proteção – tal como aqui reconstruída – pode ser lida como um regime de pré-compromisso institucional, no sentido de que o Direito, ciente do risco de “fraqueza” decisória sob pressão de eficiência, complexidade e assimetria informacional, decide amarrar o futuro por regras públicas de prudência, como avaliações de impacto, deveres de documentação e revisão humana significativa. A partir da Teoria da Racionalidade de Elster (1979), transportando-a ao Direito, trata-se de reconhecer que, quando o risco cresce, não basta confiar na virtude

do agente (público ou privado), é preciso desenhar amarras verificáveis que impeçam que a tentação da performance substitua a justificação, convertendo a explicabilidade em condição de legitimidade e não em favor eventual. A força desse arranjo está em deslocar o centro do controle da promessa abstrata de conformidade para a estrutura concreta de garantias e, com isso, tornar a proteção não episódica, mas contínua, cumulativa e institucionalmente rastreável, especialmente quando a incidência recai sobre grupos em situação de vulnerabilidade.

A tecnologia em rede, em razão de sua sistematicidade e integração às rotinas sociais, está profundamente imbricada no modo de vida atual. O uso cada vez mais naturalizado da internet, em que o usuário interage com plataformas digitais sem plena consciência de que seus dados e comportamentos estão sendo constantemente coletados, analisados e convertidos em mercadoria, leva a uma forma de relação comodificada com a tecnologia – na qual a própria experiência de uso se torna objeto de exploração econômica, tendo a vida um aspecto de *commodity*. Nesse sentido, a utilização de dados para fins colaborativos, como nos casos do Google Maps e do aplicativo de mobilidade Waze, ilustra a ambivalência dessas tecnologias. Por meio da coleta contínua de informações oriundas dos dispositivos móveis dos usuários, tais plataformas geram benefícios à coletividade, otimizando rotas e promovendo maior eficiência no tráfego urbano. No entanto, essa mesma lógica de captação e processamento de dados, quando operada sem garantias adequadas de controle e transparência, pode facilmente ser instrumentalizada de modo a violar direitos fundamentais. É precisamente essa ambivalência que exige do Estado uma postura normativa e protetiva diante dos desafios trazidos pela era digital (Han, 2018, p. 99).

Essa simbiose é marca da sociedade da informação, uma sociedade que tem seu principal alicerce em algo intangível, a interconectividade informacional. Uma sociedade insaciável por informação, agilidade, instantaneidade e comodismo ao alcance de um *click*. É o universo do dataísmo – isto é, a crença compartilhada de que mais dados e mais interações entre esses dados apontam a uma resposta adequada que encontra seu fundamento e se legitima pela fecundidade de dados –, uma realidade bioquímico-digital que oferece poderes inéditos e imensos a todas as classes de homem (Harari, 2016, p. 370-371). Além disso, nesse intenso fluxo informacional, encontra-se o ser usuário da *web*, que se põe no meio de um turbilhão inconstante de mudanças paradigmáticas recorrentes, sem base jurisdicional unitária,

sem barreiras para a aceleração ou alterações estruturais. O paradigma, para Castells (2011, p. 87), é o da ação sobre a informação, centralizando a experiência humana na valorização do fluxo informacional, colocando o homem no centro da criação e do direcionamento de algoritmos que sabem mais de um indivíduo do que este sobre si próprio (Harari, 2016, p. 394).

A magnitude da importância da informação, sobretudo a das informações sobre pessoas, dos dados pessoais portanto, alavanca os dados a papel de destaque dentro da nova economia. Os dados são alçados ao novo petróleo (*data is the new oil*), com a vantagem de ser infindável, os veículos, contudo, são a própria economia fomentada pelas Big Techs. As estradas, no entanto, são os consumidores (fornecedores de dados, a matéria-prima). O paradigma da valorização de um bem intangível não é passível de assimilação tão-somente pela velha concepção mercadológica. O ganho associado ao conhecimento e direcionamento possível com os dados só é possível na nova concepção de economia digital de rede.

A Internet consegue ainda mais. Com a emergência da possibilidade de utilização dos dados para perfilização (formando perfis para análise comportamental, maior segurança de crédito e etc.), análise behaviorista de mercado (analisar o comportamento dos usuários-consumidores na Internet, para tentar manipular suas preferências a partir disso), e fluidez quanto a assertividade de ofertas de marketing (*micromarketing* e *microtargeting*) direcionadas para pessoas com o perfil de compra de um determinado produto, conseguiu promover um verdadeiro giro copernicano: transformou o consumidor-usuário em produto, e o fornecedor de produtos em consumidor de dados, o vulnerável da relação passa a ser o detentor da riqueza buscada na relação comercial. O usuário se torna fonte única e primária de insumo de dados.

Insta analisar o estado arte do usuário na internet, sendo ameaçado enquanto produto por uma realidade de marketing direcionado e ostensivo, que se vale de digital *fingerprints*, histórico de navegações e dados mais sensíveis, para direcionar ofertas de produtos e serviços através de complexos algoritmos, inclusive, discriminando pessoas para que produtos ou serviços não sejam disponibilizados, no que se cunhou os termos *geoblocking, geopricing* e discriminação algorítmica (viés). Calha refletir, nesse sentido, se a Lei Geral de Proteção de Dados, pensada para essa tutela, consegue desempenhar uma proteção material frente à nova realidade que se

apresenta, e que em momento posterior à formatação final do presente já estará mais avançada. A ideia de resguardo dos dados, no âmbito digital, consubstancia-se na proteção das informações que podem ser coletadas, armazenadas, usadas e propagadas comercialmente no mundo virtual. Assim, insta analisar o estado arte desta nova relação, em que o ser é ameaçado enquanto um mero produto de uma ostensiva e direcionada realidade algorítmica.

O fascinante mundo dos dados, e de seu potencial transformador da realidade humana, ganha especial notoriedade quando os estudos de defesa do usuário e de proteção de dados pessoais se relacionam. Segundo os estudos de *behavioral economics* (economia comportamental), que se estabelece a partir de análises das capacidades cognitivas pessoais quando da tomada de decisão sobre amplos aspectos da vida, há questões psicológicas, bioquímicas e econômicas relacionadas quanto o comportamento humano, sobretudo com foco no *homo economicus*, isto é, o homem inserido em uma sociedade mercadológica (Daura, 2018, p. 569-570).

Um case famoso de análise comportamental é a da empresa Target. A empresa queria atingir o público dos novos pais e, para isso, começou a cruzar milhões de dados de padrão de consumo, via Big Data, que pudessem identificar possíveis grávidas, a partir da compra de produtos não diretamente relacionados à gravidez, mas frequentemente comprado por grávidas. O caso virou famoso quando uma adolescente recebeu anúncios de produtos de bebê, achou estranho, e descobriu que estava de fato grávida, isto é, o estudo da Target conseguiu descobrir que ela estava grávida mesmo sem ela ou seus pais terem se dado conta. Esse é o poder da análise comportamental e o paradigma econômico-tecnológico atual.

A proteção de dados na realidade algorítmica é um tema cada vez mais relevante e desafiador para a sociedade contemporânea. Com o avanço das tecnologias algorítmicas, a coleta e o processamento de dados pessoais se tornaram cada vez mais complexos e difíceis de serem controlados. A realidade algorítmica impõe novos desafios à proteção dos dados pessoais dos usuários, especialmente no que diz respeito à discriminação injustificada e à transparência dos algoritmos. A LGPD estabelece diretrizes importantes para a proteção dos dados pessoais na realidade algorítmica, incluindo a vedação à discriminação injustificada e o direito à transparência e explicabilidade dos algoritmos. No entanto, ainda há muito a ser feito para garantir uma proteção efetiva dos dados pessoais na era do big data. Além disso,

é importante que os usuários tenham acesso às informações sobre como seus dados estão sendo coletados e usados pelos algoritmos.

Nesse quadro, opacidade e neutralidade formal não se conciliam com igualdade substancial. A proteção de dados continua necessária, mas é insuficiente sem rastreabilidade, auditabilidade, explicabilidade e supervisão humana nas decisões que impactam direitos. O dever de proteção – dever de proibição (*Verbotspflicht*), dever de segurança (*Sicherheitspflicht*) e dever de evitar riscos (*Risikopflicht*) – exige uma moldura constitucional efetiva para transformar recomendações técnicas preocupadas com a efetivação dos direitos em obrigações jurídicas. É com essa chave que, adiante, densificam-se critérios operacionais para a jurisdição constitucional em face da discriminação algorítmica (Leal; Maas, 2022, p. 419).

Um dos principais desafios enfrentados na proteção dos dados pessoais na realidade algorítmica é a discriminação injustificada. Os algoritmos podem perpetuar preconceitos e estereótipos ao tomar decisões com base em dados pessoais, como raça, gênero e orientação sexual. Isso pode levar a uma discriminação injusta e prejudicar a vida das pessoas afetadas. Outro desafio importante é a falta de transparência dos algoritmos. Muitas vezes, os usuários não têm acesso às informações sobre como seus dados estão sendo coletados e usados pelos algoritmos. Isso pode levar a uma falta de confiança nos serviços que utilizam esses algoritmos e prejudicar a privacidade dos usuários (Paranhos, 2022).

A discriminação algorítmica é uma ameaça cada vez mais preocupante para a proteção dos dados pessoais na sociedade contemporânea. À medida que as tecnologias algorítmicas avançam, torna-se mais difícil controlar a coleta e o processamento de informações pessoais, o que pode levar a uma discriminação injustificada e prejudicar vidas e dignidades. É fundamental que a proteção dos dados pessoais seja fortalecida para impedir que os algoritmos perpetuem preconceitos e estereótipos com base em informações como raça, gênero e orientação sexual. Para garantir uma proteção efetiva dos dados pessoais na era do *big data*, é necessário que as empresas estabeleçam mecanismos eficazes de controle e fiscalização da coleta e do uso dessas informações. Os usuários também devem ter acesso a informações claras sobre como seus dados estão sendo coletados e usados pelos algoritmos, a fim de garantir a transparência e a responsabilidade das empresas.

Ações efetivas devem ser tomadas para garantir que as empresas cumpram suas obrigações legais e éticas de proteção dos dados pessoais dos usuários, sem discriminação ou violação de seus direitos fundamentais.

Para enfrentar esses desafios, é necessário desenvolver novas tecnologias que garantam a transparência e explicabilidade dos algoritmos, como a XAI. Isso pode incluir o uso de técnicas de explicabilidade, que permitem aos usuários entender como os algoritmos tomam decisões com base em seus dados pessoais. Além disso, é importante fortalecer as leis e regulamentações existentes para garantir uma proteção efetiva dos dados pessoais na realidade algorítmica. A LGPD é um exemplo importante de legislação que estabelece diretrizes claras para a proteção dos dados pessoais no Brasil. No entanto, ainda há muito a ser feito para garantir que essas diretrizes sejam efetivamente implementadas pelas empresas e fiscalizadas pelos órgãos reguladores.

Portanto, é imperioso analisar a relação dos dados no universo da defesa do usuário, que é vulnerável conceitualmente e pela própria essência de uma política de proteção. Alguns casos são extremamente pertinentes de serem abordados para desenhar o estado da arte e o risco de vazamento de dados. Um caso de vazamento de dados foi protagonizado pelo Serasa. O Procon-SP notificou a Serasa Experian pedindo explicações sobre notícias acerca de um suposto vazamento de 220 milhões de dados pessoais de cidadãos brasileiros. Após explicações, o Procon-SP entendeu que a empresa não teria conseguido implementar medidas para o cumprimento do CDC e da LGPD. O entendimento veio apoiado no fato de que a empresa não detalhou a finalidade e base legal para o tratamento de dados pessoais, a necessidade do consentimento para a coleta de dados, medidas para o atendimento da LGPD, e a segurança desses dados, tampouco uma política de armazenamento e descarte dos dados coletados. Esse caso de São Paulo aguarda desfecho, contudo a Serasa S.A. também foi ré em Brasília.

Nos autos da Ação Civil Pública nº 0736634-81.2020.8.07.0001, o Ministério Público do Distrito Federal e dos Territórios identificou, através da Unidade Especial de Proteção de Dados e Inteligência Artificial, a indevida e maciça comercialização de dados pessoais de brasileiros por meio de serviços como “Lista Online” e “Prospecção de Clientes”, oferecidos pela Serasa Experian. O TJDFT julgou procedente a ACP, no sentido de condenar a ré Serasa S.A. a se abster de comercializar dados pessoais

dos titulares por meio dos produtos denominados “Lista Online” e “Prospecção de Clientes”, sob pena de imposição das medidas indutivas, coercitivas etc. No dispêndio argumentativo, os desembargadores entenderam que a comercialização das listas viola o art. 5º, X, da Constituição Federal, bem como os correspondentes incisos I e IV do art. 2 da LGPD, todos os dispositivos tratando sobre a inviolabilidade da intimidade, da honra e da imagem dos titulares dos dados. Mas há considerações do caso a serem consideradas mais aprofundadamente.

Os dados comercializados não eram dados sensíveis, porquanto não tratavam de dados sobre origem racial ou étnica, convicção religiosa, opinião política, filiação a sindicato ou a organização de caráter religioso, filosófico ou político, dado referente à saúde ou à vida sexual, dado genético ou biométrico, como conceitua o art. 5º, II, da LGPD. Para o tratamento desses, o art. 11, I, dispõe que o tratamento somente é cabível com o consentimento do titular ou responsável, manifestado de forma específica e destacada, ressalvadas hipóteses em que os dados forem indispensáveis para questões legais (art. 11, II, da LGPD). No entanto, os dados comercializados não eram sensíveis. Ainda assim, mereciam tutela. O Serasa argumentou que por não serem dados sensíveis, não haveria ilegalidade na prática, visto que haveria dispensa do consentimento do titular, uma vez que o controlador teria interesse legítimo (art. 7º, IX, da LGPD) ou que o compartilhamento dos dados tenha finalidade de proteção do crédito (art. 7º, X, da LGPD). E de fato, há interesse legítimo e atende as finalidades da Serasa, porquanto no seu estatuto social estão delineados os objetos sociais de coleta, o armazenamento e o gerenciamento de dados, organização, análise, desenvolvimento, operação e comercialização de informações e soluções para apoiar decisões, o gerenciamento de risco de crédito e de negócios, a administração de finanças pessoais e para promover educação financeira etc.

Ainda assim, o TJDFT entendeu que o art. 7º da LGPD dispõe que o consentimento do titular é a regra maior a ser observada para o tratamento de dados pessoais, tanto é que o § 4º, daquele dispositivo, prescreve a dispensa do consentimento apenas para os dados tornados manifestamente públicos pelo titular. De modo que mesmo para os dados não sensíveis, o controlador deve obter o consentimento, exceto quando os dados foram tornados manifestamente públicos pelo próprio titular. Além disso, o cerne da LGPD é a tutela à autodeterminação do indivíduo em relação à veiculação de suas informações, de modo que apenas pelo

consentimento inequívoco que o titular dos dados consegue controlar o nível de proteção e os fluxos de seus dados, permitindo ou não que suas informações sejam processadas, utilizadas, repassadas e assim por diante. É direito dos titulares dos dados, portanto, dar anuência. E a maior preocupação com o consentimento valoriza a pessoa no mundo digitalizado (Teffé; Viola, 2020, p. 6-7).

Sobre comercialização de dados, o Ministério Público do Distrito Federal e dos Territórios também litigou com o Mercado Livre, que oferecia banco de dados e cadastro em geral pela plataforma. O Vendedor era EMARKETING011ERICAVIRTUAL, que vendia bancos de dados digitais por R$ 500,00. O Mercado Livre, nos autos da Ação nº 0733785-39.2020.8.07.0001, da 17ª Vara Cível de Brasília, teve que suspender o anúncio, em liminar, e em sentença ficou condicionada a multa de R$ 2 mil por operação desse tipo realizada novamente pelo Mercado Livre. No caso não parecia haver anuência dos titulares de dados com a comercialização do produto, portanto essa era ilegal.

Em uma cadeia de e-commerce os dados podem transitar por muitas empresas, sobretudo para a viabilização de meios de pagamento. Com efeito, entre a loja virtual e o usuário há empresas de gateway, antifraudes, adquirentes, subadquirentes, bandeiras de cartão de crédito, emissores de cartões, bancos e intermediários. Assim, há muito fluxo informacional envolvido, o que redobra o grau de seriedade com a tutela dos dados pessoais, tratamento e demais direitos estabelecidos pelo microssistema protetivo, foco especial no CDC, e na LGPD. Uma palavra é essencial para toda a cadeia comercial: confiança.

A relação comercial que ocorre na Internet precisa ter a mesma confiança, segurança e informação que a clássica, isto é, deve atender aos mesmos requisitos e ainda a outros mais específicos. O complexo normativo protetivo dos dados tem, também, esse desafio, o de transformar o ambiente virtual em um ambiente seguro, confiável e tutelado, que não promova a discriminação ou o predatismo potencializado por esses dados ilicitamente comercializados.

O Código de Defesa do Consumidor encampa essa proteção desde 1990, sobretudo com o art. 43, que estabelece uma série de direitos e garantias para o usuário em relação às suas informações pessoais presentes em bancos de dados e cadastros. Com base tão somente nesse dispositivo destacam-se os direitos de acesso aos dados, de retificação dos dados e de cancelamento deles. Além disso, o

§1º determina que os cadastros e dados de usuários devem ser objetivos, claros, verdadeiros e em linguagem de fácil compreensão, não podendo conter informações negativas referentes a período superior a cinco anos, atendendo aos princípios trazidos na LGPD como qualidade dos dados, livre acesso e transparência. Além disso, o limite legal de 5 anos beneficia muito os usuários. Nesse sentido, Bioni (2019, p. 126), destaca que o CDC já trazia a preocupação com a autodeterminação informacional.

Uma prática abusiva, nesse sentido de proteção ao crédito, é a da perfilização (*profiling*), que consiste na elaboração de perfis de comportamento de uma pessoa (ou de um grupo de pessoas) a partir de dados pessoais, disponibilizados por ela ou colhidos. Assim, através de algoritmos aliados à ciência da economia comportamental, é possível prever condições, decisões ou comportamentos futuros de uma pessoa. Isso certamente beneficia as empresas que trabalham com risco, ou que querem evitar o risco demasiado, com a técnica do *Credit Scoring* (RESP nº 1.419.697/RS). Outro uso é para o direcionamento do marketing ostensivo e pessoalizado, esse uso até pode ocorrer, mas deve ser feito mediante consentimento adequado do usuário.

A Lei nº 14.181/2021 inseriu o art. 54-D no CDC, determinando, no inciso II, que o fornecedor ou intermediário da oferta de crédito avalie, de forma responsável as condições de crédito do usuário, mediante análise das informações disponíveis em bancos de dados de proteção ao crédito, observado o disposto no CDC e nas demais leis do sistema normativo de proteção de dados.

Sobre o consentimento, que guarda relação com a autodeterminação, tem-se que a formação de bancos de dados de usuários, pela incidência em comum da LGPD e do CDC exige o consentimento expresso do usuário-titular. Esse deve ser dado a partir de uma manifestação livre de vontade, voltado a uma finalidade específica e que tenha sido informado sobre o processamento e utilização dos dados, bem como da possibilidade de não consentir e ainda assim utilizar a plataforma ou ferramenta. Há, contudo, uma grave crise de consentimento de dados. E isso pode ser facilmente explicado pelo fato de que as pessoas não leem as políticas de privacidade, e, mesmo que se lessem, talvez não as entendessem completamente, por lhes faltar conhecimento técnico. De forma que o consentimento informado pode ser uma doce ilusão (Schermer; Custers; Hof, 2014, *passim*). Mendes (2014, p. 23) comenta que há uma situação de extrema vulnerabilidade para os usuários-consumidores, que

enfrentam dificuldades para sequer controlar seus dados disponibilizados por *clicks* em *pop-ups* que surgem no meio da navegação na rede. Uma marca da modernidade é a aceleração da ansiedade, da ânsia pelo serviço, pelo bem, pela informação, e os *pop-ups* de consentimento estão no caminho e a um *click* em “aceitar” de distância.

Essas coletas podem ser chamadas de quase não consentidas. Mas há um caso interessante sobre a coleta não consentida de dados, mediante a utilização de scanners corporais pela empresa ViaQuatro de SP. O caso desenrolou a partir de um scanner de reconhecimento facial implantado para coletar registros biométricos de passageiros, e a ideia era na utilização das telas, que eram capazes de contar quantas pessoas passaram em sua frente e de analisar as emoções a partir da expressão facial, para a finalidade de passar anúncios publicitários em uma modalidade chamada *pay-per-face*, isto é, pague de acordo com a quantidade de pessoas que olharam para a tela. Na justiça de São Paulo, nos autos da Ação Civil Pública nº 1090663-42.2018.8.26.0100, movida pelo IDEC, a empresa foi condenada a pagar R$ 100 mil de multa, e se abster de utilizar a tecnologia, porquanto os usuários sequer foram advertidos da utilização do totem, e muito menos foi solicitado qualquer consentimento para os usuários dos serviços de metrô da ViaQuatro. Após o caso, a empresa se comprometeu a reforçar a transparência e se adequar à LGPD.

Contudo, o caso tem mais uma agravante. Dentre a coletividade dos usuários da ViaQuatro, poderiam, obviamente, existir crianças e adolescentes. O Código de Defesa do Consumidor destaca, em seu art. 39, IV, que é uma prática abusiva prevalecer-se da fraqueza ou ignorância do usuário, tendo em vista sua idade, saúde, conhecimento ou condição social, para impingir-lhe seus produtos ou serviços. Além disso, a LGPD, em seu art. 14, estabelece uma proteção especial às crianças e adolescentes, trazendo que o tratamento de dados pessoais de crianças e de adolescentes deverá ser realizado em seu melhor interesse. No mesmo sentido, o ECA, em seu art. 17, garante o direito ao respeito à imagem de crianças e adolescentes, preservando-as. No caso da ViaQuatro, a juíza da 37ª Vara Cível entendeu que os dados coletados eram sensíveis, além de que a violação aos direitos básicos do usuário é amplificada em termos de seus dados, sobretudo na moderna sociedade da informação. O resguardo do usuário enquanto um produto, e maior ainda quando o usuário que pode ser objetificado em seus dados é uma criança ou

um adolescente. Há vulnerabilidade dos usuários, e hipervulnerabilidade agravada quando esses são crianças ou adolescentes.

Os casos nacionais mencionados revelam uma sensibilidade protetiva, mas ainda incipiente quanto à lógica de deveres estruturais. As decisões sobre comercialização massiva de dados e reconhecimento facial em transporte público reforçam a tutela da intimidade, do consentimento e da proteção de crianças e adolescentes – avanço importante. Contudo, raramente se formula explicitamente o problema como “dever de proteção contra riscos algorítmicos de discriminação” – também por conta da época em que se deram as discussões jurídicas, em que efervescia a temática da proteção de dados, tão somente – e quase nunca se estabelecem critérios estáveis de governança algorítmica. O passo que se propõe é levar à análise tópica de novos casos uma matriz de proteção com matrizes interpretativas que exigirão, de uma maneira (critério econômico) ou outra (imposição jurisdicional) o controle contínuo, preventivo e à altura da natureza evolutiva dos sistemas algoritmizados.

À luz desse dever de proteção, a leitura dos casos revela avanços importantes, mas ainda incompletos frente à ameaça algorítmica. No caso Serasa/TJDFT, há tutela da intimidade e da autodeterminação informacional, porém o problema não é formulado como risco de discriminação algorítmica e não se exigem avaliações de impacto algorítmico, o caso ficou restrito ao consentimento. No caso do Mercado Livre, a cessação da comercialização massiva de dados foi correta, mas na decisão poderia ter havido uma imposição de deveres contínuos de monitoramento, auditabilidade e alertas de risco, dando concretude ao dever de proteção. Já no caso da ViaQuatro, o reconhecimento facial sem consentimento e sem justificativa proporcional foi coibido com foco na proteção de crianças e adolescentes; ainda assim, seria adequado vedar a reintrodução da prática enquanto não demonstradas necessidade e proporcionalidade, explicabilidade dirigida ao afetado e mecanismos de contestação/supervisão humana, materializando o dever de proibição (*Verbotspflicht*) e o dever de proteção contra riscos (*Risikopflicht*). Em síntese, as decisões caminham, mas ainda não incorporam de modo explícito e estruturado a matriz de deveres estatais exigível diante do risco algorítmico, sobretudo a partir de uma necessária tônica antidiscriminatória.

Os dados são os maiores ativos de certos ramos empresariais e esse só fato acarreta riscos à privacidade dos consumidores. A partir da capacidade de contenção e operacionalização do turbilhão informativo, através do *Big Data*, torna-se perigosa a questão dos consentimentos. A partir de simples “consentimentos” por parte dos titulares de dados pessoais é possível criar um banco de dados com informações pessoais desses e, até mesmo, traçar perfis de consumo com a seleção de ofertas que, possivelmente, irão interessar ao titular. No que diz respeito aos dados coletados de crianças e adolescentes, frequentadores assíduos da Internet, o cuidado deve ser maior. A solução encontrada pela LGPD foi, pela inteligência do art. 14, §1º, o de condicionar o consentimento do infante, por substituição, ao de pelo menos um dos pais ou pelo responsável legal. Das duas uma, portanto: ou há um poder parental ativo e presente, cuidando da navegação da web dos filhos, que são muitas vezes mais hábeis em rede do que os pais ou responsáveis, ou deverá ser reformulada a política de proteção de dados para crianças e adolescentes, fortalecendo-a.

Tudo isso para que se garanta a tutela dos dados dos usuários. Se não pelo temor às altas sanções, pelo caráter pedagógico que adota a legislação protetiva. Usuários são pessoas, dotados de direitos fundamentais, e não podem ser coisificados ou precificados por seu valor de mercado inerente aos dados pessoais que detêm. Não são barris de petróleo, mas sujeitos vulneráveis de direitos.

A questão é: isso (a proteção de dados) é suficiente?

Proteger os dados parece ser apenas uma parte mínima do problema. Quando os dados são utilizados e a discriminação já ocorreu, não se questiona mais apenas a origem dos dados, como eles foram parar ali; transcende-se a questão para o atingimento da dignidade da pessoa ofendida, não apenas pela violação de seus dados – se for esse o caso –, mas por violação de outros direitos, violação essa de cunho discriminatório de outros direitos, cuja proteção é demandada pelo Estado.

Diante desse quadro, evidencia-se que o Estado não pode se esquivar da responsabilidade de atuar de forma ativa e vigilante para evitar que mecanismos de discriminação, agora potencializados e sofisticados pelo emprego de algoritmos, se instaurarem ou se perpetuarem no tecido social. O papel estatal transcende a mera regulamentação formal, devendo envolver-se tanto na criação de diretrizes regulatórias e jurisprudenciais claras quanto na fiscalização rigorosa destes sistemas,

com o objetivo de garantir que práticas discriminatórias, estruturais ou resultantes da automação, sejam combatidas de maneira efetiva.

É imprescindível reconhecer que a inércia estatal diante das novas tecnologias, especialmente daquelas capazes de multiplicar desigualdades ou cristalizar preconceitos históricos sob a aparência de neutralidade algorítmica, representa uma violação do princípio constitucional da dignidade da pessoa humana e do direito fundamental à não discriminação. Cabe ao Estado, portanto, agir não apenas de modo reativo, após a ocorrência do dano, mas sobretudo preventivo, antecipando-se a eventuais riscos e estabelecendo mecanismos de controle, auditoria e transparência nos sistemas algorítmicos utilizados tanto pela administração pública quanto por agentes privados.

Tal dever de proteção estatal, conhecido no direito germânico como *Schutzpflicht*, implica a obrigação do Estado de tomar todas as medidas necessárias para proteger os indivíduos contra violações de direitos fundamentais por si, por terceiros (*Drittwirkung*) e em uma vinculação horizontal (eficácia horizontal) dos direitos fundamentais (*Horizontalwirkung*). Como bem pontuam Leal e Maas (2020, p. 49 e 50), trata-se de um compromisso constitucional que vincula o Estado à proteção efetiva dos direitos fundamentais, *in casu*, da igualdade, da autodeterminação e da não-discriminação. Ao reconhecer que a luta contra a discriminação algorítmica exige um contínuo processo de atualização normativa e interpretativa, o Estado reafirma sua função de guardião dos direitos fundamentais, especialmente no contexto de constantes avanços tecnológicos que desafiam as bases tradicionais da justiça e da igualdade. Dessa forma, a proteção estatal torna-se elemento central para assegurar que os benefícios da tecnologia sejam partilhados de maneira equitativa, sem tolerância a retrocessos sociais ou perpetuação de injustiças históricas sob novas roupagens digitais. Em relações entre particulares marcadas por assimetrias técnicas e informacionais, a eficácia horizontal dos direitos fundamentais reclama deveres procedimentais positivos.

Sob a perspectiva constitucional, vivencia-se um novo cenário em razão das transformações promovidas pela tecnologia digital, que introduz conflitos inéditos capazes de afetar o equilíbrio do ecossistema constitucional. Em resposta a esses desafios, são implementadas diversas medidas normativas com o objetivo de restabelecer esse equilíbrio em uma nova roupagem, a do constitucionalismo digital

(Celeste, 2019, p. 3). Gavião Filho (2023, p. 41) afirma que a descrição de um equilíbrio constitucional do constitucionalismo moderno é caracterizada pela existência de um “catálogo de direitos fundamentais e um órgão investido de jurisdição para controlar as decisões dos poderes estatais que limitam a esfera de atuação das pessoas titulares desses direitos fundamentais”. A ideia desse catálogo é a proteção dos direitos fundamentais, positivados em uma Constituição, frente a intervenções estatais que podem ser originárias da atividade do mercado, ou mesmo de leis, atos normativos, ou decisões judiciais.

Os princípios fundamentais bem como os objetivos gerais do constitucionalismo contemporâneo fazem parte também do constitucionalismo digital, concentrados num contexto peculiar a partir do surgimento da tecnologia digital. O equilíbrio constitucional, como a condição ideal produzida pela aplicação das normas de direito constitucional em uma determinada ordem jurídica, reflete duas funções básicas do direito constitucional, a proteção dos direitos fundamentais e o equilíbrio de poderes (Celeste, 2019, p. 3-4).

Nesse sentido, os direitos fundamentais recebem proteção constitucional face às intervenções do Estado. Contudo, na era da tecnologia, os conflitos ultrapassam as relações entre os titulares dos direitos fundamentais e o Estado, provocando alterações no sistema constitucional, que devem ser atendidas dentro de um renovado panorama de como os direitos fundamentais interagem nesse novo mundo de um constitucionalismo digital, e a atenção deverá ser cada vez mais contemporânea acerca de como efetivar tais proteções em contexto algorítmico (Paulo, 2024).

O Constitucionalismo Digital é um segmento do constitucionalismo contemporâneo, que utiliza de ferramentas clássicas, com um novo olhar das contrações (ou respostas) constitucionais contra os desafios trazidos pela tecnologia digital. A *Horizontalwirkung* entre o usuário de certa aplicação algorítmica e a sua desenvolvedora deverá ser reimaginada. Aqui uma relação horizontal não parece ser equitativo entre as potencialidades de respeito e de capacidades técnicas envolvidas pelas partes. Novos contextos fazem emergir novas perguntas, que traçam a busca por novas respostas.

Mapeada a ambivalência dos sistemas algorítmicos e os mecanismos de discriminação que produzem, delineia-se um núcleo normativo inequívoco: opacidade e neutralidade formal são incompatíveis com igualdade substancial. A proteção de

dados é indispensável, mas insuficiente se não vier acompanhada de rastreabilidade, auditabilidade, explicabilidade e supervisão humana nas decisões que impactam direitos. Nessa chave, o dever estatal de proteção (*Schutzpflicht*) – em suas vertentes proibitivas, organizatório-procedimentais e de gestão de riscos – fornece a moldura constitucional que converte recomendações técnicas em obrigações jurídicas para Estado e privados, estruturando uma agenda de governança algorítmica com padrões proporcionais ao risco e com ônus argumentativo escalonado conforme a gravidade e a incerteza do sistema. No capítulo seguinte, esses vetores se densificam em critérios operacionais de revisão judicial, aptos a orientar a jurisdição constitucional diante da discriminação algorítmica.

4 TUTELA JURÍDICA DA DISCRIMINAÇÃO ALGORÍTMICA: EM BUSCA DE SENTIDO, SIGNIFICADO E PROTEÇÃO NORMATIVA

> *Alguém devia ter difamado Josef K., pois, sem que tivesse feito nada de mau, certa manhã foi detido (Kafka, 2023, p. 15).*

Josef K. desperta sob acusação. Não sabe por quê. Não sabe com base em que normas, nem onde estão os autos. Em *O Processo*, a injustiça não é apenas a sentença: é o modo como se decide – um labirinto sem portas visíveis, onde o réu jamais alcança as razões que o afetam. A defesa torna-se impossível porque o direito é sonegado sob a forma de opacidade. A parábola "Diante da Lei" condensa esse paradoxo: há uma porta, há um porteiro, há uma promessa de acesso – mas o sentido fica eternamente adiado.

O mundo algorítmico multiplica essa experiência. Decisões automáticas recusam crédito, filtram currículos, modulam preços, ranqueiam perfis; e, não raro, fazem-no por correlações espúrias, e ainda mais: opacas, baseadas em dados e *proxies* que se ocultam atrás do véu do segredo comercial, do compliance formal ou da assimetria informacional. O destinatário do ato – a "parte afetada" – encontra-se em situação kafkaniana: sente o peso do resultado, mas não alcança as razões que o produziram nem o caminho que o poderia desfazer. Sem motivos inteligíveis, não há contraditório; sem trilha auditável, não há prova; sem prova, o devido processo degrada-se em ritual vazio.

Este capítulo responde a esse risco de "processo sem autos" na era algorítmica. Partindo do diagnóstico de que a discriminação estrutural (Cap. 2) encontra no ambiente automatizado (Cap. 3) um multiplicador silencioso, passamos a converter a angústia da opacidade em critérios jurídicos operacionais. Em chave constitucional, a pergunta kafkaniana – "de que sou acusado?" – transforma-se em deveres positivos de explicabilidade, rastreamento, auditabilidade e transparência proporcional ao risco, sem os quais a proibição de discriminar e o devido processo legal se tornam irrealizáveis.

A irrupção dessa racionalidade de máquina não é recente. Já em 1948, Alan Turing indagava se máquinas poderiam demonstrar comportamento inteligente, lançando, com sua *Intelligent Machinery*, as bases de uma era em que a decisão

automatizada passaria a disputar o espaço da decisão humana. Ao projetar a possibilidade de simular processos mentais com base em cálculos e correlações, Turing (1948) não previu apenas avanços técnicos, mas inaugurou uma tensão filosófico-jurídica: até que ponto decisões tomadas por máquinas – isentas de vivência, dor e alteridade – podem ser legitimadas sem trair os pressupostos da justiça constitucional? O que antes era apenas especulação técnica tornou-se, na era algorítmica, desafio normativo urgente.

A inteligência artificial carece do atributo fundamental da vivência humana. Conforme destacado, ela não se assemelha a um ser pensante desprovido de sentimento ou a um instrumento meramente objetivo e registrador; ao contrário dos seres humanos, que geram suas ideias por meio de experiências profundas e emoções intensas, como dor, paixão, desejo e consciência. A inteligência artificial limita-se a ser uma ferramenta de cálculo. A diferença não está em apenas na metafísica, embora possua mecanismos de aprendizagem, inclusive aprendizado profundo, não é capaz de experienciar sentimentos. Apenas a vivência da dor transforma a inteligência em espírito, sendo este um processo inalcançável por qualquer algoritmo, ao menos nos limites atuais. Em termos nietzscheanos, somente a dor intensa e prolongada, que exige tempo, pode libertar o espírito, levando filósofos à sua essência mais profunda e promovendo o desapego de valores anteriormente atribuídos à humanidade, tais como confiança, bondade ou moderação (Han, 2021, p. 49).

Em um mundo onde as decisões são delegadas a sistemas que operam na sombra, a experiência de Josef K. torna-se emblemática. Se antes o indivíduo enfrentava a angústia de um processo invisível, hoje ele se depara com a mesma ausência de justificativas, mas agora mediada por algoritmos. Essa transposição da lógica kafkiana ao universo digital demonstra que o desafio não é apenas técnico, mas profundamente jurídico e humano. Arendt (1999, p. 15) adverte contra a banalidade do mal, isto é, lógicas de estigmatização, operadas sob a aparência da legalidade e da melhor técnica, que permitiam a produção sistemática da injustiça. Não se trata de estabelecer um paralelismo moral entre o horror do Holocausto e as práticas algorítmicas contemporâneas, mas de reconhecer que a ausência de responsabilidade, aliada à automatização de decisões discriminatórias, pode gerar,

em uma nova passagem pelo ponto no círculo histórico, mais uma forma de crueldade: fria, técnica, silenciosa e automatizada.

A partir do diagnóstico conceitual arquitetado no capítulo 2, sobretudo acerca da discriminação estrutural, e do mapeamento fenomenológico do capítulo 3, acerca da discriminação automatizada e replicativa daquela discriminação estrutural, o presente momento opera a transição do problema apresentado em critérios normativos operacionalizáveis pelo Direito.

O eixo central deste estudo compreende dois aspectos: **i)** a reconstrução do sentido e significado do sistema jurídico diante da discriminação algorítmica, especialmente sob a ótica da igualdade material e do dever estatal de proteção dos direitos fundamentais; e **ii)** a conversão desse entendimento em interpretações aplicáveis à atuação de autoridades públicas e privadas, visando à prevenção, detecção e remediação de impactos desproporcionais sobre grupos vulneráveis.

Complementarmente, a LGPD oferece a estrutura infraconstitucional pertinente, pautada por princípios como finalidade, adequação, necessidade, qualidade dos dados, transparência, não discriminação, responsabilização e prestação de contas (art. 6º), além do dever de prevenção (art. 6º, VIII) e da possibilidade de revisão de decisões automatizadas (art. 20[8]) – o que, conjugado com a exigência constitucional de motivação adequada, converte-se em dever positivo de explicabilidade proporcional ao risco e ao impacto. O direcionamento pragmático consiste na atribuição de ônus informacional e deveres de explicabilidade proporcionais ao risco e potencial de impacto às empresas que empregam sistemas de decisão automatizada, sobretudo em situações envolvendo grupos vulneráveis, sendo intensificada a exigência justificativa, inicialmente de caráter não jurídico, mas mercadológico, conforme previsto no CDC. É, no entanto, insuficiente trabalhar com a legislação, seja em nível nacional ou internacional.

Para efeito de calibração do escrutínio, considera-se vulnerabilidade a confluência de **i)** susceptibilidade histórica a discriminações estruturais; **ii)** dependência funcional do serviço/decisão automatizada para fruição de direitos (saúde, crédito, trabalho, benefícios); **iii)** assimetria informacional elevada frente ao

[8] O titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade.

operador; **iv)** formação de perfis *proxy* – o protocolo de acesso feito pela internet – com atributos sensíveis diretos ou inferidos, possibilitando a aferição indireta de características que indicam pessoa classificada em um grupo vulnerável, ainda que por correlações espúrias, como geolocalização. A presença concomitante de tais vetores aumenta o dever de explicabilidade e a densidade do controle judicial.

A junção das novas tecnologias com a já previamente existente discriminação estrutural da sociedade contemporânea culminou no fenômeno da discriminação algorítmica, que, como visto, provoca necessárias alterações no sistema de justiça constitucional, que devem ser atendidas dentro de um renovado olhar acerca das dimensões metodológicas de proteção dos direitos fundamentais perante tal nova realidade. É um novo campo de ocorrência da discriminação estrutural, como uma nova metodologia, a de ser automatizada, mas não é um novo inimigo, e sim uma impertinente renovação deste já combatido agente estruturalmente discriminatório.

A lei é insuficiente e, nesse cenário, entre a Cila da anomia e a Caríbdis da eterna mora regulatória, surge amarrada ao poste a jurisdição constitucional, o direito prático-argumentativo-problematizante.

Nesse contexto, o Direito é convocado a enfrentar não só a opacidade técnica, mas a reconstruir a narrativa de acesso à justiça. A alusão a Kafka não é figura de estilo: é a chave heurística que demonstra porque sem inteligibilidade não há devido processo, e porque a Constituição exige transformar opacidade técnica em deveres positivos associados à transparência. A ausência de explicações transforma-se em um obstáculo insuperável ao exercício do contraditório e da ampla defesa. A tarefa jurídica, portanto, é devolver a transparência ao processo decisório, garantindo que o indivíduo não seja apenas objeto de uma decisão, mas sujeito de direitos.

Se, em Kafka, o sujeito é consumido por um processo cujo sentido jamais se revela, na era algorítmica o drama se reinventa: o julgamento já não exige tribunal, nem juiz visível, nem mesmo acusação formal. A sentença é computada em silêncio, e sua origem desaparece entre linhas de código e cláusulas contratuais. Cabe, então, ao Direito – sobretudo à jurisdição constitucional – romper essa lógica do desaparecimento. Não se trata apenas de exigir explicações técnicas, mas de restaurar a inteligibilidade da norma, o acesso ao contraditório, a possibilidade de resistência. Neste capítulo, busca-se reconstruir, a partir das promessas constitucionais e das gramáticas de proteção dos dados, da igualdade e do direito

antidiscriminatório, um caminho de reintegração do sujeito ao centro do processo decisório: não mais objeto de classificação, mas titular de dignidade. Contra a opacidade kafkiana da máquina, propõe-se a reinvenção de um Direito que saiba escutar – e que, ao escutar, proteja.

4.1 Sentido normativo do direito acerca da tutela antidiscriminatória algorítmica

Se há algo que nos distingue das máquinas, dos robôs e das artificialidades, esse algo é a consciência. A consciência humana é o fator determinante de categorização entre uma inteligência do tipo humana e uma inteligência do tipo máquina. Constatou-se que tal consciência constitui um atributo fundamental das espécies racionais, permitindo-lhes posicionar-se com discernimento em relação a cenários amplos e complexos (Russell, 2021, p. 25). Conforme demonstrado, a capacidade de processamento de dados humana permanece inferior à de sistemas avançados. Como se viu, em 2016, o supercomputador Tianhe-2 alcançou 55 quatrilhões de operações por segundo (considerando conexões entre células de processamento), enquanto o desempenho humano situa-se em torno de 10 quatrilhões de operações por segundo (em conexões neuronais). No momento, a principal vantagem do cérebro humano reside na eficiência energética, visto que o Tianhe-2 consome aproximadamente um milhão de vezes mais energia para realizar tarefas equivalentes (Nicolelis, 2020, p. 137-138).

Outra característica distingue homem de máquina: a capacidade de situação, isto é, a capacidade de se situar no tempo e de sub-alocar energia e dedicação a uma tarefa não principal. As máquinas, mesmo aquelas chamadas de Baby-AGI (recentes tentativas de Inteligência Artificial Geral), não conseguem por si sós determinar que devem ser criados novos softwares, algoritmos e *bots* para a solução de problemas que lhes são apresentados. Para tanto, é necessário que nas linhas de código esteja presente a possibilidade dessa sub-criação. O cérebro humano não é assim, o cérebro humano enquanto lócus de processamento informativo e determinante do agir humano, um ser situacional, é distinguido (Nicolelis, 2020, p. 138-139).

A autonomia privada é o poder que os particulares têm de regular, pelo exercício de sua própria vontade, as relações de que participam, estabelecendo-lhe o

conteúdo e a respectiva disciplina jurídica interrelacional. O ser humano, sozinho ou em comunhão (pacífica ou litigiosa) tem total condição de situar-se no mundo e exigir comportamentos, ou mesmo identificar quando não se age pela regra geral. Se alguém trapacear de uma máquina de *narrow AI* no Xadrez, a primeira reação da máquina é buscar a lógica por detrás daquele movimento estranho; e é capaz de encontrar e alterar dentro de sua "consciência" para uma regra de um novo jogo, um Xadrez, enviesado. Ao contrário, o impulso do ser humano será de gritar: ei, estás me roubando!

Essa é uma diferença estrutural do funcionamento dos cérebros, de um lado o orgânico e de outro o artificial. Seres humanos não são feitos de bits (Nicolelis, 2020, p. 143). E, assim como uma máquina nunca conseguirá replicar o processo de fotossíntese das plantas, ainda que se siga à risca todo o processo bioquímico-físico de quebra do dióxido de carbono em açúcar, e consiga efetivamente açúcar "para as plantas", não se conseguirá produzir oxigênio, o que falta é o fluxo orgânico e natural que, ao longo de centenas de milhões de anos, moldou a evolução para que de maneira integrada – e até mesmo mágica – o mesmo processo que alimenta e gera energia às plantas, gera o elemento tão essencial à sobrevivência humana.

A consciência humana é uma característica distintiva que separa a humanidade das máquinas e robôs, permitindo-lhes compreender e analisar perspectivas e visões gerais. Embora as máquinas possam processar mais informações do que o cérebro humano, elas carecem da capacidade de contextualização e de uma autêntica autonomia, que se funda na dignidade humana, que é para os seres humanos, e que poderia ser estendida para seres animais, mas nunca para um carro, para uma geladeira, ainda que smart e, portanto, nunca a poderia ser reconhecida às máquinas. A autonomia é uma característica distintiva dos seres humanos, permitindo-lhes regular as relações nas quais participam.

A liberdade individual, a proteção dos dados pessoais e a privacidade são garantias constitucionais que, em uma via de mão dupla, fundamentam e são fundamentadas pela autodeterminação individual. Isso implica, na modernidade, o reconhecimento do direito à explicabilidade das decisões automatizadas, como forma de garantir a proteção dos dados e da privacidade dos indivíduos, não apenas para guarnecer-lhes o direito, mas para torná-los conscientes sobre os motivos determinantes de decisões tomadas a seu benefício ou prejuízo.

É possível compreender que, desde logo, do art. 5º, caput, da Constituição, decorre a liberdade individual, do inciso X decorre a inviolabilidade da intimidade e da vida privada e do inciso XII decorre o livre desenvolvimento, a proteção dos dados pessoais (recentemente com a aprovação da EC 115/2022) e, portanto, a autodeterminação individual. Pode-se afirmar que a soma interpretativo-construtivista resulta no reconhecimento do direito implícito à explicabilidade das decisões automatizadas fundada na autodeterminação individual, na proteção dos dados e na privacidade como uma garantia que decorre do texto constitucional, como também o era antes também com os dados antes de 2022 (Mendes, 2011).

Há, portanto, além formalmente, um direito fundamental materialmente e implícito à explicabilidade e à autodeterminação na CF/88, isto é, um direito fundamental associado. Segundo Gavião Filho e Freitas (2020, p. 240-241), os direitos fundamentais implícitos são aqueles que não estão expressamente dispostos no constitucional, mas dependem de uma formulação interpretativo-argumentativa para existir. Como se propõe. Veja-se em dimensões a autonomia privada e a autodeterminação existem e se complementam.

Na Alemanha, em 1983, foi decidido acerca da autodeterminação informativa. No clássico julgamento do Tribunal Constitucional Federal Alemão (Bundesverfassungsgericht) BVerfG (conhecido como «*Census Judgment*»), por meio do direito fundamental de personalidade, deduz-se a existência de uma proteção à autodeterminação informativa, proporcionando aos indivíduos um campo de salvaguarda contra a ocorrência de discriminações por sistemas (*machine learning*, big data, inteligência artificial, v.g.) geridos por algoritmos. A falta de controle dos indivíduos acerca de seus dados, gerada pela sua comercialização desenfreada e não consentida, despertou nos legisladores de diversos países o desejo de criar regras sobre o seu tratamento. Mesmo antes da propagação de leis de proteção de dados, o Tribunal Constitucional Federal alemão (BVerfG), em 1983, já havia fixado histórico precedente em favor da «autodeterminação informativa» (*Informationelle Selbstbestimmung*), entendendo ser um direito derivado da personalidade e da dignidade humana (Schwabe, 2005, p. 233-234).

Portanto, desde a aludida data, passou a vigorar na Alemanha veredito capaz de permitir a judicialização de discriminações geradas pelo tratamento de dados – certamente que, naquela época, em meio físico. Em que pese o Brasil ter

sistematizado e incorporado uma proteção robusta à privacidade e aos dados pessoais e sensíveis, por meio da Lei nº. 13.709/18 (Lei Geral de Proteção de Dados, 2018), apenas permitindo que seu tratamento ocorra caso autorizado por alguma das suas dez bases legais (artigo 7º), ainda assim existe uma insuficiência, na legislação regente, para a efetiva tutela judicial de discriminações geradas por algoritmos no âmbito digital.

A decisão do BVerfG ficou conhecida como «*census judgment*» (julgamento do censo), visto que se referia a queixas da população alemã contra a recém aprovada Lei do Censo («1983 Federal Census Act»), que previa que, para a realização da contagem populacional, seriam coletadas diversas informações pessoais dos cidadãos, pelas quais eram indesejadas de serem fornecidas por alguns deles. A despeito do veredito ter sido, parcialmente, desfavorável aos requerentes, o Tribunal aproveitou o contexto para determinar medidas acerca do tratamento de dados, inclusive anulando partes da lei. Entendeu a Corte que o que prevê o direito ao livre desenvolvimento da personalidade estava diretamente correlacionado ao artigo que dispõe acerca da cláusula geral da proteção da dignidade da pessoa humana e, a partir de sua leitura conjunta, haveria uma noção associada de garantia à autodeterminação do indivíduo acerca das informações compositivas de sua personalidade. Renova-se, como em Leal (2003, p. XVI), que o Estado Democrático de Direito significa mais que uma resposta pela mera legalidade, mas, de uma "legalidade qualificada por valores e princípios materiais, que se afiguram como sendo a base de todo o ordenamento jurídico".

Isso autoriza o entendimento de que autodeterminação informativa, além de existir enquanto um direito fundamental, ainda não se lastreia como um direito de propriedade do indivíduo, podendo ser disposto em qualquer circunstância, mas como um direito fundamental instrumental, cujo papel na formatação constitucional é de parametrizar a interpretação e aplicação como escopo protetivo autônomo e associável.

A autodeterminação informativa engloba duas posições que o indivíduo pode ter em relação às informações processadas sobre si: o direito de saber (*right to know*) e o direito de não saber (*right to not know*). O primeiro consiste no direito do sujeito afetado em conhecer os resultados do tratamento de seus dados, ou seja, na necessidade de transparência. Já o segundo indica que, caso não queira conhecer os

resultados, a pessoa não deve ser informada sobre eles ou, se já os conhece, deve ter o poder de exclui-los. Sem a liberdade de exercer esses direitos de saber e de não saber, não é possível garantir a verdadeira autodeterminação informativa (Schwabe, 2005, p. 233-234).

A autodeterminação informativa é um direito fundamental que abrange tanto o direito de saber quanto o direito de não saber sobre as informações processadas a respeito de uma pessoa. Esses direitos são fundamentais para garantir transparência e privacidade, e a falta de liberdade para exercê-los compromete a autodeterminação informativa. No entanto, a aplicação desses direitos pode gerar conflitos com outros direitos fundamentais, como a liberdade religiosa e de expressão, como se verá no caso do padeiro Jack Phillips. Esse conflito de direitos evidencia a complexidade na aplicação dos direitos fundamentais, incluindo a autodeterminação informativa, e a necessidade de encontrar um equilíbrio que respeite todos os direitos envolvidos. Veja-se o caso.

Em 2012, o casal Charlie Craig e David Mullins foi a uma confeitaria em Colorado, nos Estados Unidos, para encomendar um bolo de casamento personalizado. No entanto, o proprietário da loja, Jack Phillips, se recusou a criar o bolo, alegando que isso violaria suas crenças religiosas. O casal processou Phillips por discriminação e o caso foi parar na Suprema Corte dos Estados Unidos. Em 2018, a Corte decidiu que Phillips não violou as leis antidiscriminatórias do Colorado, alegando que a Comissão de Direitos Civis do estado demonstrou hostilidade em relação às crenças religiosas de Phillips ao lidar com o caso. A decisão da Suprema Corte gerou um intenso debate sobre o equilíbrio entre o direito à igualdade e à não discriminação e a liberdade religiosa e de expressão.

O caso mencionado refere-se à recusa do confeiteiro Jack Phillips em criar um bolo personalizado para um casamento entre pessoas do mesmo sexo, o que gerou uma acusação de discriminação. Em 2018, a Suprema Corte dos Estados Unidos absolveu Phillips da acusação, alegando que a decisão do confeiteiro se baseava em suas crenças religiosas. Veja-se, mesmo um ato discriminatório precisa ser explicado.

Agora, necessário se faz relacionar o caso *Masterpiece Cakeshop v. Colorado Civil Rights Commission*, 584 U.S. (2018) ao direito à liberdade (de expressão e religiosa) e à privacidade, mais especificamente ao direito de ser deixado em paz *(right*

to be left alone). O confeiteiro alegou que criar um bolo para um casamento entre pessoas do mesmo sexo violaria suas crenças religiosas, o que o levou a se recusar a fornecer o serviço. Embora essa decisão possa ter sido motivada por suas crenças pessoais, ela também pode ser vista como uma afirmação de seu direito à privacidade, ou seja, o direito de se afastar de situações que violem suas convicções pessoais.

Esse caso pode ser visto como um exemplo de como as decisões automatizadas podem afetar o direito à privacidade e ao direito de saber o motivo da recusa de certo acesso a bens ou serviços. Os algoritmos que usam dados pessoais para tomar decisões automatizadas podem afetar o direito à privacidade, já que os dados pessoais dos indivíduos são usados para tomar decisões que afetam suas vidas. Além disso, a falta de transparência sobre como esses algoritmos tomam decisões pode afetar o direito de saber o motivo da recusa de certo acesso a bens ou serviços, o que configura uma injustificada discriminação.

Em suma, quando decisões — humanas ou automatizadas — afetam trajetórias de vida, não basta o resultado: é indispensável a razão. Quem decide deve explicar, e quem detém o conhecimento sobre a forma de decidir deve demonstrar que o caminho escolhido era necessário e o menos gravoso possível. Esse é o núcleo do dever de proteger: conter desigualdades, evitar exclusões e garantir meios reais de contestação. Sem isso, a aparente neutralidade da técnica encobre velhas discriminações com nova técnica. Por isso, torna-se decisivo tornar visíveis as bases de cada decisão.

Assim, é importante que haja transparência e explicabilidade nas decisões automatizadas, de modo que os indivíduos afetados – bem como o Estado – possam entender como essas decisões são tomadas e contestar decisões que possam ser discriminatórias ou violar seus direitos. Além disso, é importante que sejam estabelecidos critérios claros para o uso de algoritmos em decisões importantes, para garantir que esses sistemas sejam desenvolvidos com transparência e equidade. Assim, percebe-se com facilidade que a falta de transparência pode conduzir a situações de violação da privacidade e discriminação. Quando os dados são tratados de maneira opaca, os indivíduos não têm controle sobre o que está sendo coletado, como é utilizado e quem tem acesso a esses dados. De igual forma, quando esses dados são utilizados para a tomada de decisão para a formação de *scoring* de crédito, *profiling* para *microtargeting* e *micromarketing*, ou ainda a recusa de um serviço ou

emprego, ficando apenas nesse campo em que os dados pessoais são centralmente utilizados, se percebe uma visível necessidade de se poder sindicar a funcionalidade destes dados, isto é, se jogar luzes e vencer a opacidade da cortina de ferro, ou da *blackwall* da *blackbox*.

O tratamento não transparente dos dados e as decisões opacas reduzem o controle pessoal e podem violar a privacidade e violando a não-discriminação, contrariando a ordem constitucional. É fundamental que empresas e órgãos governamentais adotem práticas de transparência, assegurem aos indivíduos o controle sobre suas informações pessoais e implementem medidas eficazes para proteger a privacidade e prevenir qualquer forma de discriminação.

Para fins deste trabalho, emprega-se uma distinção operacional: transparência é a disponibilidade inteligível de informações essenciais sobre o sistema, responde “o que”; explicabilidade é a capacidade de reconstruir, em nível adequado ao risco, o itinerário lógico-decisório do modelo – inclusive dados e critérios de inferência, responde ao “como e por quê”; auditabilidade é a possibilidade técnica e jurídica de escrutínio independente do sistema e de seus resultados, responde ao “a quem compete verificar”; rastreabilidade é a existência de trilhas e registros que permitam relacionar entradas (*inputs*), transformações (processamento) e saídas (*outcomes*), responde ao “com base em que, em quais etapas do raciocínio lógico”. Esses quatro elementos – graduados por risco e impacto – compõem o núcleo mínimo de justificabilidade pública exigível de sistemas automatizados que afetem direitos fundamentais.

Para que se possa entender os verdadeiros riscos da nocividade da opacidade no tratamento de dados, é importante entender como e de quais formas a opacidade pode levar à discriminação e violação da privacidade, não apenas de modo direto e imediato – tema que vem dominando as pesquisas acerca da discriminação algorítmica –, mas a partir de uma matriz histórica e filosófica considerada a partir de uma leitura conjectural e analógica do atual momento para com os momentos de restrição das liberdades individuais pelas possibilidades quase ilimitadas do exercício do poder. Burrell (2016, p. 3-5) identifica três formas de opacidade algorítmica que se sobrepõem e reforçam mutuamente. A primeira é a opacidade intencional, resultante de segredos corporativos ou estatais que protegem modelos e dados sob o pretexto de concorrência ou segurança, impedindo qualquer escrutínio externo. A segunda

decorre da iliteracia técnica: a escrita e a leitura de código exigem habilidades especializadas em programação e em conceitos de ciência de dados, criando uma barreira de acesso para a maior parte dos usuários e decisores. Por fim, há a opacidade intrínseca aos próprios algoritmos de aprendizado de máquina e à escala de dados que manipulam, onde a complexidade matemática e o elevado número de variáveis transformam o processo de treinamento e inferência em uma “caixa-preta”, incompreensível mesmo para especialistas. Reconhecer cada um desses tipos de opacidade é essencial para direcionar soluções adequadas, que vão desde exigências regulatórias e auditorias independentes até iniciativas de educação em literacia computacional e técnicas de explicabilidade de modelos.

Para combater o mal da obnubilação algorítmica, há um antídoto: a XAI (*Explainable Artificial Intelligence*), que nada mais é do que uma IA que tem a habilidade de explicar as suas decisões – o que não é uma obviedade, tampouco uma garantia em se tratando de IAs. O uso de mecanismos que permitam a construção de modelos de inteligência artificial explicáveis ou cujas derivações e explanações possam ser identificáveis quando necessários resultam em um aprimoramento ótimo na busca pela identificação e possível correção de vieses discriminatórios. Claro, não há garantia universal de uma completa explicação, mas isso deve ser constantemente buscado e deve ser uma preocupação dos desenvolvedores e dos controladores. A importância da XAI é de dupla natureza. A XAI desempenha um papel fundamental na construção (*by design*) de algoritmos mais transparentes e justos, que permitem a rastreabilidade de suas decisões automatizadas, permitindo que desenvolvedores e demais agentes com aptidão técnica entendam como suas decisões foram tomadas pelos algoritmos não programados, mas também ajuda a identificar e corrigir possíveis vieses discriminatórios.

O problema está naquilo que aponta Postema (2022, p. 299-300) acerca do raciocínio jurídico artificial, inclusive ponderado se esse seria alcançável por uma máquina. Postema pondera que não há certeza quanto ao conteúdo de uma decisão tomada por um raciocínio artificial, isto é, uma decisão tomada por uma qualquer lógica que não uma lógica jurídica. O problema das decisões algorítmicas é a falta de explicabilidade quanto ao pensamento (raciocínio) que a efetivou, só que esse raciocínio robótico, se desconstruído, não passaria de mecanismos frios de engenharia de código sem qualquer necessária vinculação a critérios metaéticos,

éticos ou mesmo jurídicos. De modo que a explicabilidade que se fala é a explicabilidade em relação ao rumo decisório dos procedimentos robóticos (algorítmicos, pois não-físicos, mas cujo aspecto se físicos fossem seriam de movimento unicamente mecânicos).

Explicabilidade da decisão automatizada, a partir de uma engenharia de XAI, que permite essa explicabilidade em momento posterior, não consolida, portanto, o dever de fundamentação das decisões – falando propriamente no caso de serem essas usadas para decisões judiciais –, apenas explicita o passo-a-passo adotado pela máquina. Isso é suficiente ou desejável? A explicabilidade técnica não exonera o dever constitucional de fundamentação: ao decidir com apoio de sistemas automatizados, a autoridade deve converter os elementos técnicos em razões públicas verificáveis, articulando fatos, normas e justificativas proporcionais. A opacidade invocada como “segredo” não pode servir de escudo para decisões sem motivação adequada – núcleo do devido processo e condição de legitimidade.

A XAI se faz essencial para estabelecer a confiança do público na tecnologia, ao mesmo tempo que impõe um ônus pesado – mas vencível – ao desenvolvedor. À medida que a Inteligência Artificial desempenha um papel cada vez mais proeminente em nossa sociedade, a capacidade de explicar e compreender seu funcionamento se torna uma exigência tanto do ponto de vista ético quanto regulatório, colocando mais uma garantia acerca de que a tecnologia utilizada – aprovada nos demais testes – atue de forma responsável e equitativa, e, de igual sorte, isso decorrerá pelo próprio desenho da aplicação. A XAI está presente enquanto uma possibilidade técnica do cenário tecnológico, mas também é um excelente ponto de contato com a dinâmica jurídica, visto que sua existência e alocação enquanto direito individual ou imposição à desenvolvedora funda-se no dever de transparência, cuja dimensão individual aos credores jusfundamentais é de direito fundamental e, portanto, oponível e reclamável. A XAI não exonera o dever de motivação da decisão – ainda que meramente posterior.

O grande gargalo da proteção jusfundamental à discriminação algorítmica encontra-se nessa opacidade das aplicações algorítmicas, protegidas pelo sigilo comercial – segredo comercial, segredo industrial, segredo de programação ou, ainda, segredo de propriedade intelectual – que encontra guarida jusfundamental no direito de proteção à produção intelectual. As ferramentas algorítmicas e de Inteligência Artificial começam a fazer parte cada vez mais da vida hodierna, a partir da tomada

de decisão automatizada há um direcionamento já alavancado na vida humana, já estruturando um espaço próprio – de difícil remoção – de áreas como o mercado de empregos, planos de saúde, acesso ao crédito e mesmo acesso à justiça – e a prestação jurisdicional. Nesse sentido, a falta de transparência e de explicabilidade na forma como se dão tais decisões se mostra desarrazoada, literalmente, visto que joga sombras sobre as próprias decisões que, mesmo contra a vontade do usuário impactado, contam principalmente com a tomada de decisão pela máquina e, quando muito, com a confirmação humana, essa mesma já enviesada pelo viés de confirmação comum aos humanos, atarefados que somos.

Atualmente, há uma candente necessidade pelo advento de respostas acerca da colisão entre os direitos de transparência algorítmicas, que passam a surgir diante da nova dinâmica de discriminações algorítmicas automatizadas e a garantia de um mínimo de conservação dos direitos de propriedade intelectual sobre tais aplicações. Até o presente momento, as poucas amostras de respostas-base do poder judiciário – assim como da regulação – tem sido a de tentar encontrar na programação de tais aplicações uma caixa-preta a ser aberta e descoberta, transferindo uma relação civil de consumo para uma nova verticalização (como na eficácia Estado-Indivíduo) em que à empresa se guarda o ônus de proteção e promoção da não discriminação.

A tensão entre o direito à transparência algorítmica e a proteção da propriedade intelectual demanda análise proporcional e circunstanciada, balizada por critérios constitucionais e sensíveis às finalidades envolvidas. Em juízo, esses elementos devem ser convertidos em razões públicas verificáveis, assegurando contraditório efetivo sem dissolver, de plano, segredos técnicos: a abertura será graduada por proporcionalidade, com acesso controlado aos artefatos estritamente necessários à verificação de ausência de discriminação injustificada.

A saída naturalmente encontrada até o presente momento indica um caminho com nitidez, a relativização dos direitos de propriedade (eminentemente privados) em detrimento daqueles direitos de proteção à igualdade (eminentemente públicos), de modo a abrir a caixa algorítmica e desvelar o “princípio das dores” da falha na programação que gerou essa ou aquela decisão discriminatória. O peso será ainda maior quando a vítima da aludida discriminação seja parte de um grupo vulnerável, estrutural e historicamente já discriminado e escanteado da sociedade. Além disso, sistemas que decidem de maneira autônoma e sigilosa incrementam a discriminação

e a desigualdade que são potencializados quando vêm a ocorrer em meio digital ou a partir da interação com *softwares* que são abastecidos com dados pessoais que, anonimizados ou não, conseguem triangularizar perfis, tendências e riscos, incrementando ainda mais a prejudicialidade de grupos historicamente marginalizados.

Assim, ao mesmo tempo em que há uma importante reestruturação e releitura a se realizar na ordem democrático-jurídico-constitucional nesse novo momento histórico e jurídico em prol da busca de maior efetividade aos direitos de igualdade e não-discriminação frente aos avanços algorítmicos, em uma perspectiva estrutural, também há se que efetuar uma defesa da dignidade existente nos direitos fundamentais de proteção à propriedade intelectual, buscando-se, acima de tudo, o equilíbrio necessário ao reestabelecimento da ordem e da paz pública após o acontecimento de danos potenciais pelas aplicações automatizadas, o que deve ser resolvido com prioridade protetiva aos grupos vulneráveis, dadas as possibilidades fáticas e jurídicas de cada caso. Proteger a propriedade algorítmica não é uma carta branca à opacidade, mas é âncora estritamente necessária para a aplicação da proporcionalidade informacional. A propriedade intelectual continua relevante, mas não pode virar escudo para esconder práticas que ampliem desigualdades. Nesses casos, é razoável pedir que quem usa a tecnologia explique, em termos simples, por que decidiu assim e quais cuidados tomou para evitar discriminação, apresentando elementos que permitam conferir esses pontos.

4.2 Da gramática técnica à razão pública justificada: bases para o controle judicial de discriminações algorítmicas

Antes de ingressar no exame da jurisdição constitucional, cumpre justificar a incursão em linguagem de modelos e *feedbacks* – própria da engenharia de dados – como requisito epistêmico do controle judicial. É por essa gramática técnica que se compreende o risco, se documenta o percurso decisório e se distribui o ônus argumentativo: ciclos de realimentação, curadoria de dados, validação e métricas por subgrupos, além de *logs* e *datasheets* transformam a opacidade em material cognoscível pelo devido processo. Como a proporcionalidade opera sobre fatos e razões públicas, aferir adequação, necessidade e o balanceamento material exige

verificar desenho do sistema (incluindo supervisão humana significativa), qualidade e governança dos dados e mecanismos de explicabilidade. Sem essa base empírica – e sem o mapeamento dos ciclos de *feedback* que amplificam discriminações – a jurisdição carece de substrato para impor medidas proporcionais, convertendo a tutela antidiscriminatória em retórica. É, pois, nessa ponte entre engenharia e razão pública que se delimita o que a jurisdição efetivamente pode realizar.

Os modelos de linguagem são ferramentas poderosas para gerar e compreender textos, mas também podem reproduzir vieses e estereótipos presentes nos dados que os treinam. No entanto, um estudo recente do laboratório de inteligência artificial *Anthropic* mostrou que esses modelos podem se autocorrigir para alguns desses vieses, se receberem instruções simples em linguagem natural para isso (Firth, 2023).

Os grandes modelos de linguagem (LLMs, na sigla em inglês – *Large Language Models*) são sistemas computacionais que aprendem a gerar ou processar texto a partir de grandes quantidades de dados linguísticos. Nos últimos anos, os MLs baseados em redes neurais profundas, como BERT, GPT, desde o modelo 2, até o atual modelo 5, BARD, Chat do Bing, DeepSeek, dentre outros, têm demonstrado avanços impressionantes em diversas tarefas de processamento de linguagem natural (PLN), como compreensão de texto, geração de texto e resposta a perguntas. No entanto, esses MLs também sofrem de problemas sérios relacionados aos vieses presentes nos dados de treinamento, que podem refletir estereótipos e preconceitos da sociedade. Esses vieses podem se manifestar de forma indesejável e prejudicial no comportamento dos MLs, gerando textos (também decisões) discriminatórios.

Firth (2023) apresenta uma nova abordagem para capturar geometricamente os vieses morais e éticos dos MLs, usando uma "direção moral" que pode ser computada no espaço de incorporação dos MLs. Essa "direção moral" pode avaliar a normatividade (ou não-normatividade) de frases arbitrárias sem treinar explicitamente o ML para essa tarefa, refletindo bem as normas sociais. Além disso, Firth (2023) demonstra que o uso da "direção moral" pode fornecer um caminho para atenuar ou até prevenir a degeneração tóxica nos MLs, mostrando essa capacidade no conjunto de testes RealToxicityPrompts.

Firth (2023) indica ainda que os pesquisadores Amanda Askell e Deep Ganguli analisaram modelos de linguagem de diferentes tamanhos, que haviam

passado por diferentes quantidades de treinamento com reforço a partir de feedback humano - *reinforcement learning from human feedback* (RLHF) –, uma técnica que faz com que humanos orientem o modelo de Inteligência Artificial para produzir respostas mais desejáveis. Eles testaram esses modelos com três conjuntos de dados que foram projetados para medir o viés ou a estereotipação em relação a idade, raça, gênero e outras categorias. Eles descobriram que apenas solicitar ao modelo que garantisse que suas respostas não dependessem de estereótipos tinha um efeito positivo dramático em sua saída (*output)*, especialmente naqueles que haviam completado rodadas suficientes de RLHF e tinham mais de 22 bilhões de parâmetros, as variáveis em um sistema de Inteligência Artificial que são ajustadas durante o treinamento. Em alguns casos, o modelo até começou a se envolver em discriminação positiva em sua saída. Esse resultado é notável porque mostra que os modelos de linguagem podem aprender a evitar vieses sem precisar de uma definição formal ou explícita do que é viés. Isso sugere que eles podem captar pistas implícitas ou contextuais sobre o que é considerado aceitável ou inaceitável em diferentes situações. No entanto, os pesquisadores não sabem exatamente como os modelos são capazes de fazer isso, embora tenham algumas hipóteses.

Uma das hipóteses é a de que à medida que os modelos ficam maiores, eles também têm conjuntos de dados de treinamento maiores, sendo que nesses conjuntos de dados há muitos exemplos de comportamento tendencioso ou estereotipado. Outra é que o *feedback* humano pode fornecer uma orientação valiosa para os modelos sobre como se comportar de forma mais ética ou socialmente responsável. O estudo tem implicações importantes para o desenvolvimento e o uso de modelos de linguagem em diversas aplicações, como assistentes virtuais, *chatbots*, sistemas educacionais e jornalismo automatizado. Ele sugere que esses modelos podem ser treinados para produzir conteúdo menos tóxico ou ofensivo, simplesmente pedindo-lhes para fazê-lo (Firth, 2023), ao contrário do que defende, dentre outros O'Neil (2020, p. 15), para quem os *feedbacks* sobre aplicações ruins as pioram ou as inutilizam. No entanto, isso também levanta questões sobre quem deve definir o que é viés e como garantir que os modelos não sejam manipulados para produzir conteúdo enviesado intencionalmente (Firth, 2023).

A inteligência artificial, ao contrário do que é sugerido pelo nome, não é nem inteligente, nem artificial. É, na verdade, comandada por uma série de seres humanos

que operam em momentos específicos o funcionamento dos mecanismos algorítmicos e robóticos, de modo que não é puramente artificial. Também não é puramente inteligente. Inteligência é uma palavra derivada de *inter-legere*, isto é, eleger dentre; uma competência tipicamente humana que foi formatada ao longo de anos de experiência dos seres-humanos em saber distinguir as coisas do mundo (Han, 2021, 46).

Se, como sugere Han (2018), o digital tende a dissolver a experiência pública em uma lógica de enxame – reativa, acelerada e pobre em elaboração – então a promessa de neutralidade algorítmica opera também como empobrecimento do político: menos deliberação, mais otimização; menos razões, mais métricas artificialmente constatadas. É precisamente aí que a advertência de Morozov (2018, p. 43-48) se torna convergente, no ponto em que defende que o tecnossolucionismo converte conflitos normativos em “problemas de engenharia”, como se bastasse refinar o modelo para que desaparecessem as assimetrias que são, antes, sociais, históricas e institucionais. A consequência para o Direito é direta, visto que a tutela contra discriminação algorítmica não pode ser desenhada como correção tecnicamente eficiente de “erros do sistema”, e, por isso, deve ser uma recondução da decisão ao espaço das razões públicas, onde justificabilidade, contestação e responsabilidade voltam a ser o centro – e onde a opacidade deixa de ser destino e retorna a ser escolha, se os efeitos por ela gerados persistirem, o problema é ainda anterior.

A participação humana na construção das Inteligências Artificiais e dos algoritmos é imprescindível, cada participação humana importa. Seja no processo de formação, treinamento e programação da Inteligência Artificial, seja no processo de uso e *feedbacks* da ferramenta. O uso desses sistemas nem sempre são iguais, há diversos tipos de programações, com maior ou menor dependência humana, mas em dado grau são utilizados dados, e a seleção e o cuidado com os dados selecionados ou demonstrados para a Inteligência Artificial auxilia no desenvolvimento das discriminações. Veja-se, um ponto eticamente buscado é o da promoção e incentivo para com a diversidade nas equipes de desenvolvimento das aplicações algorítmicas, sobretudo, além da programação, na fase de tratamento dos dados e de treinamento da Inteligência Artificial com tais bases de dados. Isso tem se destacado como uma prática essencial, pois contribui para uma visão mais inclusiva e democrática durante

todo o processo de implementação de algoritmos, reduzindo a probabilidade de reprodução de preconceitos. A diversidade nas equipes de desenvolvimento é prática amplamente recomendada e na qual se apostam esforços em torno de implementação via políticas públicas e, também, conscientização.

O uso de conjuntos de dados diversificados e representativos durante o treinamento dos algoritmos pode ajudar a reduzir o viés. Técnicas de aprendizado de máquina justas e transparentes podem ser utilizadas para garantir que os algoritmos tomem decisões justas e explicáveis. No campo do direito, a criação de leis e regulamentos específicos para lidar com a discriminação algorítmica é crucial, mas a partir de um olhar mais tocado à realidade do dia a dia do desenvolvimento e envolvimento com a progressão dessas tecnologias, não há de ser ignorado o fato de que a medida de presença representativa mínima e inclusiva nas equipes de desenvolvimento de aplicações pode, se não reduzir drasticamente o problema, indicar por uma boa-fé da equipe e da empresa em buscar se adequar no campo do dever de respeito à não-discriminação – direito já consagrado na normatividade pátria, e que virá reforçado quando da aprovação do PL sobre o uso da Inteligência Artificial (nº 2.338/2023), atualmente previsto no Art. 3º, IV, do PL.

A diversidade nas equipes de desenvolvimento desempenha um papel crucial na mitigação da discriminação algorítmica. A programação de algoritmos muitas vezes reflete as perspectivas e experiências de quem os desenvolve, e equipes diversas são essenciais para garantir que essas perspectivas sejam inclusivas e representativas. A falta de diversidade pode levar a preconceitos e discriminações não intencionais nos algoritmos, uma vez que as decisões acerca da programação podem ser influenciadas pelos vieses dos desenvolvedores que, não por maldade, mas por desconhecimento fático (Hao, 2019), pode passar ao largo das microagressões que possam vir a redundar em uma violência automatizada.

Sob um viés individual do programador, no entanto, deve ser atacado o fato de que programadores indiferentes com questões sociais, por exemplo, a questão racial – ou mesmo programadores (in)conscientemente preconceituosos ou, no exemplo, racistas – não se importarão com uma base de dados pouco diversificada (O'Neil, 2020, p. 37). O desapego para com a diversidade social representada na base de dados é uma das raízes de nível mais próximo à superfície do mal da discriminação algorítmica.

Por isso, é crucial que os programadores levem em consideração a diversidade social e as questões de discriminação algorítmica durante o desenvolvimento de sistemas e algoritmos, tanto em sua programação quanto na organização de sua base de dados. É preciso tomar medidas para mitigar esses vieses e garantir que as soluções tecnológicas sejam justas e equitativas, mesmo que para tanto se exija mais do que meramente a não discriminação incidental; é necessário o combate ativo à discriminação. Para tanto, é preciso um esforço consciente para incorporar às etapas do processo de desenvolvimento a diversidade enquanto conceito a liderar e subsidiar todas as etapas. A questão é: como fazer isto?

Promover a diversidade nas equipes de desenvolvimento não é apenas uma questão ética, mas também uma estratégia eficaz na luta contra a discriminação algorítmica. Essa diversidade não se limita apenas à representação de gênero e etnia, mas também à inclusão de pessoas com diferentes formações acadêmicas, idades e experiências de vida. Equipes diversas estão mais bem posicionadas para avaliar e aprimorar a equidade dos algoritmos, reduzindo assim o impacto de preconceitos indesejados nos resultados. A inclusão de uma variedade de perspectivas no processo de desenvolvimento de algoritmos é fundamental para garantir que as tecnologias reflitam de maneira mais justa a sociedade como um todo e contribuam para a construção de um ambiente mais igualitário, que atualmente em 2019 contava com apenas 18% de mulheres na indústria da Inteligência Artificial como um todo e 2,5% e 4% de negros, respectivamente no Google, e empatados Facebook e Microsoft (Hao, 2019).

Além disso, a diversidade nas equipes de desenvolvimento também promove a inovação e a criatividade. Diferentes perspectivas e experiências podem levar a soluções mais criativas e eficazes na construção de algoritmos e na resolução de problemas complexos. Isso é especialmente importante em um campo tão dinâmico e desafiador quanto a tecnologia da informação, onde a capacidade de inovação pode fazer a diferença entre o sucesso e o fracasso. A diversidade e a curadoria documentada de dados deixam de ser ética programável para se tornarem deveres auditáveis e, portanto, evidência em juízo no atual desenho.

O cuidado com a justiciabilidade dos direitos fundamentais deve, obviamente, passar por atenciosa verificação destes pormenores técnicos, que realmente fazem a diferença na construção de um modelo racional e das tentativas de construção de vias

interpretativas para solucionar o atual problema: como tutelar direitos fundamentais frente à discriminação algorítmica automatizada.

A discriminação algorítmica exige uma nova abordagem jurídica para proteger direitos fundamentais diante das inovações tecnológicas. Será necessário analisar direitos em conflito, incluindo direitos implícitos a serem deduzidos, para equilibrar os interesses envolvidos e equalizar racional e proporcionalmente os direitos em questão. A transparência destaca-se com especial protagonismo acerca da explicabilidade, que alegoricamente reflete-se como o fogo de Prometeu, trazendo consciência e jogando luzes às ocorrências e aos desalinhamentos programáticos que resultaram na injusta agressão a direito pela aplicação robótica. Funcionando como uma cláusula fiduciária dentro da programação, garantindo desde o desenho do algoritmo, a possibilidade de que seus *outcomes* sejam explanáveis e sindicáveis, na medida e no grau que seja reputado necessário a partir da injusta agressão ocasionada, a explicabilidade é vital enquanto um direito a ser reconhecido, e como dever dos desenvolvedores.

A salvaguarda dos direitos individuais é imperativa no contexto dos algoritmos, exigindo proteção sempre que houver lesão ou risco a direitos fundamentais. Contudo, isso não significa dizer que sempre e em qualquer caso, seja qualquer o grau de afetação à direito fundamental, se torna legítima a total abertura de códigos-fonte, chaves algorítmicas, programação, caminho decisório automatizado ou demonstração do funcionamento da rede neural. Há que se ter especial cautela para que a busca da efetivação do direito fundamental de uma das partes envolvidas não cause a total desproteção do outro direito fundamental cuja medida aparentemente resolva a colisão. Não se pode extinguir o diferencial competitivo de uma empresa desenvolvedora de Inteligências Artificiais pela busca implacável de uma solução dentro da total publicização da caixa-preta algorítmica.

Há, no entanto, uma clara necessidade de revisar a mera eficácia horizontal dos direitos fundamentais nesse sentido. O domínio de mercado e o alarmante risco da opacidade algorítmica nesse mercado cada dia mais inescapável (Ferrari, 2021, p. 97) aponta para uma atualização acerca do pensamento sobre a paridade de armas, isto é, de eficácias jusfundamentais, dos direitos cuja colisão vêm se anunciando. Os direitos fundamentais, como se sabe, formam uma ordem objetiva de valores (*Wertordnung oder Wertsystem*) que iluminam o direito, a partir do direito

constitucional (Leal, 2007, p. 65). Basta, para tanto, o reconhecimento da dimensão objetiva dos direitos fundamentais e sua irradiação sobre a ordem privada, isto é, a função de estruturar padrões mínimos de proteção e de distribuir ônus justificatórios em contextos de assimetria. Quando agentes privados detêm poder estrutural e (deveriam ter) controle informacional sobre fluxos decisórios algorítmicos, emergem deveres positivos proporcionais de prevenção de vieses, documentação do processo e explicabilidade suficiente. São parâmetros que vinculam a atuação privada que deve ser exigível pelos riscos comuns à própria atividade algorítmica.

O fato é que os grandes desenvolvedores de aplicações tecnológicas de ponta, como as Inteligências Artificiais em comento, não devem ocupar um mesmo patamar que os seus usuários ofendidos. Também não é correto afirmar que devem figurar ao lado do Estado, pois a sua dimensão original não é aquele de primeira dimensionalidade, de defesa em relação ao Estado (*Abwehrrechte des Bürgers gegen den Staat*). Celeste (2019, p. 4) enxerga útil a roupagem do constitucionalismo digital para repensar a localização desses grandes *players* do mercado internacional de tecnologia na efetividade dos direitos fundamentais individuais. Requisitar maior atenção à eficácia dos direitos individuais desses conglomerados, e de outros que auxiliam esses, se torna mais importante que uma mera eficácia privada, horizontal. O controle do poder e a justificabilidade das razões sempre foram o coração do constitucionalismo; é sobretudo nesse ponto que a automatização algorítmica parece obscurecer.

Isso, no entanto, não está consolidado. A mudança é necessária. A Magna Carta figura como um dos marcos inaugurais e juridicamente relevante do movimento que submeteu não só os súditos a um texto legal vinculante, mas também os barões da nobreza e, em especial, o soberano maior do Estado, o Rei, que até então tinha poder absoluto sobre todas as esferas da vida civil, militar, religiosa e política. Este momento histórico é dado como o marco principal, ainda que em fase embrionária, do movimento chamado de constitucionalismo, que lutou, e ainda luta, pela concretização de direitos e garantias inalienáveis, certas e universais.

Esta digressão histórica não é ornamental: ela mostra como a passagem do poder absoluto ao poder limitado por razões públicas funda a gramática do constitucionalismo. No ambiente digital, plataformas operam como infraestruturas

privadas de poder; por isso, as mesmas ideias de limites, publicidade e controle precisam ser reatualizadas para sistemas algorítmicos.

Paradoxalmente, não é estranho trazer a Magna Carta para pensar sobre Inteligência Artificial e algoritmos. Nos debates que a instruíram, no longínquo ano de 1215, foram estabelecidas certas bases do direito de defesa do homem em relação ao Estado, o que vem a ser mais bem definido a partir do constitucionalismo de 1ª geração, dos chamados direitos de defesa, em que o Estado tinha limites à imposição de sua vontade ao cidadão, pois este agora tinha uma poderosa arma: direitos de proteção e de liberdade, em *status negativus* e *status activus*, respectivamente, na perspectiva de Jellinek, a romper com o preexistente *status subjectionis*, que previa uma verdadeira ilimitação do poderio estatal sobre o indivíduo. Uma liberdade galgada na possibilidade de se autodeterminar enquanto indivíduo – na perspectiva diante do Estado –, povo ou nação nos *Status europaeus* e *universalis*, posteriormente (Brugger; Leal, 2007, p. 126-131). Especificamente ao *status activus*, de reconhecimento a direitos legitimantes do indivíduo em estruturas de poder, como a política, que na hodiernidade necessita de uma releitura a partir do deslocamento do poder nessas relações algorítmicas às figuras privadas – conservando o dever de proteção jurídica do e através do Estado.

O tom aqui remonta a Hobbes (2002, p. 25-28), para quem a concepção do estado absolutista era a de que a natureza humana era a da desordem, porquanto a máxima *homo homini lupus*, de modo que “a origem de todas as grandes e duradoras sociedades não provém da boa vontade recíproca que os homens tiveram uns para com os outros, mas do medo recíproco que uns tinham dos outros”.

Antes de invocar Hobbes, convém delimitar o papel do contratualismo neste raciocínio: trata-se de um recurso heurístico – e apenas isso – para iluminar três questões que a era algorítmica recoloca sob novas formas: **i)** a origem da autoridade que passa a ordenar comportamentos quando decisões se concentram em infraestruturas privadas de grande escala; **ii)** o papel do medo e da assimetria (agora informacional) como motor de conformidade acrítica, o que ativa o dever de proteção estatal; e **iii)** a necessidade de limites, publicidade de razões e auditabilidade como condições de legitimidade. A analogia não equipara *Big Techs* a soberania nacional; ela apenas oferece um quadro conceitual mínimo para mostrar por que a transferência de poder decisório para sistemas opacos exige contrapesos jurídicos – transparência,

explicabilidade proporcional – antes de se aceitar o “Leviatã” digital como dado do mundo. Assim, no célebre O Leviatã, Thomas Hobbes (2008, p. 147) argumenta que:

> a única maneira de instituir um tal poder comum, capaz de os defender das invasões dos estrangeiros e dos danos uns dos outros, garantindo-lhes assim uma segurança suficiente para que, mediante o seu próprio labor e graças aos frutos da terra, possam alimentar-se e viver satisfeitos, é conferir toda a sua força e poder a um homem, ou a uma assembleia de homens, que possa reduzir todas as vontades, por pluralidade de votos, a uma só vontade. Isso equivale a dizer: designar um homem ou uma assembleia de homens como portador de suas pessoas, admitindo-se e reconhecendo-se cada um como autor de todos os atos que aquele que assim é portador de sua pessoa possa praticar ou levar a praticar, em tudo o que disser respeito à paz e à segurança comuns; todos submetendo desse modo as suas vontades à vontade dele, e as suas decisões à sua decisão. [...]. Feito isso, à multidão assim unida numa só pessoa se chama *República*, em latim *Civitas*, é esta a geração daquele grande Leviatã, ou antes (para falar em termos mais reverentes), daquele *deus mortal*, ao qual devemos, abaixo do *Deus imortal*, a nossa paz e defesa.

Deste modo, *auctorictas, non veritas facit legem* (Hobbes, 2008, p. 230). Estar-se-ia vivendo, agora, em um *potestas economicus datum, non veritas facit legem*?, isto é, o poder econômico do dado, e não a verdade, é que faz a lei/o direito? A comparação Estado absolutista com o poder exercido pelas *Big Techs* e demais companhias de tecnologia pode até parecer, para alguns leitores, algo totalmente ilógico e despreocupante. Quem decide deve justificar e quem mais sabe deve provar. A jurisdição constitucional não substitui o gestor tecnológico, mas exige razões públicas verificáveis e calibra a abertura do segredo técnico pelo risco e pela proteção aos grupos vulneráveis. Onde há opacidade técnica e assimetria informacional, especialmente com potencial de impacto díspar, impõe-se controle denso e proporcional de racionalidade a partir da transparência.

A produção datológica a partir da navegação do usuário na rede está a todo o momento, incluindo esse exato momento, sendo coletada, armazenada e pronta para ser utilizada, gerando valor econômico para empresas que vendem tais dados, ou meramente os utilizam, e para auxiliar outras empresas que, ao adquirir esses dados, poderão ter um real controle de influência sobre o comportamento daquele ser-alvo. Assim, uma notícia impactante: o e-mail da promoção do livro, da geladeira, da camiseta ou da bicicleta que talvez nem se tenha chegado a buscar em provedores de pesquisa como o da Google, não chegou por acaso; e não chegou por acaso naquele exato minuto em que se sabe que a caixa de e-mail individual do caro leitor é acessada com maior frequência.

Os dados possuem um grande poder, sendo o centro da economia mundial hoje, tendo seu valor independente de qualquer limitação. Os dados são abundantes, constantemente gerados e disponíveis. Existe uma relação contínua: onde há pessoas, existe economia, consumo e produção de dados — gerando uma economia fundamentada neles. Nesse contexto, decisões automatizadas utilizam os dados como principal recurso para funcionar (Hoffmann-Riem, 2022, p. 12).

Dessa forma, ao desvincular o indivíduo de seu próprio processo de autodeterminação, observa-se que, no contexto atual da vida digital o ser humano passa a ser tratado como objeto de sua própria autodeterminação. De maneira paradigmática, ele é impactado por influências personalizadas promovidas pelas redes, que inferem suas preferências, desejos e comportamentos, configurando um novo paradigma relacionado à compreensão da consciência.

Isso não lembra muito a Rousseau (2017, p. 23-25), segundo o qual todo homem nasce livre no estado natural, sendo a natureza humana boa, pacífica e próspera. A questão está na falta de autoridade central que concentre o poder da vontade e das possibilidades. Eis a fundamentação da celebração do contrato social, para que os homens possam ter segurança contra os inconvenientes da natureza e contra a arbitrariedade de homens mal-intencionados. Hoje é necessário um (re)pacto 2.0, ou melhor, 4.0. Hoje o contrato social não servirá para a proteção da natureza, que era a fonte emissora do risco, mas deverá ser das *Big Techs* e demais companhias, que passam a ter a centralidade sob o comando dos rumos de vidas e nações.

A liberdade exige limites públicos a poderes privados com capacidade de direcionar condutas – e, no ecossistema algorítmico, isso significa um pacto renovado que vincule plataformas a deveres de justificação e controle. Não se trata de estatizar a tecnologia, mas de constitucionalizar sua governança: quem estrutura escolhas deve submeter-se a razões públicas. Diante de opacidade e assimetria informacional, o Estado deve operar para evitar que isso se revele em uma assimetria de direitos, sobretudo contra grupos em situação de vulnerabilidade. Esse é o conteúdo do pacto: autonomia com transparência, inovação com responsabilidade, poder com controle.

Veja-se, na sequência dessa espécie de reconstrução, tem-se Kant (2017, p. 73-93), que afirmava serem os homens livres no domínio da razão prática, isto é, os seres humanos são dotados de liberdade e consciência moral, sendo ele, o ser

humano, o “ente” habilitado a realizar distinções entre as muitas possibilidades de seu desenvolvimento, isto é, da promoção de sua vida guiada pelo exercício de uma racionalidade livre, de um livre-arbítrio, livre de arbítrios que advinham de fora do ser, impostos de modo escrachado ou sutilmente. A liberdade, para Kant, é ser livre para optar mesmo em contradição a seus próprios impulsos. Fazer aquilo que não se quer, ou que não se tem o impulso a realizar, é a liberdade em si e de si que Kant percebe. Com Kant o conhecimento humano é reduzido à exercício da razão instrumental.

Em uma relativização da sequência histórica da análise, mas por conveniência temática, cabe ressaltar as noções trabalhadas por Rawls quanto ao seu ataque ao relativismo moral e ao utilitarismo, segundo o qual a máxima felicidade alcançada por um maior número de pessoas – lógica utilitária – é inadequada com uma concepção jurídica fundada na liberdade e, em específico, em ideias de justiça associadas a, dentro outros, *self-governing* (Rawls, 1982, p. 166-167). Aqui caracteriza-se, para dentro da tônica de Rawls – no transporte semântico proposto – a lógica utilitária das redes e desenvolvedoras de exclusão, ou limitação, da participação dos seres lesados ou interessados a participarem da construção dialógica e diacrônica dos ajustes e verificações de incidentes das aplicações algorítmicas; ganhando ainda maior relevo se analisado frente às possibilidades de criação de políticas públicas de inclusão digital, ou ainda de controle público-social, também para as aplicações. A fórmula da incompatibilidade do monismo utilitarista indica a impossibilidade de justificação de fins por meios – como a punição/exclusão de pessoas inocentes por uma vantagem para um alegado bem aos demais membros da sociedade – com seu pensamento filosófico-liberal, tampouco com sua teoria da justiça como equidade. É que, para Rawls (Johnston, 2018, p. 243-244), a estrutura básica da sociedade – no campo da vida prática – deve atender, no que for diretamente relacionado com o Estado, ou outra forma de ficção do governo, estritamente ao que reclama a justiça, sendo restritivo – pela noção de justiça fundada também na liberdade – quanto ao estabelecimento de privilégios e, no que aqui importa, privações. Ainda reconhece, Rawls, que a esse domínio da estrutura básica da sociedade não importa a regulação quanto à relação de privados e as ações individuais que não signifiquem impacto real em outros.

Com base nesses aportes de Rawls, verifica-se que, também aqui, satisfaz a necessidade de (des)horizontalização da relação entre desenvolvedoras e seus

usuários, para validar uma verdadeira e faticamente verificável estrutura vertical de poder e influência. O conteúdo da igual liberdade, a ser alcançada materialmente entre seres – sejam de pessoas naturais entre si, ou destas para com pessoas jurídicas, ou ainda entre essas últimas, no limite da possibilidade –, pressupõe um ajuste acerca do conteúdo das concepções morais de uma ética mínima compartilhada entre as partes consideradas. O método de aferição desse liame subjetivo mínimo compartilhado é o Equilíbrio Reflexivo ("ER"), que propõe – sumariamente – a identificação da objetividade moral sincronizada e, se necessário, pela inferência dos elementos morais e não morais a partir da coerência entre conceitos, crenças e objeções particulares. Nos dizeres de Rawls (1971, p. 46-47), é a partir dos resultados consistentes do ER que se está apto a determinar a justificação de práticas ou condutas a partir de um dado ordenamento jurídico. Assim, no tema em apreço, será apenas a partir da identificação de critérios mínimos – e globais no atual estágio da globalização tecnológica – a identificar o grau e o nível de exigência que o direito poderá permitir que particulares imponham às *Big Techs* e demais desenvolvedoras que estejam aquém do mínimo necessário como, por exemplo, o respeito a grupos vulneráveis e o compromisso com a erradicação de toda a forma de preconceito. Preceitos éticos identificados a partir do antifundacionalismo[9] do ER que permitem a construção de deveres pelo direito, seja pelo legislador ou pela jurisdição.

No domínio algorítmico, mantem-se a máxima analógica: quem decide deve justificar, e quem mais sabe deve provar. A parte que controla dados, modelo e aplicações decisórias deve carregar o ônus informacional correspondente ao risco e ao impacto, ainda que nunca sejam requisitados. A transparência deve ser um princípio (quase-moral) que constrange empresas a desenvolverem a explicabilidade *by design*, e controladores e usuários a terem um uso ético da IA.

Retornando à lógica da evolução histórica gradual, o constitucionalismo amadurecido pela Guerra de Independência Norte Americana, de 1776, e pela Revolução Francesa, de 1789, se deu a partir de uma noção de liberdade enquanto uma nova visão de vida e relação jurídico-econômica do ser humano para com o Estado (Grimm, 2016, p. 41-42). Hoje se necessita de uma nova relação liberal para que se estabeleça o entendimento normativo de defesa para com a tecnologia? Assim como o Estado pode ser um grande aliado da nação e do povo, também o pode a

[9] Para a diferença entre as correntes do coerentismo e do fundacionalismo, ver: Paulo, 2022.

tecnologia, mas não desprovida de obrigações, limites e deveres instrumentais. A questão, portanto, é da junção entre Estado, Mercado (tecnológico) e Sociedade Civil, mas de um modo não baseado no medo. Veja-se, o medo é avesso à condição conceitual de tratarmos de liberdade:

> Apenas aqueles que conhecem a liberdade em relação à necessidade podem apreciar por completo o significado da liberdade em relação ao medo, e só aqueles que estão livres de ambos – necessidade e medo – têm condições de conceber uma paixão pela liberdade pública (Arendt, 2019).

Para Arendt, o medo ou a ignorância fundamentam a existência de uma massa que, ao lado da propagando e do uso do terror, formam a receita completa para o surgimento e o funcionamento de regimes autocráticos totalitários. Não se está advogando a existência de um governo central da internet, longe disso. A analogia passa por compreender como que, do mesmo modo que as massas funcionaram para legitimar regimes nas décadas de 30 e 40 do século passado, também a ignorância e o comodismo tem o condão de fazer algo muito semelhante no atual tempo: legitimação de um descontrole e uma desregulamentação da aplicação diária e cotidiana das novas tecnologias. O ponto de Arendt (2012, p. 439), como se percebe, é, também, da indiferença, para além da ignorância e do medo:

> [...] lidamos com pessoas que, simplesmente devido ao seu número, ou à sua indiferença, ou a uma mistura de ambos, não se podem integrar numa organização baseada no interesse comum, seja partido político, organização profissional ou sindicato de trabalhadores. Potencialmente, as massas existem em qualquer país e constituem a maioria das pessoas neutras e politicamente indiferentes, que nunca se filiam a um partido e raramente exercem o poder de voto (Arendt, 2012, p. 439).

Na Alemanha do início da década de 1930, pode-se perceber isso. Nutridos pelo revanchismo, não se opuseram à tomada de poder pelo Partido Nacional-Socialista, que, quando não foi eleito, em abril de 1932, obtendo apenas 36,8% dos votos, realizaram golpes institucionais e crimes políticos para que chegassem ao poder em 1933, sendo referendado como chanceler com 90% dos votos no plebiscito realizado em agosto de 1933, com o Auto de Autorização (*Gesetz zur Behebung der Not von Volk und Reich*). Tal Auto permitia, por uma brecha da constituição de Weimar, que o chanceler pudesse elaborar leis à margem da Constituição. Assim, com base nessa legislação, o *Reich* com Hitler no comando iniciou uma expansão aos

territórios vizinhos, uma forte e monstruosa política de segregação, e posteriormente eliminação, com os campos de concentração, a Regulamentação 11 da Lei da Cidadania do Reich – que retirou a cidadania dos judeus – e com a criação da Gestapo, tudo pelo *Gemeinnutz von Eigennutz* (o bem comum acima de todos!) (Shirer, 2017, p. 168-315).

A chave para o “sucesso alemão", se não uma, passa pelas três comentadas acima com aporte no pensamento de Arendt. Assim, parece evidente que em nada se pode deixar cair em obnubilação e em tudo deve-se buscar a luz, a verdade e a transparência. É o ato de lançar luzes sobre a decisão automatizada, exigindo-lhe explicabilidade que se poderá conformar a realidade estruturada a partir dos avanços tecnológicos – retoricamente sempre ótimos para o desenvolvimento humano – com um tom mais crítico e exigente, em um timbre que eleve os direitos humanos e fundamentais para o seu necessário papel de destaque frente aos desatinos da atual realidade algorítmica.

Que fique, pois, cristalino: a travessia do Estado de natureza aos pactos sociais não se encerra nos marcos da filosofia política clássica, mas ressurge, renovada, no embate cotidiano entre sujeitos humanos e os novos titãs digitais. Se Rousseau clamava pela liberdade frente à arbitrariedade, se Kant vislumbrava a autonomia do ser racional e Rawls erigia balizas de justiça contra o utilitarismo excludente, hoje ergue-se a demanda por um novo pacto – agora, entre pessoas, máquinas e corporações, num terreno onde a ignorância e a indiferença pavimentam o risco do retrocesso. A centralidade humana, enquanto um movimento cultural que exige racionalidade transparente pública, é essencial como guia, princípio e vetor normativo do atual momento.

O desafio não é menor: iluminar, com o rigor da razão pública e da ética compartilhada, as engrenagens opacas dos algoritmos, exigindo explicabilidade, transparência e respeito à dignidade de cada indivíduo. Que não se repita a história de massas conduzidas pelo medo ou pelo comodismo, silenciando a liberdade diante do avanço irresistível da tecnologia. É tempo de reivindicar, para a era digital, o mesmo fervor humanista que outrora impulsionou revoluções e emancipações.

Assim, ao tocar o limiar do nosso tempo, só nos resta assumir o protagonismo na escrita desse novo contrato social. Não pela dominação ou pela resignação, mas pelo compromisso inarredável com a liberdade, a justiça e a verdade – que possam

servir, na aurora da sociedade algorítmica, como faróis a guiar a travessia de nosso destino comum.

O fio comum desses marcos converge em comandos operativos para a era algorítmica. No plano técnico-jurídico, isso se traduz em explicabilidade suficiente, avaliações de impacto, inversão dinâmica do ônus informacional e abertura graduada do segredo algorítmico. Com isso, a jurisdição constitucional deve assumir um padrão de escrutínio calibrado por risco e vulnerabilidade: quanto maior a opacidade técnica e a assimetria informacional, e quanto mais intensa a chance de reprodução de discriminações estruturais, mais denso deve ser o controle de racionalidade, necessidade e proporcionalidade da solução tecnológica adotada (de matriz alexyana). Em termos operacionais, isso se concretiza por: **i)** inversão dinâmica do ônus informacional quando o decisor se vale de sistemas opacos ou de difícil auditabilidade; **ii)** ordens estruturais e aberturas graduais – i.e., *disclosure* qualificado sob sigilo judicial, auditoria independente, logs e trilhas de decisão, obrigação de supervisão e revisão humana significativa e de implementação de salvaguardas by design; **iii)** Relatório de Impacto de Proteção de Dados (“RIPD”)/Avaliação de Impacto Algorítmico (“AIA”) como condição procedimental de legitimidade em cenários de alto risco; e **iv)** monitoramento com métricas de não discriminação ao longo do ciclo de vida do sistema.

A documentação técnica deverá observar cadeia de custódia digital, com *hashing* (conteúdo, ou valor, digital) dos documentos relevantes, trilhas de acesso e controle de versões. O fornecimento ocorrerá sob proteção de sigilo judicial e acordos de confidencialidade, assegurando contraditório efetivo sem comprometer indevidamente a concorrência. O segredo algorítmico e a concorrência permanecem protegidos, mas cedem na estrita medida necessária para viabilizar o contraditório e a verificação de vieses, com acesso controlado, sob sigilo judicial e acordos de confidencialidade.

Há que se ter cautela com o ônus argumentativo em cenário algorítmico. Quem decide por meio de sistemas automatizados cria para si um dever de justificar. Isso significa que a autoridade pública que decide com apoio de IA deve não só identificar sua decisão como fruto de uso de aplicação algorítmica, como deve expor às razões da decisão, o controlador privado que decide efeitos automáticos (crédito, emprego, preços, ranqueamento) deve, ao menos, ter a capacidade de explicar a

decisão, quando solicitado, e, sobretudo o fornecedor tecnológico, que detém informações exclusivas, assume um dever importante de publicização das razões verificáveis na medida do risco e do impacto que sua aplicação poderá promover. A regra deve ser clara, onde houver assimetria informacional sobre dados, parâmetros ou trilhas decisórias, os ônus probatórios e de explicação devem ser alocados para quem opera ou fornece o sistema.

Nessa linha, Gavião Filho (2025) analisa que a repartição do ônus da prova acompanha os subtestes da proporcionalidade. Na adequação, incumbe ao operador demonstrar, com base em auditoria técnica, a relação causal entre o critério utilizado e o fim legítimo — acesso a modelo, parâmetros e testes é *conditio sine qua non* (Wachter; Mittelstadt; Floridi, 2017). Na necessidade, cabe ao mesmo operador provar a inexistência de alternativas igualmente eficazes e menos gravosas (simulações, testes comparativos), justamente por deter o domínio técnico (art. 8º da Diretiva 2000/43/CE[10]). Na proporcionalidade em sentido estrito, adota-se um modelo bifásico: ao reclamante basta evidenciar, *prima facie*, disparidade estatística; verificada a diferença, desloca-se ao operador o encargo de justificar que os benefícios (acurácia, eficiência, custo) não superam os custos discriminatórios — solução consonante com experiências comparadas (Griggs v. Duke Power Co. (1971)[11]; *CHEZ Razpredelenie Bulgaria AD v. Komisia za zashtita ot diskriminatsia* [Comissão de Proteção contra a Discriminação] - C-83/14, TJUE, 2015-[12]). Desse modo, ao reclamante incumbe

[10] Artigo da Diretiva do Conselho da União Europeia que prevê aos processos judiciais ou administrativos, em que se alegue discriminação baseada na raça ou origem étnica, que "cabe ao demandado provar que não houve violação do princípio da igualdade de tratamento". Trata-se de uma hipótese de inversão do ônus da prova, justificado a partir da efetividade da tutela contra a discriminação. Foi consolidado pelo Tribunal de Justiça da União Europeia em casos como *Feryn* (C-54/07, 2008) e *Meister* (C-415/10, 2012), em que se reforçou que basta ao reclamante indicar elementos que sugiram um impacto desigual para que a parte demandada assuma o dever de demonstrar a legitimidade e a neutralidade de sua conduta (Gavião Filho, 2025).

[11] Decisão da Suprema Corte norte-americana que introduziu a noção de discriminação indireta ao afirmar que requisitos aparentemente neutros, como testes educacionais ou de aptidão, podem produzir impactos discriminatórios ilegítimos quando resultam, de forma desproporcional, na exclusão de determinados grupos, independentemente de intenção discriminatória explícita. Nesse caso, a Suprema Corte entendeu que o critério do nível de escolaridade e teste de inteligência, aparentemente neutros, usados para seleção, contratação e promoções de funcionários, escondiam discriminação não intencional, ante as evidências de que os negros eram em sua maioria reprovados na seleção ou não promovidos (Gavião Filho, 2025).

[12] Caso em que o Tribunal reconheceu a discriminação indireta decorrente da instalação de medidores de energia em locais de difícil acesso em bairros habitados majoritariamente por pessoas de etnia cigana. Os medidores de eletricidade foram instalados em postes a uma altura de seis a sete metros, tornando-os inacessíveis para muitos residentes. Essa prática foi justificada pela empresa como uma medida para evitar o roubo e a manipulação de medidores de energia elétrica, que eles alegaram ser predominantes na área. O Tribunal identificou essa prática como um caso de discriminação indireta,

apenas demonstrar prima facie a existência de impacto desigual, recorrendo a indicadores estatísticos como o *disparate impact ratio* ou o *equal opportunity difference*. Uma vez verificada essa disparidade, transfere-se ao operador ou desenvolvedor do sistema a carga de justificar que os benefícios obtidos, em termos de acurácia, eficiência ou redução de custos, não são desproporcionais frente ao ônus discriminatório imposto (Gavião Filho, 2025).

A inversão dinâmica do ônus informacional incide quando houver, cumulativa ou alternativamente: **i)** assimetria material de acesso a dados, parâmetros e trilhas decisórias; **ii)** indícios razoáveis de impacto desproporcional sobre grupos em situação de vulnerabilidade; **iii)** opacidade técnica não superável por meios ordinários de prova; **iv)** recusa injustificada do operador em fornecer informações essenciais ao contraditório. Nessas hipóteses, compete ao responsável pelo sistema demonstrar a adequação do desenho, a qualidade dos dados e a ausência de discriminação injustificada – sempre em grau proporcional ao risco.

Para fins probatórios, a discriminação será aferida em três chaves: direta (critério explícito), indireta (*proxies* e variáveis correlatas) e por impacto adverso (resultados desproporcionais sem justificativa adequada). A verificação técnica utilizará métricas compatíveis com o bem jurídico. Persistindo indícios robustos, incide a inversão dinâmica do ônus informacional e as medidas de abertura graduada e relativização do segredo algorítmico.

O segredo algorítmico e a proteção concorrencial permanecem resguardados, mas cedem na exata medida necessária para possibilitar contraditório efetivo, verificação de vieses e ponderação transparente – inclusive mediante perícia assistida e acesso controlado às informações sensíveis. A jurisdição não substitui o gestor tecnológico, mas impõe razões públicas verificáveis, condiciona o uso de Inteligência Artificial a explicabilidade suficiente, o que garante remédios orientados a resultados quando houver impacto desproporcional sobre grupos vulneráveis.

Um esforço adicional para tanto é a intensificação de um modelo interpretativo acerca do lócus em que empresas de tecnologia, sobretudo se *big techs*, ocupam nesse desenho jusfundamental. A esses atores é o caso de repensar se a eficácia

---

anotando que a incapacidade de acessar medidores de eletricidade não afeta apenas a capacidade dos residentes de controlar seu consumo, mas também contribui para um efeito estigmatizante na comunidade cigana, levantando questões sobre a proporcionalidade e justiça das medidas implementadas.

horizontal é suficiente. Verifica-se uma nova eficácia, “verticalizada”, com efeitos diretos de direitos fundamentais sobre controladores com poder estrutural dominante no mundo digital (Celeste, 2024, p. 42). Quando tais agentes deixam de ser meros ofertantes e passam a operar como infraestruturas de intermediação – controlando acesso, dados e arquitetura de escolhas –, a incidência dos direitos fundamentais adquire feição objetiva, com deveres positivos correlatos. Nesses contextos, a igualdade material e a autodeterminação informativa não se exaurem na bilateralidade privada: reclamam padrões mínimos de prevenção de vieses, documentação do processo decisório, supervisão humana significativa, explicabilidade proporcional ao risco e trilhas auditáveis, sob ângulos de abertura graduados por proporcionalidade. Não se trata de transpor a lógica público-estatal ao setor privado, mas de reconhecer que determinadas funções – *gatekeeping* informacional, dominância de mercado de dados, centralidade de plataformas e domínio sobre a existência cibernético-automatizada – reconfiguram a medida de tutela exigível pela Constituição, visto que reconfiguram, ou talvez desconfigurem verdadeiramente, a arquitetura social e mercadológica, imprevista até então.

Essa leitura não nasce isolada. Experiências normativas e regulatórias contemporâneas vêm convergindo para arranjos de governança por risco, avaliações de impacto, deveres de transparência e mecanismos de supervisão e controle com participação de autoridades independentes – justamente para compatibilizar proteção concorrencial e proteção de direitos em ambientes tecnologicamente opacos. É imperioso verificar, portanto, em que medida os arranjos normativos e regulatórios ensaiados em outras jurisdições – avaliações por risco, deveres de explicabilidade e documentação, supervisão humana significativa e mecanismos de auditoria – podem dialogar com a moldura doméstica, orientando a interpretação constitucional e a execução prática da tutela antidiscriminatória que ora se propõe.

4.3 Sentido normativo e diálogos institucionais: uma proposta de busca de sentido da tarefa interpretativa com as experiências nacionais e internacionais acerca do tratamento jurídico da discriminação algorítmica

A tarefa interpretativa reclama uma arquitetura de diálogos entre jurisdição, legislador/regulador e órgãos de controle, na qual o sentido normativo da tutela

antidiscriminatória é coproduzido e depois internalizado em padrões operacionais. Parte-se de três marcos já presentes no trabalho: **i)** o núcleo constitucional de igualdade material e de proteção contra discriminações; **ii)** a gramática LGPD de prevenção, responsabilização e revisão de decisões automatizadas; e **iii)** as experiências normativas e regulatórias comparadas que convergem para exigências de supervisão humana, documentação do processo decisório e explicabilidade proporcional ao risco. O método é incremental e prudencial: os tribunais extraem diretrizes mínimas a partir desse triângulo normativo e, em seguida, acionam o circuito institucional (ANPD, PROCONs, Poder Judiciário, Ministério Público, etc.) para refinar salvaguardas e padronizar a interpretação do sentido normativo do direito em casos semelhantes. Para tanto, precisa se verificar a suficiência da regulação, o sentido das tentativas regulatórias e o esforço do direito diante do novo fenômeno. E, para isso, ainda, necessário verificar a situação atual.

A Internet hoje é o grande meio de existência do ser humano, estruturante de sociabilidade, de circulação de mercadorias, informações e até de fruição de direitos. É o espaço não-físico da plataforma de relevante percentual do relacionamento humano. É local com espaço para riscos institucionais que o Direito não pode mais ver como periferia. É na Internet que se negociam bens e serviços, que se encontram pessoas, profissionais e empresas, que se conhece até mesmo o verdadeiro amor. É na internet que a vida, de fato, vem acontecendo; isto é, sem a internet, a vida do lado de fora das telas seria completamente diferente. Esse protagonismo da Internet ainda ganhou um tempero maior em 2020, com a decretação da pandemia por conta da circulação do vírus Sars-Cov-2, e, com isso, o início das medidas de isolamento social.

Assim, mais do que nunca, a sociedade civil se comunicou pela Internet, participando de aulas de instituições públicas e privadas por plataformas como Google Meet e Zoom, vendo transmissões pelo Instagram, ou ainda pelo Youtube e Facebook, reunindo-se para reuniões de trabalho. Tudo isso dependente da criação de perfis com informações pessoais, como nome, e-mail, telefone, por vezes até cartão de crédito, nome da mãe, do pai, se tem filhos, qual a profissão e a faixa de remuneração etc. Tomemos como exemplo extremo o número de dados que as pessoas informam na rede quando se cadastram em *e-bankings* ou quando preenchem cadastros de seleção de vagas de emprego. Até mesmo o Poder Judiciário, resistente acerca da expansão das audiências e instrução para o meio digital migrou completamente suas

operações para o meio digital, vindo a exigir como regra geral a presencialidade apenas em 2023.

Dessa forma, nota-se uma complexização da vida. E, como sempre, o direito deve acompanhar. A mudança, contudo, e positivamente, veio dois anos antes desse novo boom no protagonismo da Internet na nova realidade que impôs a pandemia, e antes do segundo grande boom da década, o avanço e popularização do acesso às ferramentas de Inteligência Artificial. Naturalmente, o fluxo intenso de informações de dados pessoais não começou em 2020. Com efeito, a preocupação acerca da proteção desses dados pessoais, que já era existente, ganhou uma importância a mais e foi acelerada com o episódio da Cambridge Analytica ("CA"). O episódio é o fato que expôs os efeitos da perfilização e do micro-direcionamento ao mundo.

Ao que se sabe, e o que uma parte ficou demonstrado no documentário *The Great Hack*, do Netflix, a CA era uma empresa britânica de análise de dados, contratada, dentre outros, pelo então candidato à presidência norte-americana Donald Trump, para influenciar eleitores indecisos em 2016. A estratégia era a de avaliar os dados existentes em redes sociais, como o Facebook, montando perfis de eleitores, os influenciando através do *micromarketing* e *microtargeting* dessas plataformas através de estratégias de postagens que cada perfil receberia, conquistando eleitores indecisos pelo perfil, ou fazendo com que opositores fracos não votassem. A mesma empresa teria atuado no *British Exit from EU* (BREXIT). Tudo isso possibilitado e potencializado pelo *Big Data* (Nicolelis, 2020, p. 153). Na máxima Orwelliana (2009):

> Noite e dia as teletelas massacravam os ouvidos das pessoas com estatísticas que provavam que hoje a população tinha mais comida, mais roupa, melhores casas, melhores opções de lazer – que vivia mais, trabalhava menos, era mais alta, mais saudável, mais forte, mais feliz, mais inteligente, mais culta do que as pessoas de cinquenta anos antes (ORWELL, 2009, p. 226).

Assim, uma nova preocupação com a defesa das pessoas em rede, usuários de produtos e serviços na web, deu início à corrida legal por mecanismos de proteção normativa. Bem na verdade, a tutela era a do consumidor que tem uma vulnerabilidade expandida e acentuada em rede digital. Além disso, a proteção paradigmática do consumidor-usuário, que de sujeito de consumo, passou a sujeito fornecedor de dados a título semigratuito, precisava ser fortemente recompassada. Essa preocupação faz a UE aprimorar a normativa anterior com a *General Data Protection Regulation*

("GDPR"). Dois anos mais tarde, com inspiração na GDPR, o Brasil promulga a Lei Geral de Proteção de Dados ("LGPD"). O escândalo da CA, que também envolveu, como se sabe, o Facebook, não foi a gênesis da preocupação com a proteção de dados, mas foi o combustível que faltava para o impulso final, ainda que não definitivo, desta proteção.

Até 2018, ano marcado pelo ato legiferante da LGPD, a proteção de dados era tratada apenas reflexamente pelo ordenamento jurídico pátrio, por leis esparsas que não garantiam a tutela adequada desses novos *digital assets*, porquanto não eram legislações centradas nisso. Em linhas gerais, que serão densificadas posteriormente, o Brasil contava com proteção de dados em diversas leis, destacando-se o Código de Defesa do Consumidor, a Lei do Habeas Data, a Lei do Cadastro positivo, a Lei de Acesso à Informação e o Marco Civil da Internet.

Em termos de atualização normativa, pode-se dizer que o Brasil está atrás de países latino-americanos como o Chile, que tem legislação sobre proteção de dados pessoais desde 1999, a Argentina, desde o ano 2000, o Uruguai, desde o ano 2008, o México, desde 2010, o Peru, desde 2011, e a Colômbia, desde 2012. Além disso, a primeira normativa que tratou do tema data ainda do ano de 1970, quando a Alemanha, através da Lei do Estado de Hesse, começou a tutelar os direitos individuais frente ao tratamento de dados pessoais realizado pelo poder público e reconhecia, dentre outras disposições, um conjunto de princípios essenciais para nortear o tratamento de dados pessoais (Mendes; Bioni, 2019, p. 164).

Esse paradigma, que se pode chamar de paradigma alemão da proteção de dados, o qual tutelava relações eminentemente verticais (Estado-cidadão), serviu de base para o movimento internacional seguinte: os *Fair Information Practice Principles* (FIPPs). Os FIPPs são um conjunto de princípios desenvolvidos para servir de base para o tratamento de dados pessoais por todos os países. Os princípios dos FIPPs são a transparência, participação individual, especificação de propósito para o tratamento dos dados pessoais, minimização de dados pessoais utilizados, limitação de uso dos dados, controle de qualidade dos dados, integridade da atividade, segurança e uma dupla possibilidade-responsabilidade de que o tratamento seja auditável, e que se conduza o processo tendo-se em vista a necessidade de *accountability* da atividade. Nos EUA, esses princípios foram a base do *Privacy Act* de 1974. Também merece destaque a lei nacional de proteção de dados sueca, que

foi o Estatuto para bancos de dados de 1973 (*Data Legen* ou *Datalag*) (Mendes; Bioni, 2019, p. 162-164).

Afora essa questão, a partir da Lei do Estado de Hesse, foram confeccionadas diversas normas protetivas no mundo. Segundo Martins (2014, p. 231), tais normas podem ser classificadas em gerações, de acordo com suas dimensões internacionais. A primeira geração é a da gênese da década de 70, e tinha por objetivo a regulação do processo eletrônico de dados pessoais pelas administrações públicas, e algumas empresas privadas, e ganhou destaque, para além da Lei de Hesse, com Estatuto de Proteção de Dados do Estado Alemão e Lei Federal de Proteção de Dados da Alemanha. Posteriormente, a segunda geração diz respeito à preocupação com o consentimento e a liberdade de escolha do cidadão. A terceira geração, por sua vez, cuidou da questão relativa à autodeterminação informativa*,* pela qual o titular controla individualmente o processamento de seus dados pessoais. E, finalmente, a quarta geração se preocupou em melhorar o controle dos dados fornecidos pelo próprio titular, bem como "estabelecer temas de dados pessoais intangíveis a qualquer acesso e, por fim, partir de metodologia de formatação de normas gerais com complementação por normas setoriais".

Ainda em linhas históricas, há de se destacar o avanço que teve a proteção de dados pessoais enquanto um direito constitucional reconhecido na Constituição de Portugal, de 1976, no pós-Revolução dos Cravos (Limberger, 2008, p. 141). E, igualmente, a Constituição da Espanha, de 1978, trouxe a ressalva que com o avanço da tecnologia, e a necessidade de proteção dos direitos fundamentais, como os dados pessoais, a constituição hispânica asseverou que o Estado deveria limitar o uso da informática para garantir a honra e a intimidade pessoal e familiar dos cidadãos, bem como garantir o pleno gozo e exercício dos direitos.

Na Alemanha, embora já houvessem dispositivos constitucionais que tutelassem a privacidade e o livre desenvolvimento, o direito fundamental à proteção individual ao uso indiscriminado dos dados pessoais por terceiras pessoas, ou pelo Estado, veio através do caso BVERFGE 65, 1, 1983, conhecido como o caso da Lei do Censo Alemão (*Volkszählungsgesetz*), no qual pode-se dizer que houve o nascedouro forte da proteção de dados como um compromisso dos Estados nacionais, com assento de garantia constitucional, e, portanto, com status de direito fundamental. O caso discutia a constitucionalidade da lei que determinava o

recenseamento geral da população, com o levantamento de dados sobre a profissão, moradia e local de trabalho para fins estatísticos, e foi levado à Corte máxima alemã por conta do §9º, que previa, dentre outras questões, a possibilidade de uma comparação dos dados levantados com os registros públicos e também a transmissão de dados tornados anônimos a repartições públicas federais, estaduais e municipais para determinados fins de execução administrativa (Schwabe, 2005, p. 233-234).

As alegações nas reclamações constitucionais recebidas pelo Tribunal Constitucional Alemão (*Bundesverfassungsgericht*) (“BVerfG”) eram de violação direta a alguns direitos individuais da Lei Fundamental da Alemanha (*Grundgesetz*) (“GG”), sobretudo ao direito ao livre desenvolvimento da personalidade (art. 2, I, da GG). Ao final, no mérito, o BVerfG julgou as reclamações no sentido de confirmar a constitucionalidade da Lei do Censo, contudo declarando nulos os dispositivos sobre a comparação e trocas de dados e sobre a competência de transmissão de dados para fins de execução administrativa. Assim, o que o BVerfG asseverou na oportunidade foi o direito individual contra o levantamento, armazenagem, uso e transmissão irrestrito de dados pessoais, porquanto o indivíduo é credor de um direito geral de personalidade (art. 2, I, e 11, ambos da GG) e, por isso, tem o poder de decidir sobre a exibição de seus dados pessoais (Schwabe, 2005, p. 234).

Além disso, Schwabe (2005, p. 235) assevera que a partir desse caso ficou consignado no direito alemão que as restrições ao direito à autodeterminação sobre a informação pessoal são permitidas apenas em casos de interesse predominante da coletividade, e mesmo nesses é necessária a observância à existência de uma legislação específica para aquela finalidade, bem como base constitucional para a lei, e que seja observada a proporcionalidade, devendo ser tomadas todas as precauções organizacionais e processuais necessárias para que riscos de violação aos direitos de personalidade sejam evitados. O Tribunal ainda distinguiu os dados levantados anonimamente dos dados individuais, determinando que acerca dos dados levantados para propósitos estatísticos (anonimamente), não se pode exigir uma vinculação estrita e concreta de propósito dos dados. Ainda assim, dentro do sistema de informação devem existir barreiras para compensação, em contraposição ao levantamento e uso da informação.

Posteriormente, a União Europeia preocupou-se em tutelar de forma uniforme a matéria, tutelando adequadamente os dados pessoais dos indivíduos em sua macro

jurisdição, habilitando-os para exercerem suas liberdades e seus direitos da personalidade, assim como para facilitar a transferência internacional de dados. Para tanto, em 1995, o Parlamento Europeu e o Conselho Europeu lançaram a Diretiva nº 46, disciplinando o tratamento e a proteção de dados das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados no continente. Essa diretiva viria a ser substituída pela GDPR em 2016.

Na Espanha, a Diretiva nº 46/95 deu origem à Lei Orgânica de Proteção de Dados ("LOPD"). A LOPD ampliou a proteção ofertada pelo texto constitucional de 1978, trazendo inovações quanto ao âmbito de proteção legal, ampliando-o para que alcançasse todos os bancos de dados, informatizados ou não, bem como que houvesse proteção aos dados pessoais com o escopo de garantir liberdades públicas e direitos fundamentais individuais (Limberger, 2008, p. 145). Assim, a proteção de dados caminhava já para ser reconhecida, pelo menos em alguns ordenamentos, a tutela com assento constitucional alcançado.

Contudo, o ano de 2000 tem especial importância para o avançar transcontinental da proteção dos dados pessoais. Com efeito, é nesse ano que a Carta dos Direitos Fundamentais da União Europeia é organizada. No capítulo II – Liberdade, no art. 8º, consta que a União Europeia terá proteção de dados, sendo todas as pessoas credoras do direito à proteção dos dados que lhes digam respeito, que os dados devem ser objeto de um tratamento leal, para fins específicos e com o consentimento da pessoa interessada ou com outro fundamento legítimo previsto por lei, e que por isso todas os indivíduos têm o direito de acessar os dados que lhes digam respeito, obtendo a respectiva retificação. E, por último, que o cumprimento destas regras fica sujeito à fiscalização por parte de uma autoridade independente.

Em 2016 a Diretiva nº 46/95 foi substituída pelo Regulamento nº 679 de 2016, a GDPR, do Parlamento Europeu e do Conselho Europeu, entrando em vigor no ano de 2018. Assim, com o avanço da legislação europeia, bem como o reconhecimento da proteção dos dados pessoais, ou da autodeterminação dos dados pessoais, em nível continental, a matéria está sendo delineada e desenvolvida a partir das decisões do Tribunal de Justiça da União Europeia, que influenciam a matéria em nível mundial, e fortemente no Brasil.

No Brasil, em 2018 foi aprovada a LGPD, que entrou em vigor em 2020, com disposições valendo apenas a partir de agosto de 2021. A Lei traz disposições acerca

do tratamento de dados pessoais, inclusive nos meios digitais, por pessoa natural ou por pessoa jurídica de direito público ou privado, com o objetivo de proteger os direitos fundamentais de liberdade e de privacidade e o livre desenvolvimento da personalidade da pessoa natural. O dever de proteção (*Schutzpflicht*) aqui aparece e, com ele, podemos entender que o Brasil não estava desprotegido na matéria, visto que a partir de elementos normativos constitucionais e infraconstitucionais legislador e judiciário poderiam intervir para salvaguarda deste direito fundamental (Leal; Maas, 2020). Ainda assim, formalmente não havia proteção em assento constitucional até a EC nº 115/2022, oriunda da Proposta de Emenda à Constituição (PEC nº 17/03/2019), alterando o texto constitucional para que conste no artigo 5º, inciso LXXIX, que "é assegurado, nos termos da lei, o direito à proteção de dados pessoais, inclusive nos meios digitais", no artigo 21, inciso XXVI, que compete à União "organizar e fiscalizar a proteção e o tratamento de dados pessoais", bem como que o artigo 22, inciso XXX, passe a constar que compete privativamente à União legislar sobre "proteção e tratamento de dados pessoais".

Ainda assim, já era possível compreender que, desde antes, do art. 5º, *caput*, da Constituição, decorre a liberdade individual, do inciso X decorre a inviolabilidade da intimidade e da vida privada e do inciso XII decorre o livre desenvolvimento, a autodeterminação individual e, portanto, os dados pessoais – o ponto não é mais controverso; felizmente. Pode-se afirmar que seja pela alteração constitucional, seja pela soma interpretativo-construtivista que resultava no reconhecimento do direito implícito à inviolabilidade dos dados pessoais, vê-se que ela não é suficientemente adequada para tutelar integralmente os dados pessoais. É que a tutela de algo não depende unicamente da CF (Mendes, 2011, n.p.).

Como aludido, o Brasil já contava com legislação que resguardava os dados pessoais. Leis como o Código de Defesa do Consumidor, a Lei do Habeas Data, a Lei do Cadastro positivo, a Lei de Acesso à Informação, o próprio Código Civil e o Marco Civil da Internet formavam uma espécie de microssistema protetivo aos dados pessoais. Ainda que, como mostra Santos (2020, n.p.), a tentativa de formatação de uma lei de proteção aos bancos de dados remonte ao ano de 1999 no Brasil, apenas em 2018 teve-se a aprovação da LGPD.

O Código Civil, por exemplo, determinava que a vida privada da pessoa natural é inviolável, e o juiz, a requerimento do interessado, adotará as providências

necessárias para impedir ou fazer cessar ato contrário a esta norma, conforme inteligência do seu art. 21. Uma lei mais específica, a do Cadastro Positivo (nº 12.414/2011), já disciplinava a formação e consulta a bancos de dados com informações de pessoas naturais ou jurídicas, com o fito exclusivo de formar seu histórico de crédito. Essa lei trouxe questões importantes, como o princípio da qualidade dos dados (art. 3º) e os direitos de acesso, retificação e cancelamento de dados (art. 5º). Ainda, a lei delimita as finalidades para as quais os dados podem ser coletados (art. 7º). Por todas essas disposições, Mendes (2014) entende que a referida lei consolida a evolução de um conceito de autodeterminação informativa no ordenamento pátrio.

Pouco tempo depois, em 2012, a Lei Carolina Dickmann (nº 12.737) criminalizou as condutas de invadir dispositivo alheio com o fim de adulterar, obter ou destruir dados ou informações sem a autorização do titular dos dados, bem como, no presente escopo, criminaliza a conduta do desenvolvedor de software capaz de facilitar as práticas supra. Ainda, A Lei do Marco Civil da Internet (nº 12.965, de 2014) (“MCI”) prevê, em seu art. 3º, que o princípio da proteção dos dados pessoais integra o rol dos princípios que disciplinam o uso da Internet no Brasil, juntamente com o princípio da proteção da privacidade e outros. Do mesmo modo, dispõe, o art. 11 diz que os direitos à privacidade, à proteção dos dados pessoais e ao sigilo das comunicações privadas e dos registros deverão ser obrigatoriamente respeitados em qualquer operação de coleta, armazenamento, guarda e tratamento de dados, desde que pelo menos um ato ocorra em território nacional. Os princípios da finalidade e da adequação dos dados pessoas também já estavam presentes no MCI. A lei traz expressa vedação, no art. 16, à guarda dos “registros de acesso a outras aplicações de Internet sem que o titular dos dados tenha consentido previamente” e à guarda de “dados pessoais que sejam excessivos em relação à finalidade para a qual foi dado consentimento pelo seu titular” (Maciel, 2019, n.p.).

Tais leis, ainda que juntas, se mostraram insuficientes para o trato e proteção dos dados pessoais nas relações que ocorrem na Internet, sobretudo no espectro do direito do consumidor; o que se falar acerca da suficiência para tutelar acerca de discriminação algorítmica. Nesse sentido, o Código de Defesa do Consumidor (nº 8.078, de 1990) (“CDC”) já demonstrava uma ampla proteção aos dados pessoais dos consumidores. E não só isso, mas o direito do consumidor engloba, em seu escopo

interpretativo, um viés aberto e mais amplo de proteção dos consumidores, porquanto aplicado como uma política de Estado. O CDC já previa que o consumidor terá acesso às informações existentes em cadastros, fichas, registros e dados pessoais e de consumo arquivados sobre ele, bem como sobre suas respectivas fontes (art. 43, *caput*), prevendo que as informações devem ser verdadeiras, claras e objetivas (art. 43,§1º); que a abertura do cadastro ou registro deve ser comunicado por escrito ao consumidor quando não solicitada por ele (art. 43, §2º); que o consumidor poderá exigir a correção de informação inexata. Assim, com o acréscimo da LGPD, que surgiu para regulamentar de forma abrangente o tratamento de dados pessoais, ampliou-se o já composto o microssistema protetivo de dados.

Todo esse aparato normativo, que já vinha fortalecido e cristalizado pela LGPD, foi respaldado pela Emenda Constitucional nº 115/2022, que fez constar explicitamente a proteção dos dados pessoais como um direito fundamental. Mesmo quando não havia a EC 115/2022, já se ventilava a possibilidade do reconhecimento do direito fundamental implícito à proteção de dados e à autodeterminação na CF/88, isto é, um direito fundamental associado. Advogando a existência desde antes da tutela constitucional, Doneda (2020, p. 8) defendeu nos autos da ADI 6.389, que a Carta da República deve ser interpretada à luz de novos desafios, buscando o reconhecimento da constitucionalidade da proteção dos dados, de modo a garantir privacidade, liberdade e autodeterminação ao cidadão. Essa preocupação é reflexo do atual contexto tecnológico, que ceifará, se não houver zelo, a própria esfera cidadã e democrática da liberdade de conduzir a sua própria vida. É por isso, assevera Doneda, que a interpretação constitucional deverá ser apontada a solução dos problemas dos novos desafios que estão por vir.

Assim, ao menos, foi reconhecido pelo Supremo Tribunal Federal ao julgamento conjunto das Ações Diretas de Inconstitucionalidade nº 6.387, 6.388, 6.389, 6.390 e 6.393 (Brasil, 2020a). Na hipótese, o STF reconheceu o direito à proteção de dados como direito fundamental implícito ao suspender a Medida Provisória n. 954, que determinava o compartilhamento dos dados pessoais dos usuários de telefonia pelas empresas do ramo ao Instituto Brasileiro de Geografia e Estatística (IBGE). A MP 954, como se sabe, foi justificada por conta da pandemia da COVID-19, para que o governo tivesse uma base epistêmica mais sólida para acompanhar o avanço da pandemia (Mendes; Fonseca, 2020, n.p.).

Para além dessas questões, insta referir que o conceito de dado adotado pela legislação brasileira abarca um conceito expansionista e dinâmico de dado pessoal, a norma brasileira tutela os dados referentes a pessoas naturais, identificadas ou identificáveis, ainda que anonimizados em algumas circunstâncias, tratados por bancos de dados públicos ou privados (Bioni, 2019, p. 74). Dessa forma, abrange não somente os dados pessoais que se associam à pessoa natural, como também os que permitem à identificação, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião do tratamento.

A LGPD define marcos para utilização dos dados, com foco em garantir os direitos da personalidade, notadamente o direito à privacidade e à autodeterminação informativa. A LGPD amplia, densifica e concretiza a tutela dos dados pessoais que não era inexistente, mas era, no mínimo, desordenada. E, ainda que talvez não inove em questões protetivas, operacionaliza um sistema protetivo que o Brasil era carente (Bioni, 2019, p. 110). Além disso, após os escândalos da Cambridge Analytica, a LGPD, como a GDPR, encontrou meios para oferecer caminhos adequados entre um não cerceamento dos direitos comunicativos e um oferecimento de bases adequadas para o desenvolvimento da economia da informação, baseada nos vetores da confiança e da segurança (Mendes; Doneda, 2018, p. 470). Adicionalmente, sobretudo pelo rápido avanço das tecnologias algorítmicas mais avançadas, a base de novos direitos estabelecida pelo movimento da LGPD dá raízes fortes para uma nova tutela especializada acerca das decisões automatizadas, resguardando-se, de modo mais direto, direitos como o da vedação à discriminação injustificada, o direito à transparência e à explicabilidade dos algoritmos e o itinerário lógico que seguiu para a tomada de decisão. Para além da tutela dos dados, agora se faz necessária uma tutela também sobre Inteligência Artificial, problemas que de nenhum modo são isolados, mas com implícito nascedouro na relação Estado/Mercado e Sociedade Civil.

Proteger dados é condição necessária, mas não suficiente: a LGPD organiza o “como” do tratamento, porém a discriminação algorítmica emerge, muitas vezes, não de ilícitos no *input*, e sim de inferências, correlações-*proxy* e retroalimentações que afetam desigualmente grupos vulneráveis no *output* decisório. Em outras palavras, mesmo sob conformidade formal com princípios de finalidade, adequação e minimização, subsiste o risco de resultados desproporcionais, o que desloca o foco

da tutela do tratamento de dados para o uso de sistemas e seus efeitos. Daí a necessidade de standards adicionais – explicabilidade proporcional ao risco, registros auditáveis, RIPD/AIA e supervisão humana significativa – como condições de legitimidade material nas decisões automatizadas. É precisamente esse ponto de inflexão que exige atenção reforçada da jurisdição constitucional à igualdade em sua dimensão efetiva.

É nesse contexto que a discussão sobre discriminação algorítmica ganha relevo, especialmente diante do avanço de sistemas de tomada de decisão automatizada, que desafiam os marcos tradicionais de proteção jurídica. A discussão, cada vez mais complexa, revela que a proteção dos dados não se limita ao controle sobre a coleta ou ao consentimento do titular, mas se estende à análise crítica dos efeitos produzidos pelo uso dos dados em sistemas algorítmicos. É preciso, assim, ir além dos instrumentos tradicionais e pensar em mecanismos de controle que abarquem também a transparência dos processos decisórios, a auditabilidade dos sistemas e a fiscalização dos resultados, especialmente quando envolvem grupos vulneráveis ou interesses difusos.

O legislador brasileiro já protege a coleta e uso de dados pessoais, com destaque para a Emenda Constitucional nº 115/2022, que tornou a proteção de dados um direito fundamental, e para a Lei Geral de Proteção de Dados, que regulamenta o tratamento desses dados por pessoas físicas e jurídicas, visando à liberdade, privacidade e desenvolvimento pessoal. Mesmo antes dessas normas, a Constituição já garantia tutela dos dados pessoais, igualdade e proibição de discriminação injustificada, conforme os artigos 5º e 3º. À luz dessas balizas constitucionais e legais, a tutela contra impactos díspares produzidos por decisões automatizadas reclama uma engenharia probatória compatível com a assimetria informacional típica desses sistemas.

Em casos de discriminação algorítmica indireta, a prova do impacto díspar deve ser facilitada ao demandante, transferindo-se ao operador/desenvolvedor o encargo de demonstrar a legitimidade, adequação e necessidade dos critérios utilizados, inclusive a existência de alternativas menos gravosas. Essa alocação dialoga com a lógica de inversão ou flexibilização do ônus já reconhecida no direito antidiscriminação comparado e é normativamente coerente com o art. 20 da LGPD (direito de revisão/explicação) e com a técnica processual de distribuição dinâmica do

ônus probatório. No plano estrutural, trata-se de aplicar, ao ambiente algorítmico, a exigência de cultura da justificação: demonstrado *prima facie* o efeito desigual, incumbe ao responsável pelo sistema comprovar que o meio empregado fomenta o fim legítimo e que não havia alternativas menos discriminatórias, nos termos do teste de proporcionalidade."

O 6º artigo da LGPD traz o princípio da não discriminação, segundo o qual não se poderiam utilizar dados pessoais para discriminar ilicitamente e abusivamente pessoas. No âmbito pessoal, essa discriminação ocorre(ria) a partir das técnicas de perfilização, p.e., utilizadas para restringir crédito, mesmo com os CPFs não negativados nas listas de proteção ao crédito, na qual poderiam ser utilizados informações como gênero, etnia, endereço residencial para dar uma maior ou menor pontuação a partir do risco ofertado à empresa de crédito.

Algumas legislações, como é o caso do GDPR – da União Europeia – e da LGPD, por inspiração a essa, trazem o direito de contestação às decisões tomadas com base em algoritmos. Eis uma grande possibilidade de expor os motivos determinantes das decisões tomadas de maneira automatizada ou como suporte por mecanismos de algoritmos. Na LGPD esse direito está contido no art. 20 e regula uma hipótese importante para a regulação *a posteriori* e combativa da discriminação. Contudo, essa não é a melhor forma final de resposta da legislação para cumprir com o dever de proteção estatal (*Schutzpflicht*) do direito à não discriminação (Leal; Maas, 2020, p. 50).

O constitucionalismo contemporâneo tem a tarefa de ser o guardião das portas de *Ἴλιον* (Ílion/Troia): examinar cada presente, sondar cada cavalo que se ofereça ao espaço público digital. O dever de proteção estatal contra discriminações algorítmicas exige exatamente a recusa da ingenuidade. É dever da jurisdição constitucional abrir o cavalo antes que o inimigo se espalhe pela cidade, verificando se a promessa de inovação não oculta a semente da exclusão.

No âmbito da União Europeia, desde 2021, tramita Proposta de Regulamento da Inteligência Artificial (*AI Act*), a qual tem por um de seus objetivos o estabelecimento de regras harmonizadas em matéria de inteligência artificial, melhorando as previsões, otimizando as operações e personalizando o fornecimento de serviços. Em sentido mais específico, e considerando que a Proposta apresenta natureza horizontal, isto é, exige plena coerência com a legislação da União Europeia,

garante um complemento – não um prejuízo (Oliveira; Dias, 2023, p. 152) – à matéria de proteção de dados, notadamente o GDPR. Trata-se de desenvolver e organizar regras harmonizadas aplicáveis a ambas as temáticas.

Nessa moldura, o AI Act densifica a tutela de direitos fundamentais por duas vias complementares: (i) um elenco taxativo de proibições – técnicas de manipulação subliminar ou enganosa; exploração de vulnerabilidades; avaliação e classificação social; avaliação preditiva de risco de atividade delitiva; criação de bases massivas de reconhecimento facial; inferência de emoções em ambientes laborais e educacionais; categorização biométrica discriminatória; e identificação biométrica remota em tempo real – e (ii) uma arquitetura graduada de risco que condiciona o uso de sistemas de IA segundo a magnitude do impacto sobre saúde, segurança e direitos fundamentais (López, 2024; Andrés, 2024).

A classificação regulatória estrutura três eixos: (a) sistemas de alto risco, para os quais se impõem obrigações reforçadas de governança, documentação, transparência e explicabilidade; (b) modelos de uso geral, admitidos com deveres proporcionais de supervisão e informação; e (c) modelos de risco sistêmico, cujos outputs podem razoavelmente perturbar setores críticos, afetar processos democráticos ou amplificar desinformação e conteúdos ilícitos, exigindo controles de robustez e accountability ampliados (López, 2024; Andrés, 2024).

No Brasil, a LGPD oferece a gramática mínima (art. 6º e 20), mas a tutela antidiscriminatória efetiva requer proteções adicionais de governança por risco em linha com o padrão comparado. A governança algorítmica, nessa chave, reclama um ciclo contínuo de gestão de riscos – avaliação prévia, mitigação e monitoramento – e a incorporação de Privacy by Design e Privacy by Default como salvaguardas estruturais. O regime de proteção de dados funciona, assim, como dique de contenção da dignidade, inclusive no enfrentamento de *sesgos*/viés. A explicabilidade (XAI) opera como instrumento de auditabilidade das cadeias inferenciais, permitindo reconstruir a decisão e controlar a racionalidade dos *outputs* – condição para prevenção e correção de impactos discriminatórios (Álvarez, 2024).

No plano organizacional, os riscos associados à IA são transversais – regulatório, reputacional, operacional e de responsabilização – e impõem a adoção de programas de *compliance* algorítmico, com políticas de gestão de risco, documentação técnica, trilhas de auditoria e mecanismos de resposta a incidentes

(Santomil et al., 2024). Na prática jurídica, a incorporação de sistemas de IA demanda vigilância a vieses e limites explicativos, sob pena de fricções com o devido processo e a integridade argumentativa (Luna Salas; Perona; Carrillo De La Rosa, 2023).

Do ponto de vista técnico, subsiste a máxima *garbage in, garbage out*: modelos treinam dados pretéritos para predizer o futuro – e, sem intervenções deliberadas, tendem a reproduzir assimetrias históricas. Daí a centralidade de curadoria de bases, de métricas de equidade e de avaliações de impacto como medidas de contenção de vieses e de prevenção de discriminações indiretas (Manjikian, 2023). Exemplificando, o fenômeno da invisibilidade datológica evidencia que a ausência sistemática de mulheres nos ciclos de coleta, teste e validação produz *blindspots* que degradam a acurácia e geram riscos desproporcionais. Como demonstra Criado-Perez (2020), a sub-representação feminina em ensaios clínicos e em bases de projeto levou a diagnósticos incorretos e a padrões técnicos calibrados para o corpo masculino – exemplo eloquente de discriminação indireta de base estatística que sistemas de IA tendem a reproduzir quando não há curadoria e métricas de equidade.

No direito administrativo, a especialidade da tarefa é variável decisiva de confiabilidade – quanto mais específico o escopo, mais previsível e verificável o resultado; quanto mais genérico, maior a propensão a delírios. A exigência de motivação – sobretudo quando haja discricionariedade – não pode ceder ante a automatização: a confiança difusa na técnica não substitui o dever de justificar. Sem transparência e linguagem acessível, corre-se o risco de um desaparecimento prático da discricionariedade, com decisões opacas e não controláveis (Wilhelm, 2023).

Em paralelo, experiências nacionais têm densificado direitos digitais para assegurar a dignidade também no ambiente informacional. A Agenda España Digital 2025 e a *Carta de Derechos Digitales*, articuladas à LOPDGDD, consagram posições como identidade no entorno digital; não localização e não perfilização; igualdade e não discriminação no digital; liberdade de informação; desenvolvimento tecnológico e entorno sustentável; e, de maneira basilar, “*derechos ante la inteligencia artificial y la neurociencia e garantía de los derechos en los entornos digitales*” – oferecendo base normativa para políticas públicas e controle jurisdicional consequentes (Álvarez, 2023).

Em síntese, a arquitetura jurídico-material consolidada a partir de proibições e gradação de risco no AI Act, *compliance* algorítmico e avaliação de impacto, privacy *by design/default* e XAI, motivação e transparência das decisões automatizadas e o reconhecimento expresso de direitos digitais – compõe um padrão normativo de controle das aplicações de IA orientado à igualdade e à não discriminação, que servirá, adiante, de premissa para a análise dos critérios de justificação de decisões automatizadas em contextos de colisão de direitos.

A Proposta da União Europeia já serve como marco jurídico inspirador para a regulamentação da temática em diversos ordenamentos jurídicos ao redor do mundo, incluindo o Brasil. Efetivamente, a ideia que permeia o *AI Act* está em tentar ultrapassar, ou ao menos minimizar, a tensão que necessariamente existe entre proteção de dados pessoais e inteligência artificial. A tendência (necessária) é a de centralizar o ser humano como fim em si mesmo dentro da lógica das funcionalidades da Inteligência Artificial, bem como nos escopos interpretativos de sua prática (Cantarini, 2022, p. 35).

Embora o AI Act não nomeie expressamente “discriminação estrutural”, sua arquitetura por risco e as proibições objetivas internalizam a lógica estrutural: deslocam o foco do ato individual para mecanismos sistêmicos de produção de desigualdades e impõem salvaguardas *ex ante*, visto que exigem documentação, avaliação de impacto em direitos fundamentais, supervisão e explicabilidade mais rigorosas quanto maior o potencial de impacto díspar em razão de vulnerabilidades discriminatórias.

O ordenamento jurídico brasileiro conta, desde 2020, com a Resolução nº 332 do Conselho Nacional de Justiça, que, dentre outras providências, dispõe sobre a ética, a transparência e a governança na produção e no uso de inteligência artificial no Poder Judiciário. Em 2025 o CNJ atualizou a Resolução nº 332 pela Resolução nº 615, que amplia os graus de exigência sobre o uso de IAs no Poder Judiciário, sobretudo a partir do aumento das aplicações generativas de texto. Em um sentido mais amplo há, também, tramitando no Congresso Nacional, o Projeto de Lei nº 21/2020, que visa a criação do Marco Legal da Inteligência Artificial, a partir do estabelecimento de fundamentos, princípios e diretrizes para o desenvolvimento e a aplicação da inteligência artificial no Brasil. É de se destacar, também, a Portaria nº

4.671/2021, do Ministério da Ciência, Tecnologia e Inovação, que instituiu a Estratégia Brasileira de Inteligência Artificial.

A normatividade interna vem acompanhando dos esforços internacionais para a construção de parâmetros jurídicos de contenção da discriminação algorítmica. A Resolução nº 332/2020 do Conselho Nacional de Justiça representou o primeiro gesto institucional nesse campo, formulando princípios de ética, transparência e governança no uso de sistemas de inteligência artificial. Contudo, seu alcance permaneceu limitado, voltado sobretudo a aplicações de gestão procedimental e eficiência administrativa, sem enfrentar em profundidade o problema estrutural dos vieses replicados por tecnologias de aprendizado de máquina e, menos ainda, o fenômeno mais recente das inteligências artificiais generativas.

Esse déficit inicial é significativamente corrigido pela Resolução nº 615/2025, que inaugura um novo patamar normativo. O documento não apenas reafirma valores genéricos de transparência, mas os traduz em deveres jurídicos concretos: impõe a vedação expressa à criação ou reprodução de discriminações abusivas ou ilícitas, densificando a vinculação ao princípio constitucional da igualdade. Além disso, exige explicabilidade, auditabilidade e a contestabilidade como atributos obrigatórios do uso de IAs. Determina ainda que a supervisão humana seja efetiva, periódica e adequada, interditando sistemas que eliminem a possibilidade de revisão e, de forma inédita, impõe a necessidade de representatividade dos dados de treinamento, de modo a mitigar a perpetuação de desigualdades estruturais. A normatividade, antes principialista e abstrata, converte-se em arcabouço procedimental robusto, apto a operar como filtro jurídico-material diante da opacidade algorítmica, ainda que de modo, ao sentir do presente esforço, ainda insuficiente.

Essa evolução demonstra que a institucionalidade brasileira não ignora o problema da discriminação algorítmica e da opacidade. A Resolução nº 615/2025, nesse sentido, já não trata a inteligência artificial como mera ferramenta de gestão, mas como tecnologia de decisão (ainda que potencialmente e de modo não desejável), cujo uso exige uma governança em conformidade com direitos fundamentais. Ainda assim, permanece a questão crucial de saber se tais diretrizes administrativas bastam para enfrentar o fenômeno no plano sistêmico das relações jurídicas, sobretudo porque os riscos de viés discriminatório não se limitam à esfera

do processo judicial, mas atravessam outras dimensões da vida social. Dimensões a que o judiciário será chamado a enfrentar.

Não se trata apenas de um problema “informacional”, mas de um problema estrutural de igualdade, visto que sistemas de decisão automatizada podem reproduzir assimetrias históricas, com maior dano justamente para grupos em situação de vulnerabilidade. Por isso, o eixo da tutela precisa sair do insumo, os dados, e alcançar o resultado da decisão. Sem esse deslocamento, a proteção permanece reativa e insuficiente para conter vieses sistêmicos.

É nesse ponto que emerge o limite da insuficiência da Lei Geral de Proteção de Dados para tutelar essas situações, em que pese o princípio da não discriminação esteja delineado. Veja-se, o problema está na tessitura normativa do direito à revisão de decisões automatizadas, visto que permanece centrada na proteção da privacidade e da autodeterminação informativa, sem um caráter de proteção estrutural. O problema é estrutural, não é apenas informacional. Não se trata apenas de garantir que indivíduos possam acessar, corrigir ou revisar dados pessoais, mas de impedir que os próprios sistemas de decisão automatizada reproduzam e amplifiquem desigualdades sociais preexistentes. Assim, a legislação no atual paradigma opera em chave individualista, fundada na reação *ex post* do titular frente a uma decisão que já produziu efeitos. Falta-lhe, portanto, uma dimensão preventiva que assegure diversidade nos dados, auditabilidade nos processos e mecanismos substantivos de aferição de impactos coletivos. E isso falando apenas de dados.

Além disso, a assimetria técnica entre o titular e os controladores de sistemas complexos mina a efetividade do direito de revisão previsto no art. 20. Como reivindicar a reanálise de uma decisão automatizada sem acesso a explicações compreensíveis? Como questionar a licitude de um *outcome* se não há instrumentos jurídicos para avaliar os vieses embutidos nos *datasets* de treinamento ou nos critérios e nas codificações utilizadas? A ausência de parâmetros normativos de inteligibilidade transforma o direito em promessa vazia, insuficiente para lidar com as formas sutis e silenciosas de exclusão, como é o caso do presente esforço.

Também no plano da colisão de direitos, a LGPD revela sua insuficiência. Os conflitos que emergem da utilização de IA não se reduzem a tensões entre privacidade e inovação tecnológica, mas envolvem transparência, igualdade, dignidade, livre iniciativa e sigilo empresarial. A lei, todavia, não fornece instrumentos de ponderação

material que permitam calibrar a efetividade desses direitos em graus proporcionais. A normatividade infraconstitucional limita-se a comandos gerais, incapazes de orientar o intérprete em situações nas quais a proteção contra discriminação algorítmica colide com interesses econômicos ou tecnológicos de grande peso social.

O percurso normativo revela um movimento crescente de densificação da tutela de dados pessoais e de enfrentamento do fenômeno da discriminação algorítmica, ainda que marcado por insuficiências estruturais. Do paradigma alemão ao reconhecimento constitucional brasileiro, verifica-se a consolidação de um léxico jurídico de proteção, voltado a direitos de privacidade, igualdade e não discriminação. Todavia, o quadro normativo atual permanece essencialmente reativo e fragmentado, sem ainda oferecer um método consistente para calibrar, em chave constitucional, os efeitos de decisões automatizadas sobre grupos vulneráveis. A esse diagnóstico soma-se a percepção de que, diante da opacidade técnica e da intensidade dos riscos, apenas a jurisdição constitucional pode assumir a tarefa de estruturar os limites materiais de legitimidade do uso de sistemas algorítmicos. É precisamente nesse ponto de tensão que se abre a passagem ao capítulo seguinte, no qual se buscará delinear um modelo decisório racional capaz de graduar a efetividade dos direitos fundamentais em colisão.

Em suma, o padrão comparado converge para proibições claras, governança por risco, explicabilidade suficiente e métricas de não discriminação ao longo do ciclo de vida. É esse arcabouço que informa a moldura constitucional brasileira, especialmente quando grupos vulneráveis estão em jogo. No capítulo seguinte, operacionaliza-se essa moldura em um modelo decisório que gradua a efetividade dos direitos fundamentais em colisão, alocando ônus, provas e aberturas conforme risco, opacidade e vulnerabilidade.

5 A CONSTRUÇÃO DE UM MODELO DECISÓRIO RACIONAL BASEADO EM GRAUS DE EFETIVIDADE DOS DIREITOS FUNDAMENTAIS EM COLISÃO

> *Não há neles nada de mágico. A magia está apenas no que os livros dizem, no modo como confeccionavam um traje para nós, a partir dos retalhos do universo. (...) A primeira, como eu disse, é a qualidade da informação. A segunda, o lazer para digeri-la. E a terceira, o direito de realizar ações com base no que aprendemos da interação entre as duas primeiras (Bradbury, 2009, p. 121-123).*

Em *Fahrenheit 451*, Bradbury (2009) narra um mundo em que os bombeiros não apagam incêndios: acendem-nos para queimar livros, que são queimados como símbolos da capacidade humana de questionar, compreender e resistir. O gesto de queimar é mais que censura: é apagamento da memória, supressão do dissenso e extermínio da possibilidade de crítica. É o desaparecimento de fontes do pensamento crítico resolutivo e transformador. No contexto da discriminação algorítmica, a opacidade informacional exerce função análoga. Na metáfora de Bradbury, a opacidade é nova forma de *book burning* – não se queimam livros; queima-se a razão de compreendê-los. Ao tornar indecifráveis os critérios de decisão utilizados por sistemas automatizados – especialmente em contextos de relevância jurídica –, elimina-se a possibilidade de defesa, de contraditório e, em última análise, de justiça. Trata-se de um novo tipo de queima: não mais de livros, mas de transformações a partir de direitos.

Há um contraste decisivo aqui. Para sistemas de IA, trabalhar com o passado significa ajustar-se às frequências pretéritas: generalizar a partir do que já aconteceu, comprimindo o presente em padrões históricos. Sem correções explícitas, isso fossiliza assimetrias e reproduz tais padrões automaticamente. Para o leitor humano, trabalhar com o passado é outra coisa: é interpretar, criticar e reordenar sentidos; os livros não acorrentam o homem ao que foi, pelo contrário: abrem horizontes para o que pode vir a ser. Este capítulo se debruça sobre a necessidade de reconstrução jurídica da tutela antidiscriminatória a partir da racionalidade jurisdicional. O desafio, aqui, é operar a conversão do diagnóstico em critérios normativos: pensar como o Poder Judiciário – em sua função garantidora – pode impor standards materiais de

proteção frente à atuação algorítmica estatal e privada. A exigência de transparência, auditabilidade e rastreabilidade não é um adereço técnico, mas um imperativo democrático. Ao Estado-juiz impõe-se o dever de não permitir que a lógica das máquinas substitua o julgamento humano sem que sejam previamente delineadas as fronteiras ético-jurídicas da automação. É preciso, parafraseando Bradbury, "parar de queimar os livros" – isto é, romper com a lógica de ocultação e restituir ao processo decisório sua inteligibilidade constitucional.

É tempo de apagar as fogueiras da opacidade, que queimam as razões de decidir, e de reacender o lume da explicação. Transparência, nesse contexto, não é só técnica – é requisito de controle, é, antes, condição de liberdade –. É preciso, em outras palavras, parar de queimar pensamentos humanos desvencilhantes. Montag aprende com Faber que o valor dos livros não está no suporte, mas em três condições: qualidade da informação, tempo para digeri-la e liberdade para agir conforme o aprendido. Transportado à era algorítmica, isso significa: **i)** qualidade informacional do sistema (curadoria de dados, métricas por subgrupos, validações reprodutíveis); **ii)** tempo e meios de escrutínio público (registros, *logs*, perícia independente, contraditório efetivo); **iii)** liberdade institucional para corrigir, suspender e reparar (poder regulatório e jurisdicional com remédios proporcionais). A comunidade dos “homens-livro” reaparece como consórcio cívico de auditores, autoridades e sociedade civil que preserva a memória verificável contra a obsolescência programada da explicação. Buscar o conhecimento crítico, aqui, é dever jurídico: converter modelos e estatísticas em razões públicas auditáveis.

A aplicação do modelo que se propõe é integralmente dependente do caso concreto — da prova disponível, da intensidade do impacto, da vulnerabilidade dos atingidos, da opacidade residual e da assimetria informacional. Não há — nem deve haver — regra *ex ante* de abertura do algoritmo; oferece-se, isto sim, um procedimento graduado e controlável (A1–A3), ancorado na lei material da ponderação e calibrado pelo grau de certeza epistêmica, para organizar o juízo e distribuir o ônus argumentativo e probatório conforme o grau de abertura necessária.

Se a IA tende a reencenar o passado, o Direito deve instaurar novos e melhores futuros. Em termos do atual problema, isso se traduz na tríade de abertura da caixa-preta dos algoritmos em uma escala em três etapas, A1 (documentação qualificada que torna visível o passado), A2 (meios efetivos para interpretar e testar

esse passado) e A3 (capacidade de agir para corrigir a aplicação). Quanto mais intensa a abertura (até A3), maior a exigência de certeza das premissas epistêmicas. Ao mesmo tempo, quanto maior a intensidade, mais o ônus argumentativo/informacional desloca-se ao operador, sobretudo se *big techs*, acompanhando a progressividade da escala.

A análise da discriminação algorítmica revela a complexidade e a profundidade dos desafios enfrentados na era digital, onde a tecnologia pode tanto perpetuar quanto agravar desigualdades estruturais. Diante desse cenário, é imperativo que o Direito se posicione de maneira proativa e eficaz para proteger os direitos fundamentais de modo proporcional. A seguir, necessário verificar como o sistema jurídico pode responder a essas novas formas de discriminação, examinando as ferramentas e os mecanismos disponíveis para garantir uma proteção adequada e justa. A judicialização emerge como um caminho crucial para enfrentar as injustiças algorítmicas, promovendo a transparência, a responsabilidade e a equidade no uso das tecnologias emergentes e, como visto nas decisões europeias, uma resposta aparentemente apontada pelas Cortes se dá exatamente na demonstração dos fluxos informacionais e lógicos dos algoritmos, permitindo participação transparente dos seres humanos (titulares e controladores), no processo de tomada de decisão automatizada – essa nova fenomenologia que o Direito deve responder com fôlego.

A efetividade dos direitos fundamentais diante da inteligência artificial depende de métricas para avaliar impactos, danos e resposta institucional; apenas uma parcela dessa última dimensão é foco na presente. Para decidir racionalmente, é necessário reconhecer conflitos entre esses direitos e explicitar critérios objetivos para a ponderação, como impacto algorítmico sob viés antidiscriminatório, vulnerabilidades, alternativas técnicas e domínio de mercado. O processo decisório deve contar com certeza epistêmica e apoio de técnicos, auditorias independentes e experts para confirmar possibilidades diante de cada caso, isto é, no atual foco, cada discriminação automatizada injustificada ocasionada, visto que cada IA e cada caso comportam significações particulares. A transparência exige explicabilidade, participação do ser envolvido na aplicação e possibilidade de controle. O que possibilita promover verificações e correções. Autoridades regulatórias e o Judiciário têm papel central em garantir publicidade, rastreabilidade e revisão das decisões automatizadas, cumprindo a proteção constitucional contra arbitrariedades

tecnológicas. O foco está em construir um modelo decisório para quando a aplicação algorítmica já operou no mercado e já teve um *outcome* injustificadamente discriminatório; a tese e os *insights* aqui delineados podem até auxiliar no desenho de uma arquitetura regulatória, seja por técnica ou por economicidade, o que não é visado pelo presente esforço.

Desta feita, questiona-se se é possível o estabelecimento de uma relação de eficácia integrativa e em graus de direitos fundamentais de defesa de interesses de consumidores-usuários – e ainda mais tutelar quando esses forem titulares de posições jurídicas de múltiplas vulnerabilidades sociais, dentre elas com prioritária especialidade os pertencentes a grupos vulneráveis – de aplicações algorítmicas e de agências e empresas que disponibilizam tais aplicações no mercado. A aferição de graus adequados de sobreposição da eficácia em casos é tanto possível – demonstrando quanto dos direitos empresariais devem ser relativizados para a satisfação dos direitos dos usuários – como benéfico à sociedade, aos usuários e ao próprio mercado – que não terá toda a aplicação algorítmica exposta ou inutilizada em eventuais investigações e condenações a prejuízos ocasionados e ilícitos cometidos.

A partir da lei material da ponderação – enquanto uma vertente central do escrutínio da proporcionalidade, no formato do modelo alexyano, considerando ainda estudos e críticas mais recentes da sua operacionalização – é cabível analisar os diferentes graus de adequação de medida judicialmente imposta às empresas fornecedoras de aplicações algorítmicas em casos de constatada discriminação automatizada, levando-se em conta a substancialidade dos parâmetros verificados em previsões de impacto algorítmico, ou então de avaliação do impacto algorítmico ocasionado. Igualmente, na ausência de tais exigências legislativas, a análise a partir do grau de intervenção proporcional ao grau de afetação dos direitos fundamentais em questão se mostra a medida apta a equacionar o nível e a intensidade das medidas judiciais a serem comandadas nesta nova realidade, com o zelo de que cada uma destas decisões estará formando e formatando o sentido interpretativo-normativo do direito acerca da natureza jurídica e do tratamento jurídico constitucional e substancialmente adequado na questão. O produto dessas preocupações será o de uma possível mensuração a posteriori em cada caso – e gradativamente mais a priori à medida que casos forem decididos – dos impactos que a própria operação empresarial poderá sofrer a partir dos danos ocasionados, bem como dará ao usuário

garantias de proteção, sobretudo quando esse se encontrar em situação de vulnerabilidade, e com especial gravosidade para discriminações multidimensionais e estruturais, a justificar uma maior intervenção do direito à explicabilidade em afetação ao direito ao segredo comercial-algorítmico das desenvolvedoras.

5.1 A colisão central dos direitos fundamentais da tutela à discriminação algorítmica: transparência algorítmica e segredo algorítmico

Para que se possa mensurar o grau de satisfação do direito fundamental à explicabilidade e à transparência algorítmica, bem como o grau de restrições ao segredo comercial sejam feitas de modo racionalmente justificado é necessária a utilização de procedimento lógico-argumentativo que demonstre porque um determinado direito fundamental deverá receber precedência frente a outro direito fundamental. Há uma fina relação entre a transparência algorítmica – diante da nova dinâmica de discriminações automatizadas – e a garantia da propriedade intelectual sobre as aplicações, argumentando que a medida do desvelar das programações, protegidas pelo sigilo, deve ser proporcional ao bem jusfundamental preponderante na colisão. Nessa hipótese ocorrerá a chamada colisão de direitos fundamentais, em sentido amplo, que é quando os direitos fundamentais colidentes se consubstanciam de direitos cujos bens juridicamente tutelados são coletivos (Gavião Filho, 2011, p. 41).

Na colisão aqui delineada, a pessoa atingida ostenta os direitos à explicabilidade proporcional ao risco, à transparência e à proteção de dados, os quais integram o seu âmbito de proteção informacional e processual de defesa; no polo oposto, a desenvolvedora invoca a livre iniciativa, a função social da empresa e o segredo comercial-algorítmico. Não se trata de antinomia absoluta, mas de graduação proporcional: o núcleo essencial de cada direito — e os limites aos limites — orienta o quanto de abertura informacional é exigível, sempre sob o dever estatal de proteção.

A ponderação entre transparência/explicabilidade e segredo comercial exige explicitar *trade-offs* entre eficiência/acurácia e igualdade material. A literatura técnico-normativa reconhece que restrições de "*fairness*" podem tensionar objetivos de desempenho (*fairness–utility trade-off*), impondo escolhas justificadas entre valores constitucionais relevantes (Kleinberg; Mullainathan; Raghavan, 2019). Por isso, a

decisão não pode apoiar-se apenas em métricas internas *(accuracy, precision, recall, F1-score*) (Gavião Filho, 2025), que não captam efeitos distributivos, nem tampouco em indicadores isolados de *fairness*. Impõe-se metodologia híbrida — quantitativa e qualitativa — que integre métricas técnicas a parâmetros normativos (dignidade, igualdade, não discriminação) e instrumentos de Avaliação de Impacto Ético, Social e de Direitos Humanos (HRESIA), verdadeiramente de direitos fundamentais, garantindo que ganhos marginais de acurácia não sirvam de pretexto para custos sociais desproporcionais (Barocas; Hardt; Narayanan, 2019).

Sempre que a desenvolvedora alegar perdas de desempenho, o juízo deve explicitar os *trade-offs fairness-utility*, recusando tanto decisões apoiadas apenas em métricas internas de acurácia quanto soluções de *fairness* isoladas. Impõe-se metodologia híbrida: métricas técnicas adicionadas a parâmetros normativos (dignidade, igualdade, não discriminação) e avaliação de impacto adequada ao risco, para impedir que ganhos marginais de acurácia legitimem custos sociais desproporcionais.

Devido aos fatores já expostos até aqui, percebe-se que as *black boxes*, isto é, as aplicações algorítmicas são frequentemente usadas em áreas de relevante interesse comercial e social. As empresas de crédito, por exemplo, utilizam algoritmos para tomar decisões sobre a concessão de crédito, entretanto, o público não possui acesso à lógica interna destes algoritmos, podendo resultar, inclusive, na restrição à direito fundamental (O'Neil, 2020), como se viu. Tais lógicas foram debatidas em casos do Tribunal de Justiça da União Europeia, sobretudo a decisão do Processo C-817/19 – no sentido de que, dentre outras questões, a coleta e o armazenamento em massa de dados pessoais, sem garantias adequadas de proteção e supervisão, é incompatível com os direitos fundamentais, há, portanto, que se ter um cuidado em relação aos dados coletados, ao uso/tratamento destes dados, bem como uma accountability acerca da forma como as decisões automatizadas são trabalhadas nesta nova lógica.

Exemplarmente, considere-se um sistema de análise de crédito que, sem empregar variáveis sensíveis, utiliza "local de residência" como atributo decisivo. Em cidades com segregação histórico-espacial, tal variável pode elevar a taxa de rejeição de candidatos de bairros periféricos, mantendo-se equivalência quanto a condições financeiras. O escrutínio proporcional evidencia: (i) legitimidade do fim (redução da

inadimplência), (ii) adequação apenas se demonstrada contribuição técnica idônea do critério para o fim, (iii) necessidade negativa quando existirem alternativas individualizadas menos gravosas (histórico de pagamentos, comprometimento de renda) e (iv) proporcionalidade em sentido estrito desfavorável quando o ganho marginal de acurácia não compensa o custo social de perpetuação de exclusões. Nessa hipótese, a prática não passa no teste e o impacto díspar carece de justificação (Gavião Filho, 2025, n.p.).

Não há, por exemplo, como admitir, ao menos por enquanto, que juízes e tribunais decidem a partir de instrumentos algorítmicos. O'Donnell (2025, n.p.) descreve uma adoção inicial e heterogênea de Inteligência Artificial generativa por magistrados nos EUA, cotejada com episódios concretos de erro: **i)** em Minnesota, a corte federal descredenciou o testemunho de um professor de Stanford por citações inexistentes geradas por Inteligência Artificial em ação sobre *deepfakes* eleitorais; **ii)** em julho, juízes federais retiraram/reescreveram decisões após advogados apontarem imprecisões factuais e trechos que "não constavam" dos precedentes citados; e **iii)** a Corte de Apelações da Geórgia anulou ordem de primeira instância que se apoiava em casos fictícios, alucinados pela Máquina. Esses exemplos ilustram o risco sistêmico de alucinações não detectadas em decisões, além de uma assimetria de accountability: sanções são claras para advogados, enquanto a retratação de equívocos quando partem do próprio julgador é mais custosa e pode afetar a confiança pública.

No plano operativo, emergem diretrizes prudenciais: a fronteira entre tarefas delegáveis (instrumentais e repetitivas) e o núcleo prudencial do julgar é porosa; modelos generalistas podem produzir resumos plausíveis, porém cronologias e referências equivocadas. A adoção responsável requer **i)** uso restrito da Inteligência Artificial a funções de apoio – pesquisa, triagem e organização de autos, transcrição preliminar, geração de perguntas – com verificação humana obrigatória; **ii)** *checklists* de validação *ex ante* (conferência de citações em bases oficiais, reconciliação de cronologias com os autos, rastreabilidade do texto); **iii)** segregação funcional entre rascunho assistido e formação do juízo, vedado o emprego para decisões que envolvam discricionariedade substancial, com mais seriedade e importância quanto mais sensível for a afetação possível à direito; **iv)** mecanismos céleres de correção e transparência quando erros forem detectados. Se há de se incorporar as Inteligências

Artificiais, que seja feito como apoio cognitivo incremental e auditável, mantendo humana a responsabilidade decisória e a justificação pública – especialmente porque o problema das “alucinações” permanece não resolvido (O’Donnell, 2025, n.p.)

Nesse mesmo sentido, mas em atividades mercadológicas, outros 3 casos da Corte de Amsterdã complementam muito bem o sentido da juridicidade diante destas novas relações algorítmicas, como trazem Fischer e Viollier (2023, p. 162-163): **i)** Caso Ola Netherlands BV de 11 de março de 2021 Rb. Amsterdã-C/13/689705/HA RK 20-258. Caso em que condutores que usavam o App Ola para compartilhar viagens pediram acesso à informação para saber como seus dados pessoais eram utilizados, sobretudo de quais os dados utilizados (e do modo de utilização) para se aferir ao "grau de probabilidade de fraudes", que era dito como um diferencial da plataforma. O tribunal determinou o compartilhamento das informações para que os titulares dos dados entendessem o mecanismo e pudessem corrigir e determinar um processamento de seus dados legalmente tratados; **ii)** Caso Uber C/13/692003/HA RK de 11 de março de 2021. Nesse caso, semelhante ao próximo, a Uber demonstrou que as decisões automatizadas passavam pelo pessoal da área de riscos para serem efetivadas, de modo que foi absolvido de qualquer indenização pela participação humana – tão somente; **iii)** Caso Uber C/13/696010/HA ZA de 21 de abril de 2021. Nesse caso, de desativação de contas de modo automatizado pela plataforma, a partir de uma atuação integralmente robótica do software mastermind – que detecta fraudes –, a Corte de Amsterdã entendeu que o uso da aplicação sem intervenção humana feria o artigo 22 da General Data Protection Regulation, colocando os titulares em uma situação kafkiana quando têm seus direitos infringidos por alegações infundadas sem a capacidade de se defender delas, visto que não sabem e não tem a origem e a rastreabilidade do raciocínio algorítmico que decidiu se tratar de situação de fraude.

A explicabilidade das Inteligências Artificiais já é lei. Na França (2016), o Decreto 2016-1321, de 7 de outubro, sobre a República Digital, estabelece que os cidadãos devem ser informados quando uma decisão individual é adotada por meio de um algoritmo. O Decreto 2017-330, de 14 de março (França, 2017), sobre direitos das pessoas sujeitas a decisões individuais adotadas com base em tratamentos algorítmicos, obriga a administração a informar de forma inteligível sobre o grau e o modo em que o algoritmo contribuiu para o tratamento automatizado da tomada de decisões, incluindo os dados tratados, a sua fonte, os parâmetros de tratamento e a

sua ponderação, bem como as operações realizadas. Já na Itália (2023), o Conselho de Estado exige um prévio conhecimento pleno do módulo utilizado e dos critérios empregados para adotar a decisão. Esse conhecimento do algoritmo deve ser garantido em todos os aspectos, desde seus autores até o procedimento empregado para sua elaboração, o mecanismo de decisão, as prioridades atribuídas no procedimento de avaliação e os dados selecionados como relevantes. Isso permite verificar se os critérios, condições e resultados do procedimento automatizado cumprem as prescrições e finalidades estabelecidas pela lei ou pela própria administração, além de esclarecer as regras e as modalidades de programação. No Brasil, a previsão da revisão de decisão automatizada está no art. 20 da LGPD, em que pese ainda não haja a previsibilidade e a exigibilidade da explicabilidade.

A simples exigência e demonstração pela necessidade de transparência tende a inibir certas práticas corporativas e governamentais que direcionam programações discriminatórias – o uso de programas discriminatórios pode vir a configurar o crime de racismo, por exemplo – e sob uma perspectiva mediata, a transparência objetiva permitir a abertura da caixa-preta dos algoritmos e auxiliar em um ajuste mais fino e sofisticado nas ferramentas, entregando respostas cada vez mais coerentes para com o núcleo do Estado Democrático de Direito, que atualmente vem sendo ofendido.

A transparência na Inteligência Artificial enfrenta desafios devido à dificuldade de compreender como os algoritmos tomam decisões, caracterizando a chamada *black box*. Modelos explicáveis e divulgação dos dados de treinamento são essenciais para superar esse obstáculo. A transparência algorítmica é necessária para avaliar resultados, exigindo ponderação entre ética, direito e segredo comercial, a fim de proteger informações sensíveis sem comprometer direitos constitucionais. Restrições ao segredo devem ser justificadas por argumentos sólidos e avaliação dos direitos fundamentais.

Para evitar que a opacidade se converta em barreira probatória intransponível, adota-se um modelo bifásico de alocação do ônus da prova. Primeiramente o autor prova *prima facie* o impacto injustificado por indicadores idôneos, que comprovem a discriminação direta ou indireta (algorítmica). Feita essa prova, o ônus *probandi* é invertido em desfavor do réu, a quem caberá demonstrar, com documentação técnica auditável, que a medida atende a um fim legítimo, é adequada e necessária, e que os

benefícios não são desproporcionais ao custo imposto a grupos vulneráveis. Sem essa demonstração, o critério é inválido.

A essencialidade da promoção de um elevado grau de transparência algorítmica, de modo a possibilitar o conhecimento daquilo que determina o resultado pela ferramenta, é substancial (Lordelo, 2020, p. 149). O que é alcançável a partir da inclusão destas preocupações sociojurídicas e éticas na etapa de desenvolvimento e treinamento das aplicações algorítmicas. Não realizadas ainda nessas etapas, resguarda-se ao Poder Judiciário – se não antes ao Poder Legislativo – que identifique a necessidade da promoção de políticas de ética, privacidade e não-discriminação *by design*. O que implica também uma explicabilidade possível de ser exigida da Inteligência Artificial desde seu desenho, permitindo, por exemplo, uma maneira que os programadores terão de corrigir a ética das decisões automatizadas tomadas na fase de treinamento e, em um segundo momento, que a própria máquina, ao notar o potencial resultado é discriminatório, acione o humano para questioná-lo acerca da assertividade e da ética de resultado, permitindo-lhe o uso adequado do *feedback* – ferramenta com a qual a máquina pode entender a adequação ou não de sua resposta, reforçando as boas respostas e descartando as respostas indesejáveis (O'Neil, 2020, p. 207).

É fundamental garantir transparência para compreender como ferramentas algorítmicas produzem resultados. Isso se alcança ao incorporar preocupações sociojurídicas e éticas no desenvolvimento dessas aplicações. Vale lembrar que há diferentes níveis de automação: alguns controláveis, outros não, sendo necessário definir limites programáveis e o papel do controle humano. Contudo, a arquitetura dos protocolos não é o centro da ausência de luz, a opacidade está nos pormenores das múltiplas engrenagens que ficam atrás das paredes da sala algorítmica (Polonski, 2018, n.p.). Por isso que o grau de satisfatoriedade nestas dimensões parafuncionais dos algoritmos é tão dependente do mesmo fator de seu sucesso estrito: a habilidosa capacidade do programador humano em construir e fortificar um algoritmo capaz de entregar o melhor resultado deixando mais visíveis os rastros lógicos de sua funcionalidade (Steiner, 2012, n.p.). Eis aqui o ponto chave da ocorrência dos vieses e o autêntico *lócus* de colisão entre transparência algorítmica e segredo algorítmico.

Acerca do segredo algorítmico cabe uma rápida menção aos desavisados, do porquê ele interessa tanto aos desenvolvedores. Ora, dados são poderosos, e muito

mais do que o petróleo. O petróleo tem um valor inexoravelmente ligado à sua disponibilidade, visto que é finito e de certa forma difícil de ser extraído. Os dados, por sua vez, estão à disposição, são potencialmente inesgotáveis e mais: enquanto houver homem, haverá economia, haverá consumo, haverá dados e haverá economia de dados aliada à realidade da internet das coisas, do *Big Data* e da economia comportamental e dos dados, o homem passa a ser objeto de sua própria autodeterminação, em que no influxo atual consome e é levado a consumir o que propriamente fornece aos *players* do mercado, podendo ter, inclusive, negado certos bens e serviços a partir dos dados fornecidos. Paradigmaticamente, passa a sofrer influxos personalizados pelas redes e começa a inferir suas predileções, vontades e atitudes em um mito moderno, que é o da consciência. Por trás de tudo isso estão as decisões automatizadas, cujo combustível são os dados (Hoffmann-Riem, 2022, p. 12).

Nesse cenário mercadológico, de necessária e constante evolução das técnicas, o segredo comercial, ou segredo do código-algorítmico é um direito fundamental das empresas desenvolvedoras, sendo essencial para garanti-lhes a livre concorrência e a inovação, destacando-as no mercado e, inclusive, possibilitando que lucrem e queiram continuar desenvolvendo tais tecnologias, que também são suas propriedades intelectuais. O artigo 170 da Constituição da República, mais especificamente no inciso IV, traz como ditames da interpretação e da aplicação do direito à livre concorrência na atividade econômica.

A livre iniciativa é um princípio fundamental do sistema econômico capitalista, que pressupõe a liberdade de os indivíduos empreenderem e realizarem suas atividades econômicas sem intervenção excessiva do Estado. Nesse sentido, a livre concorrência se apresenta como um pressuposto da livre iniciativa, uma vez que, para que haja liberdade de mercado, é necessário que não haja monopólios ou oligopólios que possam interferir na concorrência. A garantia da livre concorrência é, portanto, um direito fundamental que visa assegurar a livre iniciativa e a preservação do mercado competitivo (Peixoto, 2020, p. 135-137).

Romper com essa lógica é renunciar à proteção fundamental gozada por empresas e empreendedores de sua higidez financeira, a partir do princípio da conservação empresarial. A higidez financeira das empresas também é uma preocupação legítima, uma vez que a manutenção da saúde financeira das empresas

é essencial para a continuidade de suas atividades econômicas e para a criação de empregos, geração de renda e continuidade no desenvolvimento tecnológico. Nesse sentido, a proteção do segredo comercial é vista como uma forma de assegurar a competitividade das empresas, que podem desenvolver tecnologias e inovações sem o risco de que seus concorrentes copiem seus modelos de negócio ou estratégias comerciais, podendo perder competitividade (Paranhos, 2022, p. 93).

Há, portanto, que se reconhecer a existência de um direito fundamental associado que deve ser derivado dos princípios liberais postos na Constituição da República Federativa do Brasil. Um direito fundamental que signifique uma proteção do capital privado às incursões – ainda que protetivas e contramajoritárias – para que se dê dimensões protetivas ao negócio empreendido, sobretudo quanto ao desenvolvimento tecnológico, que necessita de monetização para ocorrer. Uma lógica jurídica que respeitem efetivamente os valores constitucionais (Reis, 2020, p. 97).

Assim, transparência algorítmica deve ser equilibrada com a proteção dos segredos comerciais das empresas de tecnologia; porém não poderá haver uma única proteção polarizada de um único direito fundamental. A proteção aos direitos fundamentais é feita conjuntamente e em graus, tanto quanto possíveis. Aqui se contrasta o direito fundamental à proteção à propriedade intelectual sobre o desenvolvimento algorítmico e ao segredo algorítmico, com o direito fundamental à transparência algorítmica e à não-discriminação; o que evidentemente poderá, em casos específicos, ganhar ainda mais destaque.

O segredo comercial é um direito fundamental das empresas e é essencial para garantir a livre concorrência e a inovação. Portanto, é necessário encontrar um equilíbrio adequado entre a transparência algorítmica e o segredo comercial, levando em consideração o impacto de uma decisão automatizada na vida das pessoas. A LGPD é um passo importante para garantir a privacidade dos dados pessoais dos indivíduos e a transparência algorítmica. No entanto, é necessário que a autoridade nacional responsável pela aplicação da lei (ANPD) tenha os recursos e a capacidade necessários para fiscalizar e fazer cumprir as disposições da LGPD. Além disso, é importante que a sociedade esteja consciente da importância da privacidade e da transparência algorítmica e exerça seus direitos em relação a essas questões.

Não obstado por tanto, o artigo 170 da Constituição da República, mais especificamente no inciso IV, traz como ditames da interpretação e da aplicação do

direito à livre concorrência na atividade econômica. A livre concorrência é um pressuposto jurídico pensado a partir de uma compreensão de direito pautado na valorização do trabalho humano, na livre iniciativa, na existência digna da nação. É, sem dúvidas, uma previsão a ser levada como objetivo e princípio norteador de todo o ordenamento jurídico.

A proteção ao segredo comercial-algorítmico – projeção da livre iniciativa e da livre concorrência – não opera como salvo-conduto, mas como limite com função: preservar a vantagem competitiva até onde não inviabilize a verificação de premissas que comprimem direitos fundamentais. A ponderação dispensa absolutos: quanto maior o risco de dano e a vulnerabilidade dos atingidos – e quanto mais baixa a explicabilidade residual –, mais denso é o ônus de abertura sob confidencialidade, com salvaguardas de perícia assistida e compartimentalização de acesso. A tutela concorrencial legítima permanece, mas cede na exata medida necessária à reconstrução pública do itinerário lógico-decisório – nunca o contrário.

O que será necessário é que a solução dos embates entre os direitos fundamentais cuja colisão se está propondo de modo que a inescapável relativização dos direitos de propriedade intelectual não o inviabilize totalmente (como poderá ocorrer com decisões que mandem abrir totalmente o código e a programação de uma aplicação que ocasionou discriminações automatizadas). A desproteção jusfundamental das empresas de tecnologia, nesse sentido, abre o flanco para arbitrariedades que são nocivas não apenas às empresas, mas à coletividade que, levadas às últimas consequências, poderá ver uma empresa desenvolvedora de tecnologia (que poderia beneficiar o ser humano) ruir por perda da competividade em um cenário de desproporcional proteção aos direitos individuais (Paranhos, 2022, p. 93) – ainda que haja justificativas jurídicas para tanto, eis que hão de haver também para a preservação do segredo comercial, salvo se a conservação do segredo cause uma desproporcional e insuficiente proteção à não-discriminação e em uma ineficaz transparência.

A inteligência artificial é, em essência, um reflexo do conhecimento humano, de alcance instantâneo – mas não necessariamente correto nem eficaz. O ser humano, por vezes, torna-se ignorante diante do próprio conhecimento, limitado para uns, mais abundante para outros. A Inteligência Artificial não é totalmente abstrata: é uma montagem virtual sobre o real já experienciado – um relance, uma sombra. Deve

ser utilizada como ferramenta, e não como um fim em si. Do contrário, o ser humano torna-se prisioneiro de um universo distorcido por um mero reflexo algoritmizado da realidade externa. A Inteligência Artificial pode ensejar retrocesso explicativo-intelectual, embora represente um avanço tecnológico significativo. Os pés devem estar sempre fincados no presente concreto, para que o universo virtual da algoritmização robótica sirva como meio de aperfeiçoar o mundo real.

A proteção aos direitos individuais é necessária e urgente em contexto algorítmico. A proteção deve vir sempre e em todo o caso em que identificada lesão ou ameaça de lesão a direito fundamental pessoal ou coletivo. Isso não significa dizer, como aludido, que sempre e em qualquer caso, seja qualquer o grau de afetação à direito fundamental, se torna legítima a total abertura de códigos-fonte, chaves algorítmicas, programação, caminho decisório automatizado ou demonstração do funcionamento da rede neural.

O reconhecimento do direito fundamental associado à proteção do segredo comercial algorítmico é importante para que se garanta paridade de armas em contexto de jurisdição constitucional entre ser-usuário e empresa-desenvolvedora. A proteção aos direitos individuais é necessária e urgente em contexto algorítmico. A proteção deve vir sempre e em todo o caso em que identificada lesão ou ameaça de lesão a direito fundamental pessoal. Isso não significa dizer que sempre e em qualquer caso, seja qualquer o grau de afetação à direito fundamental, se torna legítima a total abertura de códigos-fonte, chaves algorítmicas, programação, caminho decisório automatizado ou demonstração do funcionamento da rede neural.

Para que essas restrições ao segredo comercial sejam feitas de modo racionalmente justificado é necessária a presença de algum tipo de procedimento lógico-argumentativo que demonstre porque um determinado direito fundamental deverá receber precedência frente a outro direito fundamental. Nessa hipótese ocorrerá a chamada colisão de direitos fundamentais, em sentido amplo, que é quando os direitos fundamentais colidentes se consubstanciam de direitos cujos bens juridicamente tutelados são coletivos (Gavião Filho, 2011, p. 41).

O teste da proporcionalidade é utilizado para decidir qual direito fundamental deve prevalecer em casos de conflito. Esse método precisa estar ligado aos conceitos de proibição do excesso (*Übermassverbot*), vedação da proteção insuficiente (*Untermassverbot*), ao núcleo essencial dos direitos fundamentais e aos chamados

limites aos limites (*Schranken-Schranken*), além de considerar o dever estatal de proteção (*Schutzpflicht*) durante a interpretação das normas constitucionais. O teste de proporcionalidade funciona como uma ferramenta lógica para verificar se a solução encontrada para um conflito entre direitos fundamentais está de acordo com a Constituição. Ele é dividido em três etapas: adequação (ou idoneidade técnica), necessidade e proporcionalidade em sentido estrito (ou ponderação). Essas etapas permitem avaliar, de forma gradual, se o meio usado é apropriado, imprescindível e em que medida deve ser maximizado ou restrito, levando sempre em conta as possibilidades práticas e jurídicas (Gavião Filho, 2019, p. 239).

Não se detendo a cada uma delas, o que interessa é analisar que a partir das verificações subsequentes dos testes, o Estado (legislador e juiz) deverá escolher o meio menos restritivo de direitos para que possa ser considerada constitucional a restrição do outro direito. No caso da restrição do segredo comercial, enquanto um direito fundamental associado à saudabilidade da empresa.

No âmbito da explicabilidade e da transparência algorítmica, o teste da proporcionalidade pode ser um instrumento útil para avaliar se a restrição ao segredo comercial é justificada pela proteção de outros direitos fundamentais dos titulares dos dados pessoais ou da sociedade em geral. Por exemplo, pode-se questionar se a divulgação do código-fonte ou dos parâmetros de um algoritmo é adequada, necessária e proporcional para garantir o direito à informação, à participação, à não discriminação ou à responsabilização dos agentes envolvidos no tratamento dos dados pessoais. Nesse sentido, deve-se ponderar os interesses e os valores em jogo, bem como as possíveis alternativas e salvaguardas para conciliar o segredo comercial com a explicabilidade e a transparência algorítmica.

Por outro lado, o teste da proporcionalidade também pode servir para limitar ou condicionar a exigência de explicabilidade e de transparência algorítmica, quando esta representar uma ameaça indevida ao segredo comercial ou a outros direitos fundamentais. Por exemplo, pode-se argumentar que a revelação de certos aspectos técnicos ou funcionais de um algoritmo pode comprometer a sua segurança, a sua eficiência ou a sua inovação, ou ainda violar direitos autorais, patentes ou outros direitos de propriedade intelectual. Nesse caso, deve-se verificar se há uma razão legítima e suficiente para exigir a explicabilidade e a transparência algorítmica, bem

como se há meios menos invasivos ou prejudiciais para atingir os objetivos de tutela de direitos atingidos pelas aplicações algorítmicas automatizadas.

Dessa forma, o grau de satisfação da explicabilidade e da transparência algorítmica em face do segredo comercial deve ser avaliado caso a caso, com base no teste da proporcionalidade e nos princípios constitucionais aplicáveis. Não há uma resposta única ou definitiva para essa questão, mas sim um equilíbrio dinâmico e contextualizado entre os diferentes interesses e valores em jogo. O importante é que o processo de tomada de decisão seja transparente, fundamentado e sujeito ao controle jurídico e social, de modo a garantir a legitimidade e a efetividade da proteção dos direitos fundamentais envolvidos, sem atribuir pesos diferenciados *a priori* de violações de direitos e em abstrato. Como critério de gravidade, deve-se considerar prioritariamente as situações de restrição direta à fruição de direitos fundamentais de indivíduos pertencentes a grupos vulneráveis.

O teste de proporcionalidade, aqui, cumpre função análoga à de Laocoonte – sacerdote troiano do deus Apolo que advertiu os troianos a desconfiar do presente grego. Sua prudência foi silenciada, mas permanece como símbolo da necessidade de racionalidade crítica. Ponderar, neste contexto, é precisamente desconfiar: abrir a estrutura, avaliar seus riscos, mensurar os graus de proteção de direitos que nela se escondem. Cada algoritmo é, em potência, um cavalo de Troia; e a ponderação, como régua central do teste, é o único instrumento capaz de expor seus ocupantes à luz do dia (Homero, 2013).

Ainda, a cultura da justificação impõe que limitações a direitos sejam amparadas por razões públicas robustas. Sempre que o benefício invocado for marginal — p. ex., ganho mínimo de acurácia — e o custo antidiscriminatório elevado, revela-se desproporcional manter a prática. A justificativa deve, pois, explicitar porque não bastariam medidas menos intrusivas para equalizar desempenho e igualdade. Além do mais, há dignidade no direito fundamental à propriedade intelectual, seja na faceta de seu segredo algorítmico/comercial ou então na conservação dos interesses socioeconômicos que, em colisão, devem ser corretamente colididos e ponderados na holística da atual conjuntura para que se decida racionalmente acerca do grau de sua relativação nos casos de discriminações algorítmicas.

Do ponto de vista da colisão central, a pessoa atingida ostenta direitos à explicabilidade – que não se confunde com a previsão do art. 20 da LGPD (voltada à

revisão de decisões) –, à transparência e à proteção de dados; de outro lado, a desenvolvedora invoca livre iniciativa, função social da empresa e segredo comercial-algorítmico acerca do programa da aplicação. A tensão entre explicabilidade e segredo não é antinomia absoluta, mas problema de desenho acerca das eficácias jusfundamentais que, *in casu*, devem ser colididas e ponderadas acerca de uma preponderância no caso. Nesse sentido, há uma clara e nítida correlação entre a exigência por explicabilidade e transparência e a conservação do segredo da programação pelos algoritmos, sendo ambas vistas como soluções possíveis ao problema do enviesamento algorítmico (Maggiolino, 2019, p. 4-5).

Em termos operativos, a colisão não se resolve por absolutos, mas por graduação proporcional: quando houver indícios de afetação relevante a direitos fundamentais, especialmente de grupos vulneráveis, a explicabilidade prepondera na medida estritamente necessária para reconstituir o itinerário lógico-decisório; preserva-se o segredo comercial por cuidado com o nível de abertura da escala de A1 a A3, conforme o risco e o impacto.

Nessa escala, A1 identifica hipóteses de baixo impacto, em que os efeitos são reversíveis e a identificabilidade de titulares é reduzida; a arquitetura do sistema, em geral, oferece alta explicabilidade e trilhas de auditoria suficientes para reconstituir o percurso lógico-decisório. A2 corresponde a impacto médio, típico de contextos em que a decisão automatizada pode repercutir no acesso a serviços essenciais ou oportunidades, operando sob opacidade moderada e com reversibilidade possível, porém onerosa – em custos, tempo ou fricções institucionais. A3 designa alto impacto, com efeitos graves ou irreversíveis, frequentemente incidentais sobre grupos em situação de vulnerabilidade, em cenário de baixa explicabilidade/auditabilidade e, por vezes, histórico de falhas ou elevado poder de profiling.

A alocação do caso concreto em cada classe não é automática, mas decorre da ponderação conjunta de critérios: a gravidade e a irreversibilidade dos efeitos; a extensão do impacto sobre direitos e liberdades; a incidência qualificada sobre grupos vulneráveis; o grau de opacidade ou inauditabilidade do artefato; o nível de identificabilidade e de combinação de dados; e o histórico de responsabilização ou incidentes. Essa graduação orienta a calibragem das exigências de transparência e explicabilidade – e, correlatamente, do espaço legítimo de segredo comercial – segundo o risco e o impacto efetivamente demonstrados.

Para que se possa garantir decisões de base algorítmica mais justas, imparciais e menos discriminatórias, é necessário que se cuide com **i)** a qualidade dos dados utilizados pelos algoritmos, isto é, os dados devem ser verdadeiros, precisos, relevantes, imparciais e diversificados, devendo refletir a diversidade da sociedade, e não podem permitir a perpetuação de preconceitos ou estereótipos. Uma boa decisão automatizada começa com uma boa base qualitativa de dados.

Um segundo cuidado a se ter é o da **ii)** transparência do funcionamento dos algoritmos, algo que ainda é uma caixa-preta regulatória, jurídica e de programação, uma vez que o funcionamento dos algoritmos é protegido pela propriedade intelectual da empresa desenvolvedora. Exige-se, nesse ponto, um necessário novo olhar a partir da colisão existente entre o direito à propriedade intelectual e o segredo de fábrica, de um lado, e o direito à transparência algorítmica, à fundamentação algorítmica e à não discriminação – quando essa for ocasionada diretamente pela ausência de claridade de sua programação e de sua base de dados. Algo a ser respondido a partir da proporcionalidade entre a proibição de proteção insuficiente (*Untermaßverbot*) e de "proibição de excesso" na restrição ou limitação (*Übermaßverbot*) (Leal; Maas, 2020, p. 50).

Outro ponto crucial a se alinhar é **iii)** o cuidado com a ética e a responsabilidade social das empresas de desenvolvimento para com o resultado de seus produtos. Ter práticas de ESG (*environmental, social and governance*), p.e., pode vir a ter um papel importante nos rumos do desenvolvimento da programação e dos algoritmos, ao se trazer maior diversidade às equipes de programação, comungando e somando experiências sociais e geográficas de diferentes origens e perspectivas, levando em consideração a maior soma possível da diversidade existente na sociedade.

Os três pontos sinalizados supra comungam para a necessidade da regulamentação do desenvolvimento e do uso dos algoritmos, em processos que sejam auditáveis, transparentes e que tenham o condão de promover o direito à não discriminação algorítmica. Deverá ser encontrado um equilíbrio entre a valorização dos avanços tecnológicos e das inovações pelas empresas, resguardando seus interesses comerciais, ao mesmo tempo em que se tenha a valorização do ser humano, não enquanto um fornecedor (semi-inconsciente) de dados, mas enquanto utilizador e consumidor final dos algoritmos. O ser deve ser tratado como um fim em si, e não como um meio para o melhoramento da máquina.

Dessa forma, tem-se que a discriminação algorítmica é uma preocupação jurídica atual e que seguirá para o futuro próximo. Algoritmos e inteligências artificiais têm especial facilidade em se tornar relevantes e integrados na vida humana do cotidiano. O fato de que podem perpetuar preconceitos e estereótipos, levando a decisões injustas e desiguais, é fato relevante o suficiente para que o embate jurídico e o combate à discriminação estejam autorizados a existir. Além disso, é importante garantir a qualidade dos dados, a transparência dos algoritmos, a diversidade nas equipes de desenvolvimento, a responsabilidade das empresas para que se chegue a uma fórmula de regulamentação e de regulação adequada à (nova) realidade promovida pelos algoritmos.

Assim, para corretamente compreender a forma com a qual os algoritmos funcionam e qual a dinâmica discriminatória de sua atuação no mundo, se deve voltar um olhar atento às formas de funcionamento dos algoritmos e da sua tomada de decisão automatizada. Definir em mínimos conceitos acerca de como se estrutura a lógica por detrás da identificação dos *inputs* e as entregas dos *outcomes* é um passo ainda superficial, o que se deve ter em maior atenção e zelo é a dinâmica da construção das respostas pedidas às Inteligências Artificiais a partir das suas programações e dos insumos disponíveis (os dados armazenados e a base de treino datológica destas Inteligências Artificiais).

À luz da colisão enunciada, desloca-se o problema da hipótese para o método de como, a partir da ordem constitucional, deverá operar-se as decisões para converter tais parâmetros em decisões cuja razão pública seja legítima. A eficácia do modelo depende de uma jurisdição constitucional capaz de estabilizar o sentido interpretativo alinhado ao sentido normativo do direito, sobretudo dos direitos fundamentais, isto é, fixar limiares de escrutínio conforme o risco, distribuir ônus argumentativos e probatórios, densificar o dever estatal de proteção e controlar, com racionalidade prática, a extensão admissível do segredo frente às exigências de transparência e não discriminação. O passo seguinte – e apenas assim – consiste em mostrar como a jurisdição, por meio deste modelo proporcional, transforma princípios em critérios aplicáveis, produzindo previsibilidade institucional e cultura de justificação. É essa passagem do desenho teórico-operativo à praxis jurisdicional que se passa agora.

5.2 Jurisdição constitucional frente à realidade das discriminações algorítmicas: a formatação do sentido interpretativo alinhado ao sentido normativo do Direito

A jurisdição constitucional, chamada a tutelar direitos em ambiente automatizado, deve articular coerência interna (do caso) e coerência externa (do sistema), convertendo indícios técnicos em padrões de decisão. O parâmetro é a proporcionalidade material: quanto maior o impacto e a vulnerabilidade, maior o ônus explicativo do operador e mais intensa a abertura controlada das caixas-pretas, sem anular, por isso só, a fruição e a efetividade dos direitos da desenvolvedora, como o segredo comercial-algorítmico e a livre iniciativa. O que deve ser considerado em uma análise casuística.

A jurisdição constitucional, no entanto, tem compromisso reforçado com a proteção de grupos vulneráveis, o que decorre da dignidade da pessoa humana, da igualdade – em vertente material, evitando-se discriminações –, de uma jurisprudência que admite deveres positivos de proteção e medidas estruturais quanto à vulnerabilidade. No novo cenário algorítmico, esse dever reforçado se traduz em escrutínio ainda mais denso, na medida necessária para prevenir ou reparar discriminações injustificadas e produzir razões públicas verificáveis, que deverão nortear a interpretação e aplicação do direito nas hipóteses.

O direito manifesta-se argumentativamente, em problemas práticos, através de sua racionalidade construtiva. A materialização do sentido normativo-jurídico se dá a partir da definitividade decisória dos juízes, a qual deve manifestar, em um raciocínio adequado, corrigido e justificado, o mandamento que imperará mudanças no mundo fático. No entanto, quando muitas decisões são demandas do Poder Judiciário sobre questões que lidam com questões relevantemente semelhantes ou sobre mesmas bases principiológicas, um outro dever surge anexo ao dever judicial de motivação e fundamentação da prática argumentativa de decidir: o dever de coerência, tanto a coerência narrativa (de correção interna da argumentação da decisão em si) quanto da coerência normativa (de correção e adequação externa da argumentação dispendida na decisão).

Desse modo, percebe-se que a prática jurisdicional é um resolver de problemas práticos discursivamente desenvolvido a partir do oferecimento judicante

de razões, fundamentos e condicionantes que devem ter o condão de não apenas conduzir a respostas conclusivas do que deve ser feito no mundo real, apresentadas de modo racional e lógico, isto é, que decorram da fundamentação apresentada, mas também que estejam alinhadas com a perspectiva jurídica externa do discurso jurídico judicativo que pretende ver a lide solucionada. Ao primeiro momento denomina-se justificação interna da decisão judicial, ao passo que o segundo trata da justificação externa da motivação judicial.

Esses dois momentos justificativos da decisão judicial devem estar presentes no *decisium* para que ele possa ser considerado minimamente justificado. Assim, juízes e tribunais, por mais livres que pareçam estar para decidir sobre questões que lhes cabe competentemente resolver, estão adstritos a algumas questões, como à moldura normativa e à margem de apreciação, ao requisito teórico-metodológico da coerência normativa, bem como à material intencionalidade da concretização de um justo teleologicamente intencionado (Fuller, 1963, p. 145).

A partir desse duplo dever de justificação, a jurisdição constitucional, em ambiente de decisões automatizadas, deve articular a coerência narrativa do caso concreto com a coerência normativa do sistema, de modo a calibrar a intensidade do controle sobre caixas-pretas algorítmicas. O parâmetro operativo é a proporcionalidade, em sua dimensão material (lei da ponderação), a permitir que, constatada afetação relevante a direitos fundamentais de pessoas em situação de vulnerabilidade – mormente quando houver indícios de discriminação estrutural replicada por modelos de Inteligência Artificial –, as exigências de transparência e de explicabilidade (para além dos moldes dos art. 6º e art. 20 da LGPD) prevaleçam sobre o segredo comercial na medida estritamente necessária à reconstituição do itinerário lógico-decisório, sem esvaziar indevidamente a livre iniciativa e a concorrência. Esse juízo, caso a caso, evita pesos absolutos *a priori* e fornece standards replicáveis para decisões futuras, contribuindo para a estabilização do sentido jurídico da tutela antidiscriminatória em ambiente algorítmico.

Em termos de técnica decisória, a fundamentação exige dupla ancoragem – justificativa interna do caso e justificativa externa de compatibilidade normativa –, sob pena de nulidade. O julgador deve explicitar, com precisão, o objeto da ponderação, as premissas fáticas reputadas verdadeiras, as alternativas menos gravosas efetivamente consideradas e as razões de prevalência, nos termos do art. 93, IX, da

Constituição e do art. 489, § 1º e § 2º, do CPC. Esse roteiro não é formalismo de superfície: é condição epistêmica da decisão, sobretudo quando o resultado depende de processos de inferência opacos. Eliot (1934) pergunta onde está a sabedoria que perdemos no conhecimento – um lembrete de que explicabilidade não é só revelar parâmetros, mas dar razões normativamente relevantes. A decisão que se limita a invocar “eficiência”, “segurança” ou “segredo” – sem demonstrar a idoneidade, a necessidade e a proporcionalidade em sentido estrito – incorre em déficit de justificação e viola a cultura da motivação – entendida como a prática cultural-institucional de oferecer razões públicas que sejam reconstruíveis, contestáveis e justificadas sobre sólidas bases fáticas, técnicas e normativas – (Paulo, 2023).

Sistemas que decidem de modo opaco excluem os atingidos – e, não raro, os próprios técnicos – da possibilidade de recompor a cadeia causal dos erros e desvios cometidos. Essa assimetria informacional agrava estruturas de discriminações, amplificando-as de modo automatizado. À jurisdição, portanto, cumpre mapear os conteúdos constitucionais relevantes e exigir demonstrações verificáveis das razões técnicas, sob pena de insuficiência de motivação.

Esse encargo, contudo, encontra um obstáculo relevante: os marcos normativos atuais carecem de critérios substantivos para orientar a ponderação entre valores em colisão. A LGPD, por exemplo, ancora-se em princípios como finalidade, adequação, minimização e não discriminação, mas não fornece métricas de integração proporcional entre direitos em conflito. O mesmo déficit se observa nas resoluções administrativas e nos projetos legislativos em trâmite, que indicam riscos e impõem deveres de transparência, mas não estabelecem metodologias claras para graduar intensidade de danos, vulnerabilidade dos grupos atingidos e extensão dos efeitos. É nesse espaço de ausência – entre comandos gerais e situações concretas – que emerge a necessidade de um modelo decisório racional, capaz de converter razões normativas dispersas em parâmetros verificáveis de tutela. O deslocamento, portanto, é inevitável: da normatividade descritiva para a construção de critérios de justificação constitucional, mediante a lei material da ponderação.

Essa ausência de critérios substantivos impõe ao direito constitucional a tarefa de estruturar o espaço normativo de proteção. A experiência demonstra que, diante da opacidade técnica e da assimetria de poder informacional, apenas uma hermenêutica constitucional orientada pela proporcionalidade material pode oferecer

respostas adequadas. A questão já não é apenas de compliance com a proteção de dados, mas de justiça algorítmica: garantir que as tecnologias não sejam instrumentos de replicação de desigualdades, mas submetidas a filtros de racionalidade jurídica capazes de integrar, em graus variados, a tutela dos direitos fundamentais em conflito. É a partir desse movimento que se justifica a busca, no capítulo seguinte, por um modelo decisório racional, apto a graduar efetividades e a estruturar soluções constitucionais diante dos dilemas postos pela era da discriminação algorítmica.

Para enfrentar a assimetria informacional típica dos sistemas algorítmicos, impõe-se a distribuição dinâmica do ônus argumentativo e probatório. Quanto maior a opacidade residual e o poder de configuração do operador, maior o encargo de trazer aos autos documentação técnica mínima – relatórios de validação, métricas de desempenho estratificadas, logs relevantes, descrição dos critérios decisórios – apta a permitir contraditório qualificado. Trata-se de concretização da lei epistêmica da ponderação: sem confiabilidade das premissas, não há racionalidade na restrição de direitos. Em litígios com hipossuficiência técnica da parte atingida – muitas vezes em situação de vulnerabilidade interseccional –, a dinamização do ônus informacional deixa de ser faculdade para converter-se em exigência estrutural do devido processo.

Outro ponto de muita importância na atual sistemática de pensamento acerca da discriminação algorítmica diz respeito à responsabilidade dos desenvolvedores de inteligência artificial (IA), no que diz respeito ao risco algorítmico associado de cada aplicação. Ainda não há uma regulamentação específica para a responsabilidade civil por danos causados por Inteligência Artificial no Brasil, mas a Lei Geral de Proteção de Dados (LGPD) estabelece princípios e diretrizes que podem ser aplicados a essa questão. Hoje, os princípios da segurança e da prevenção da LGPD, aliados à competência regulamentar da Autoridade Nacional de Proteção de Dados (ANPD), permitem que sejam fixados padrões de mínima diligência capazes de estabilizar expectativas e possibilitar a responsabilização nos marcos do art. 42 da LGPD, para o mau uso ou conservação dos dados, como no caso de eles serem vendidos, vazados ou usados especificamente para um final danoso, não abarcando a matriz de risco sobre uso das Inteligências Artificiais. Além disso, as pessoas desenvolvedoras devem garantir que os sistemas de Inteligência Artificial sejam seguros e resistentes a ataques cibernéticos, garantindo que não sejam manipulados ou hackeados por

terceiros que podem manipular o funcionamento do sistema, como se vê, distante do que se é necessário a partir da Inteligência Artificial e de algoritmos.

Mas, na abordagem do risco, pode-se modular o conteúdo das normas de acordo com a intensidade dos riscos criados pelos sistemas de Inteligência Artificial, previsões trazidas aos Arts. 13 a 18 e 22 a 26 do Projeto de Lei nº 2.338/2023. A realização de avaliações de impacto algorítmico (AIA) serve para categorizar o risco em Risco Excessivo – cuja implementação é vedada –, Alto Risco e as que não apresentam riscos classificadas nas duas categorias anteriores. A AIA é composta por várias etapas que visam avaliar e mitigar, preventivamente, potenciais impactos negativos decorrentes da exploração de atividades de risco. O PL também prevê a criação de um órgão regulador específico para a Inteligência Artificial, que terá como função principal fiscalizar o uso desses sistemas e garantir que eles sejam desenvolvidos e utilizados em conformidade com a regulação, a ANPD pode vir a, eventualmente, assumir esse papel.

Acerca da responsabilidade, o PL nº 2.338/2023 traz em seu texto original, no Art. 5º, inciso V, a previsão do direito à não-discriminação e à correção de vieses discriminatórios diretos, indiretos, ilegais ou abusivos. E traz um regramento maior no Art. 12, prevendo que as pessoas afetadas, inclusive em **i)** decorrência do uso de dados pessoais sensíveis ou de impactos desproporcionais em razão de características pessoais como origem geográfica, raça, cor ou etnia, gênero, orientação sexual, classe socioeconômica, idade, deficiência, religião ou opiniões políticas; **ii)** função do estabelecimento de desvantagens ou agravamento da situação de vulnerabilidade de pessoas pertencentes a um grupo específico, ainda que se utilizem critérios aparentemente neutros; terá, salvo se demonstrado pela empresa que a "diferenciação realizada se dê em função de objetivos ou justificativas demonstradas, razoáveis e legítimas à luz do direito à igualdade e dos demais direitos fundamentais" (previsão do Parágrafo único do Art. 12), incidente a matriz de responsabilidade civil do Art. 27 e ss., sendo seu direito ver reparados os danos ocasionados, seja de natureza patrimonial, moral, individual ou coletiva.

Para a aferição da natureza quanto à responsabilidade civil, os §§1º e 2º do Art. 27 disciplinarão que ao se tratar de inteligência artificial de alto risco ou de risco excessivo, o fornecedor ou operador respondem objetivamente pelos danos causados, na medida de sua participação no dano, e quando não se tratar de

aplicação de alto risco, a culpa do agente causador do dano será presumida, aplicando-se a inversão do ônus da prova em favor da vítima. Desse modo, percebe-se que os desenvolvedores têm um dever de cautela que decorre da perspectiva mais econômica possível, a produção antecipada de provas – seja pela XAI ou por outro meio – para conseguir desabonar-se de eventuais lides que venha a integrar.

Em resumo, o PL 2338/2023 busca estabelecer um marco regulatório para a Inteligência Artificial no Brasil, com o objetivo de promover uma aplicação segura e responsável desses sistemas, seja pelo controle prévio, com a AIA, seja pelo controle administrativo, pela ANPD ou nova autoridade/agência, ou então pela via da reparação via poder judiciário. Não é de se olvidar, também, que mesmo após aprovado o Marco Regulatório da Inteligência Artificial, o poder judiciário, sobretudo na figura do Supremo Tribunal Federal, não deixará de ser atuante na definição da interpretação e aplicação do direito e, sobretudo, da nova lei a partir de uma leitura constitucional.

O art. 17 da Resolução 615/2025 representa um avanço no desenho de governança da inteligência artificial no Judiciário, no que se refere à avaliação de risco e impacto algorítmico. Ele determina que a atualização da categorização de riscos considere um conjunto de critérios materiais que vão além de parâmetros abstratos. Entre eles estão o impacto sobre direitos fundamentais e acesso a serviços essenciais, a ocorrência de danos materiais ou morais, inclusive por discriminações diretas ou indiretas, a repercussão diferenciada sobre grupos vulneráveis e a irreversibilidade dos efeitos produzidos. Soma-se a isso a avaliação do histórico de responsabilização de cada sistema, a transparência e a auditabilidade efetivas do modelo, bem como o grau de identificabilidade de titulares de dados quando múltiplas bases são combinadas. Trata-se de um salto qualitativo: a opacidade deixa de ser tolerada como falha técnica e passa a ser reconhecida como fator jurídico de risco, legitimando medidas de correção e até a suspensão de sistemas persistentes na sua falta de explicabilidade.

Esse filtro do art. 17 não atua de modo isolado, mas acoplado a um procedimento escalonado. A primeira etapa recai sobre o próprio tribunal que desenvolve ou contrata a aplicação, pois deve realizar avaliação inicial de riscos. Em seguida, a classificação em níveis de risco determina obrigações reforçadas de monitoramento e auditoria. Quando uma solução é classificada como de “Alto Risco”,

deve ser submetida a avaliação de impacto algorítmico. O procedimento desenhado funciona como coluna vertebral da governança: à categorização do risco se acoplam deveres de transparência proporcionais ao grau de perigo detectado, como critério definidor da atuação.

A avaliação de impacto algorítmico, por sua vez, recebe densidade normativa expressiva. Ela deve ser contínua, sujeita a revisões periódicas e acompanhada por auditorias externas. Há previsão de participação social e dos órgãos da justiça e essenciais à justiça sempre que possível, com acesso a relatórios técnicos inclusive. O resultado das avaliações e eventuais medidas corretivas precisa ser publicizado em linguagem clara, o que reforça a *accountability* institucional e abre espaço para controle democrático.

Identificado impacto díspar estatisticamente relevante, o procedimento de governança não se esgota no monitoramento. Impõe-se um dever de busca ativa por alternativas menos discriminatórias, consistente em testar, documentar e comparar arranjos de modelagem (como supressão de variáveis (*proxy*)) antes da manutenção do *status quo*. Esse dever é a tradução operacional do subteste da necessidade: entre vários meios adequados, somente se legitima aquele que impõe a menor restrição possível. Eventual decisão de manter um critério mais gravoso deve vir acompanhada de justificativa técnica robusta, sujeita a controle externo.

Apesar dos avanços, permanecem insuficiências estruturais. As categorias decisórias ainda se apoiam em expressões vagas como “alto potencial danoso” ou “baixa transparência”, sem indicar metodologias públicas de mensuração ou padrões probatórios uniformes. Esse espaço interpretativo pode gerar assimetrias entre tribunais, justamente nos contextos de maior sensibilidade. Além disso, o § 2º do art. 17 condiciona a adoção de medidas corretivas à “possibilidade técnica”. Esse limite, embora razoável em tese, acaba servindo de válvula de escape em casos em que a indisponibilidade decorre de modelos comerciais fechados e não de limitações científicas. Sem uma regra clara de inversão do ônus argumentativo – exigindo que fornecedores justifiquem a alegada impossibilidade – corre-se o risco de a opacidade se converter em escudo protetivo para o próprio sistema que gera risco.

Além disso, verifica-se que não há uma incorporação de elementos de combate à discriminação estrutural. Até se reconhece riscos e vulnerabilidade, mas a Resolução não a internaliza de modo operacional. As desigualdades históricas que

moldam os resultados algorítmicos até podem ser reconhecidos dentro de métricas de “alto potencial danoso” ou “baixa transparência”, mas há que se fazer isso de modo direto. Ao tratar a reprodução de assimetrias como uma exceção probatória, a ser verificada no caso em concreto, a opacidade tende a proteger o próprio sistema que reproduz discriminações, a não ser que busque ativamente incorporar o cenário de combate à estrutura da desigualdade.

O núcleo do controle previsto no art. 17, portanto, recai sobre ponderações complexas à luz de critérios múltiplos, mas sem metodologia definida de como integrá-los. É nesse espaço entre o veto categorial e a autorização condicionada que emerge a necessidade de um modelo decisório racional, capaz de graduar a efetividade dos direitos fundamentais em colisão. A Resolução, nesse artigo, fornece um léxico de razões públicas, mas não um método de articulação dessas razões em função de intensidade proporcionais ao risco. Faltam instrumentos que transformem, por exemplo, a baixa explicabilidade em barreira progressiva ao uso, ou que tratem a elevada identificabilidade como gatilho automático de revisão humana prévia. Do mesmo modo, seria necessário adotar matrizes de impacto que combinem a intensidade do dano, a extensão do público afetado e a assimetria informacional, para produzir graus de tutela mais finos do que as atuais respostas binárias.

Propõe-se, portanto, que a categorização de risco do art. 17 seja acoplada a um módulo obrigatório de proporcionalidade operativa. Cada classe de risco desencadeia deveres graduados de justificação: quanto maior o risco, mais denso deve ser o ônus de explicabilidade, testes de alternativas e ponderação documentada entre benefício e custo, com auditoria externa e participação social. A opacidade, aqui, é tratada como fator jurídico de risco que autoriza medidas corretivas e suspensão de sistemas persistentes na falta de explicabilidade.

Para garantir controle judicial e administrativo efetivo, propõe-se um protocolo de proporcionalidade para sistemas algorítmicos. Esse protocolo vai além da categorização de riscos e detalha etapas essenciais, como definir o objetivo legítimo do sistema com dados, mapear critérios e variáveis com documentação transparente, testar empiricamente o impacto dos critérios por meio de simulações A/B, explorar alternativas menos prejudiciais e incorporar monitoramento contínuo para identificar discriminações. O sistema deve demonstrar que utiliza a opção menos gravosa possível para os direitos individuais, documentando escolhas e evidências. Por fim, é

necessário registrar e justificar publicamente o balanço entre ganhos de eficiência e custos distributivos, garantindo transparência em todas as decisões e métodos empregados.

Como regra prática, a motivação deve acompanhar o grau de abertura informacional. Em A1, basta a entrega documental robusta e a demonstração pelo operador de que alternativas menos intrusivas foram testadas; em A2, acresce-se a exposição, sob sigilo, de auditoria e amostras, com justificativa explícita sobre por que o *trade-off* proposto supera o custo social; em A3, a decisão somente se legitima com perícia assistida e encadeamento probatório que confira alta confiabilidade às premissas colhidas pelo profissional e levadas aos autos para apreciação das partes (para fins de contraditório) e do juízo[13]. Em todos os níveis, compete ao julgador articular, de modo transparente, a passagem pelos subcritérios de finalidade, idoneidade, necessidade e proporcionalidade em sentido estrito – evitando tanto a tecnocracia acrítica quanto a retórica vazia de princípios.

Como visto, não há legislação específica para lidar com a discriminação algorítmica com suficiência em consideração ao caso em concreto, ou seja, *a posteriori*, diante de uma discriminação. Eis o *lócus* do Direito. Com base no dever de proteção e com bases nas linhas técnicas colhidas das áreas da ciência e da engenharia de dados, é possível defender ações positivas de programadores e controladores de dados para mitigar e/ou erradicar a falta de cuidado para com os resultados socialmente prejudiciais que advêm do processamento puro e simples de algoritmos sem tal preocupação.

Nesse contexto, a criação de regulamentações específicas para o uso de algoritmos em decisões importantes e o estabelecimento de critérios claros para o desenvolvimento desses sistemas são medidas essenciais para garantir a justiciabilidade dessa nova realidade. Afinal, é fundamental considerar as dimensões

[13] O que deverá ser feito aos moldes do apontado racionalmente à luz do voto da Ministra Nancy Andrighi no REsp nº 2.124.424/SP, a partir de um uso analógico da figura do perito, administrador ou inventariante digital, isto é, pessoa determina que em um incidente processual próprio, sob sigilo, poderá acessar, no caso mencionado, os bens digitais do falecido, que reportará aquilo unicamente demandado pelo julgador para fins de identificação de interesses patrimoniais transmissíveis. Os de cunho pessoal não poderiam ser abertos ou reportados. Analogicamente, ao caso do presente esforço, considera-se que a mesma lógica pode ser aplicada à sindicabilidade da explicabilidade algorítmica. No contexto da discriminação algorítmica o perito especializado deveria acessar os sistemas de forma controlada, reportando ao juízo apenas os achados estritamente relevantes ao caso, evitando o devassamento de informações de cunho essencial à competitividade de engenharia de software ou de códigos da Ré.

humanas em que a privacidade e os direitos antidiscriminatórios possam ser afetados pelo processamento de dados, o mau uso, e o uso mal-intencionado destas ferramentas; daí a importância da tutela à autodeterminação informativa, como escudo e controle pelos indivíduos perante as decisões automatizadas que se valem de seus dados.

Existir uma lacuna legislativa para colmatar as consequências discriminatórias da atuação dos algoritmos na sociedade não significa necessariamente a ausência de normas jurídicas que poderiam conduzir hermeneuticamente a soluções. Nesse compasso, é imperioso o debruçar-se sobre a construção de um raciocínio guiado pela eficácia dos direitos fundamentais que possam demonstrar assertividade e proporcionalidade na condução das tentativas judicantes de solver o problema da discriminação algorítmica.

E é exatamente nesse ponto, portanto, que fica ainda mais clara a importante tarefa a ser exercida pelo poder judiciário e pela jurisdição constitucional. Há vácuo normativo-protetivo em pontos sensíveis do ciclo de vida algorítmico, mas não há vazio jurídico: a Constituição – pela igualdade material, pelo devido processo e pela exigência de motivação adequada – e a LGPD – pela prevenção, não discriminação, responsabilização e revisão de decisões automatizadas – já autorizam e exigem uma resposta jurisdicional orientada à eficácia dos direitos. Incide aqui o dever estatal de proteção na sua dimensão positiva: diante de riscos previsíveis e de assimetrias informacionais próprias de modelos opacos, compete à jurisdição constitucional condicionar a interpretação acerca do uso de sistemas automatizados a salvaguardas verificáveis, sob pena de insuficiente proteção na ocorrência de violações a direitos.

Para além de assimetrias informacionais, assiste-se à privatização da própria capacidade social de aprender acerca de como verdadeiramente os dados pessoais são utilizados por ferramentas algorítmicas, e qual o risco disso. Trata-se de um deslocamento de poder que redistribui a própria privacidade como faculdade decisória, concentrando-a na infraestrutura de plataformas e seus sistemas de previsão e modulação comportamental. O poder instrumentário opera por economias de movimentação e transformação social. O resultado é a conversão de opacidades técnicas em dominação estrutural silenciosa, com efeitos discriminatórios e antidemocráticos sobre quem decide e sobre o que pode ser decidido. Se o conflito não é meramente entre segredo algorítmico e transparência algorítmica, em um macro

quase-extrajurídico, se dá entre a configuração de um novo poder e a salvaguarda dos direitos fundamentais (Zuboff, 2019, p. 189-222).

Esse condicionamento, por sua vez, não implica substituição do gestor tecnológico pelo juiz, mas elevação calibrada do escrutínio conforme o risco e a vulnerabilidade dos atingidos: quanto maior a opacidade técnica e o potencial de reprodução de discriminações estruturais, maior o ônus argumentativo de quem opera o sistema. Em termos operativos, isso se traduz em: **i)** inversão dinâmica do ônus informacional quando a parte detém exclusividade de acesso a dados, parâmetros e trilhas decisórias – e daí, portanto, o dever necessário da explicabilidade, seja pela XAI ou por outros métodos; **ii)** exigência de RIPD e AIA prévia e de registros auditáveis (*logs*, *datasets* de treino e *feedbacks*) aptos a reconstruir a cadeia causal da decisão – reconstituindo o processo lógico-artificial derivativo do *output*; **iii)** supervisão humana significativa, com competência e autonomia para intervir e reverter a decisão a qualquer momento – o que não exclui o risco do viés de confirmação e homologação, inclusive alerta para tal risco; e **iv)** soluções graduais e proporcionais – *disclosure* qualificado sob sigilo, auditoria independente, correções técnicas com prazos e metas, até suspensão condicionada em casos de alto dano, ainda que não fosse de alto risco. O ângulo exato de abertura da caixa-preta deve ser determinado com base em uma análise proporcional entre o bem afetado pela discriminação automatizada e o grau de complexidade – ou dificuldade explicativa – associado à identificação do motivo decisório empregado pela máquina na decisão discriminatória. O segredo algorítmico permanece tutelado, mas cede na medida estritamente necessária para viabilizar contraditório efetivo e a ponderação informada, mediante acesso controlado (perícia assistida, depósito em juízo, compromissos de confidencialidade), de modo a compatibilizar a proteção concorrencial e a proteção de direitos. Efetivando a máxima dos direitos em colisão.

Para estabilizar expectativas e evitar tanto deferência acrítica quanto ativismo descoordenado, deve-se verificar a proporcionalidade informacional, isto é, o grau de abertura do código deve corresponder à escala de risco algorítmico, de vulnerabilidade do sistema, ou de danos causados efetivamente. Com isso, fecha-se a ponte entre a moldura constitucional e infraconstitucional já vigente e a execução prática da tutela antidiscriminatória. Necessário agora verificar a colisão enunciada: transparência algorítmica, e direitos associados, *versus* segredo algorítmico, e direitos

associados. A partir desta colisão de eficácias jusfundamentais verificar-se-á a lei material da ponderação para sua definição e para a gradação do ônus informacional a desencadear a racionalidade acerca da resposta do ângulo de abertura da aplicação algorítmica.

A abertura deverá seguir ao menos três níveis proporcionais, que futuramente, diante das necessidades, poderá ser ampliado: A1 (documental) – entrega de políticas da empresa, protocolos de manejo, descrição do fluxo decisório e métricas globais, rodada de testes, relatórios de segurança técnica da aplicação algorítmica, etc.; A2 (técnico-controlado) – acesso sob sigilo a *logs* de uso relacionados ao(s) episódio(s) discriminatório(s), amostras de dados, critérios de decisão, relatórios de validação, etc.; A3 (de profunda análise) – acesso assistido determinado pelo Poder Judiciário a investigação em toda a aplicação algorítmica, com segredo judicial, parâmetros e código relevante ao resultado impugnado, com encadeamento das evidências e confidencialidade. A escolha do nível deverá decorrer do binômio gravidade do impacto discriminatório e opacidade residual persistente – isto é, após exercidos os direitos imediatos de revisão e explicabilidade administrativamente –, à luz do teste da proporcionalidade.

A abertura graduada observa quatro gatilhos cumulativos: (i) intensidade do impacto; (ii) posição de vulnerabilidade/interseccionalidade; (iii) opacidade residual após explicações voluntárias; (iv) assimetria informacional/posição de mercado — e se traduz em níveis proporcionais:

> **A1 (documental)** – políticas, fluxos decisórios, métricas globais e relatórios de validação;
> **A2 (técnico-controlado)** – acesso sob sigilo a logs e amostras, critérios de decisão e testes específicos;
> **A3 (profunda análise)** – perícia assistida sob segredo judicial sobre parâmetros e código relevante ao resultado impugnado, com encadeamento probatório.

A escolha do nível decorre da proporcionalidade e deve ser motivada nos termos do art. 93, IX, CF e do art. 489, § 1º e § 2º, CPC. A opção entre A1–A3 observará cumulativamente o teste de proporcionalidade, a gravidade e extensão do impacto, a vulnerabilidade dos grupos afetados, a opacidade residual após explicações e *logs* fornecidos; a posição dominante do operador (empresa) e a viabilidade de proteção do segredo por meios menos intrusivos, como perícia técnica, respeitando o subteste da necessidade, como se verá.

O arco está traçado: entre a opacidade que fere e a transparência que tutela, a Constituição reclama razões verificáveis e meios proporcionais. A abertura graduada (A1–A3) não é um gesto de curiosidade técnica, mas a tradução institucional do velho imperativo do constitucionalismo: limitar o poder onde quer que ele se concentre – inclusive quando se disfarça em códigos e dados. Se a discriminação se aloja em correlações obscuras, a resposta não pode ser cega: exige desenho, evidência e responsabilidade.

A motivação deve ser reforçada, demonstrando a finalidade constitucionalmente legítima, a idoneidade da medida escolhida, a necessidade, com exame de alternativas menos intrusivas, a proporcionalidade em sentido estrito, explicitando por que o ganho pretendido supera o custo antidiscriminatório, a documentação dos elementos fáticos que sustentam o grau de abertura (A1–A3) e, por fim, escrutinando o contraditório técnico e os meios de proteção do segredo judicial (fundamentado no segredo comercial) quando cabível e justificado.

Daqui em diante, o problema deixa de ser apenas o que a jurisdição pode exigir e passa a ser como a sociedade organiza a sua própria defesa. É nesse ponto que o sentido normativo se converte em arquitetura operacional jurídica, de viés constitucional. Em outras palavras: do dever em tese ao dever em prática. Se é da vontade que a tutela antidiscriminatória saia do papel, é preciso refletir acerca de uma possível proposta interpretativa, que parta do ponto de vista constitucional em direção à eficácia dos direitos. É com essa chave que se abre o raciocínio para um modelo decisório racional acerca do ângulo proporcional de abertura das caixas-pretas algorítmicas.

Rapidamente permita-se adentrar no eixo da argumentação acerca da proporcionalidade sobre os elementos a serem avaliados para essa graduação: **i)** a adequação exige que o operador demonstre a capacidade de que, naquele caso, a abertura escalonada surtirá efeito prático; **ii**) a necessidade exige que o operador demonstre que as alternativas são mais gravosas aos direitos da desenvolvedora; **iii)** a proporcionalidade em sentido estrito exige do juízo que a decisão pela gradação da escolha (A1-A3) esteja fundamentada em mínimos de cognoscibilidade exigíveis e justificáveis à luz de provas documentais ou, então, da necessidade da perícia algorítmica diante de inconsistências e lacunas a serem por ele reconstituídos à razão públicas, mitigando a opacidade, antes protegida pelo segredo algorítmico.

Quem detém os dados e as aplicações detém o ônus. Sem a entrega proporcional do fluxo, dados e métricas, o contraditório torna-se impossível. Dessa feita, o controle judicial deve exigir explicações auditáveis como condição de legitimidade da decisão automatizada. Mas o que deve ser justificado?

A parte responsável por desincumbir-se do ônus probatório e de explicabilidade deve disponibilizar, em escala proporcional ao risco, a ser verificada, elementos e evidências que possam descrever o funcionamento e o fluxo decisório da aplicação, a instrução de direcionamento da aplicação, se houver, a forma de utilização dos dados, e a própria base de dados em que a aplicação foi treinada, bem como outros parâmetros importantes de treinamento ou recebidos por *feedbacks*. Também é necessário que seja demonstrada, na maior medida do possível, sob pena de não restar alternativa a não ser realmente abrir a caixa-preta, elementos que permitam reconstruir a cadeia causal do episódio discriminatório, bem como se nesses casos havia validação humana a ser dada ao final, ou não. O viés homologatório, que é um viés humano, também deve ser encarado com seriedade nesse sentido.

O dever precisa ser ainda mais importante quando se estiver diante de uma grave opacidade, naquelas hipóteses em que nem os desenvolvedores sabem quão longe o *machine learning* levou a aplicação, também quando o impacto da aplicação for relevante a violação de direitos fundamentais e, com ainda maior importância, quando esses impactos a direitos fundamentais atingirem titulares pertencentes a grupos vulneráveis; nessa hipótese, será maior o ônus quanto maior for o grau de vulnerabilidade e de interseccionalidade. Por fim, aposta-se em uma também necessário embrutecimento do ônus *probandi* e de explicabilidade quanto houver recusa dos detentores do ônus de empreender em explicações iniciais, previstas na legislação.

No plano processual, duas ferramentas são imediatamente chamadas a dar efetividade ao dever de explicação. A inversão dinâmica do ônus da prova, quando o operador detém a informação indispensável, bem como a necessidade de que submeta a aplicação à perícia técnica, que a fará com registro da cadeia causal de procedimentos e etapas de verificação, assegurando acesso proporcional a dados sobre o uso da IA. A motivação judicial deve explicitar porque o grau A1–A3 é necessário e quais salvaguardas protegem o segredo.

A Lei Geral de Proteção de Dados (LGPD) foi criada como resposta ao compartilhamento de dados pessoais, mas, após quase cinco anos de sua promulgação, o foco parece estar mais na gestão de riscos sobre vazamento de dados do que na perspectiva da autodeterminação informativa. Assim, é preciso olhar para a tensão entre os direitos empresariais das empresas de tecnologia e os efeitos resultantes aos direitos dos usuários.

Com isso, surge a necessidade de afirmar a existência de um direito fundamental à transparência e à explicabilidade algorítmica como decorrência do direito à autodeterminação individual. Do outro campo, promovendo a proteção aos interesses dos programadores e das empresas de tecnologia, há o segredo comercial, fundamentado na livre iniciativa e na livre concorrência, originando o fenômeno das *black boxes*. As intersecções entre a explicabilidade e o segredo comercial, devem ser interpretados e aplicados a partir do escrutínio do teste da proporcionalidade, que poderá avaliar o choque aparente entre o direito fundamental à transparência algorítmica e o segredo comercial, considerando, em casos, o impacto de uma decisão automatizada na fluência de um direito fundamental, a importar no grau de intervenção e na possível relativização do segredo comercial para garantir a explicabilidade, cumprindo-se o artigo 20 da LGPD.

O § 2º do art. 20 da LGPD assegura que, não sendo prestadas informações claras e adequadas sobre os critérios e os procedimentos utilizados na decisão automatizada – observados os segredos comercial e industrial sobre a aplicação algorítmica –, a ANPD poderá realizar auditoria para verificação da conformidade e exigir as medidas cabíveis. Para tanto, contudo, é necessário que se reconheça, à luz do teste da proporcionalidade, que estão em colisões o direito fundamental associado da transparência algorítmica e da explicabilidade de um lado, vinculado à autodeterminação algorítmica, e de outro o direito fundamental associado do segredo comercial, vinculado à saudabilidade da vida financeira da empresa e à livre concorrência e à livre iniciativa. Avaliações que devem ser resolvidas caso a caso, sem a possibilidade de atribuição de pesos absolutos *a priori*. Ferramenta metodológica essa que deverá ser utilizada não só pelo poder judiciário, na repressão e atendimento a demandas propostos, mas também pelos poderes executivo e legislativo na avaliação de impacto jusfundamental enquanto legiferam sobre o tema.

Para tanto, necessário se faz construir um *iter* decisório. Ele começa pela identificação do bem jurídico afetado e da eventual vulnerabilidade do titular. Em seguida, descreve-se com precisão o resultado automatizado e os indícios de discriminação. Definido o quadro, fixa-se o grau de abertura inicial (A1) e as salvaguardas correspondentes. Se a explicação ofertada não bastar, amplia-se a abertura de modo proporcional, passando-se a A2 – diante da gravidade e da falta de resposta do operador pode ser determinado o nível A2 desde logo – e, se persistir a opacidade, a determinação justificada de aplicação do nível A3, com perícia e acesso controlado às informações indispensáveis, sob confidencialidade, quando possível. Em sendo necessário avançar, sob o comando do juízo responsável, o nível A3 admite adentramento ainda mais profundo na aplicação. Reconstituído o nexo causal entre dados, regras e resultado, sendo possível atribuir uma responsabilidade, adotam-se medidas proporcionais de correção e de prevenção de novas ocorrências e dão-se seguimentos às devidas repercussões.

Nesse sentido, a jurisdição constitucional, que tutelará tais direitos, tanto nesse primeiro momento quanto em última instância, deverá ser capaz de encontrar as bases para exigir, ainda que minimamente, a proteção a direitos fundamentais prejudicados pelas aplicações, com especial relevo para a descoberta das razões acerca do cometimento, ou mero acontecimento, de discriminações automatizadas. Há um dever jurídico intrinsicamente motivado por substratos morais que determina a proteção e a defesa de usuários em uma nova dinâmica de putativa soberania cibernética de empresas e desenvolvedoras de aplicações algorítmicas.

Além disso, é importante destacar que o teste da proporcionalidade pode ser usado para avaliar o equilíbrio entre a transparência e a explicabilidade algorítmica e o segredo comercial. Essa avaliação deve ser feita caso a caso, levando em conta as circunstâncias específicas de cada situação. A transparência e a explicabilidade dos algoritmos são fundamentais para garantir o uso ético e responsável da inteligência artificial, sem comprometer a inovação e o desenvolvimento tecnológico, e ambas devem ter sua efetivação buscada no máximo grau de sua possível satisfação. É preciso garantir a transparência e explicabilidade dos algoritmos para que essas ferramentas possam ser usadas de forma ética e responsável, sem que isso onere ou rompa com a lógica de mercado que promove os avanços tecnológicos em primeiro lugar; tal garantia não deve ser alcançada a qualquer custo.

A esta altura, percebe-se que o déficit não é apenas regulatório, mas metodológico-operativo, chamando, mais uma vez, a atuação do Direito pelo Poder Judiciário. Faltam critérios públicos de mensuração que integrem, de modo verificável, a intensidade do dano, a vulnerabilidade dos atingidos, a extensão dos efeitos e o grau de opacidade. Sem tais balizas, a regulação corre o risco de operar no vazio, e a abertura das caixas-pretas (A1–A3) fica sujeita a variações substanciais. Isso impõe um passo adicional no argumento, qual seja, o de compreender o ambiente jurídico-socio-técnico que repercute contra a opacidade e os vieses replicados, pois é nele que os critérios devem enraizar-se e mirar. Em outras palavras, antes de fixar pesos, é necessário reconhecer que os algoritmos herdam e cristalizam padrões históricos, e que a falta de métricas não é neutra, mas favorece a reprodução automatizada de desigualdades. É justamente por isso que é necessário verificar o papel real de transformação e interação das IAs para que se entenda o real poder de enfrentamento e transformação do Direito e da Jurisdição Constitucional em realizar exigências proporcionais e mensuráveis de explicabilidade, auditabilidade e correção de viés. O ponto aqui é da certeza epistêmica acerca da efetividade dessas exigências e da certeza normativa acerca da possibilidade prático-jurídica desta realização judicializada.

Conforme observado, existe uma constante relação dialética entre a realidade da experiência humana e as programações e ajustes realizados no universo algorítmico, não apenas pelos dados gerados, coletados e fornecidos continuamente, mas também pelas alterações promovidas pelo pensamento dos *prompters* (programadores). Observa-se um processo de algoritmização das atividades humanas, acompanhado pela atualização sistemática de programas e dados nos algoritmos. No entanto, é importante destacar que os algoritmos permanecem condicionados aos registros históricos das ações humanas.

De igual forma, um mundo algoritmizado é um mundo estagnado no passado. É um mundo conservador pela atuação do algoritmo e pela dependência humana ao algoritmo. Ainda que os dados sejam sempre renovados, uma confluência da cultura para com os algoritmos enquanto mecanismos altamente presentes no dia a dia dos seres humanos potencializa o viés estagnador e discriminatório. Cazuza nunca esteve tão presente (“Eu vejo o futuro repetir o passado, Eu vejo um museu de grandes novidades”), ou, como ironiza Machado de Assis na “Teoria do Medalhão” (2022),

quando a cultura transforma o lugar-comum em virtude pública e condecora a repetição: premiam-se a previsibilidade, a aparência de neutralidade e a segurança das platitudes – exatamente o que os algoritmos, treinados para otimizar o passado, tendem a reproduzir como norma, convertendo o enviesamento histórico em critério técnico de seleção e exclusão. Para além de enfrentar o problema da discriminação algorítmica, seja ela consciente ou inconsciente por parte dos programadores, além de exigir que as empresas procurem alterar suas culturas promovendo maior equilíbrio de diversidade em suas equipes e plantéis, é imperiosa a criação e implementação de leis e regulamentações que possam ajudar a garantir a transparência, a responsabilidade e a justiça no desenvolvimento e no uso de algoritmos, tendo por base os princípios e os direitos humanos e fundamentais diretamente relacionados à discriminação dos algoritmos.

A nova forma de discriminação, que na verdade é apenas uma roupagem automatizada dos próprios estigmas humanos ainda presentes, está renovando tratamentos explícitos que já deveriam ter sido erradicados, o que passa também por um distanciamento existente entre a ponta de criação das programações e os usuários, que em regra se dão na origem destas *Big Techs* e outras empresas de tecnologias do meio norte do mundo (Zuboff, 2019, p. 206). A colaboração entre empresas, governos, organizações e sociedade civil é fundamental para enfrentar os desafios da discriminação algorítmica, desde a criação de espaços de diálogo e cooperação para discutir e debater os efeitos dos algoritmos até a implementação de fóruns permanentes e de autoridades competentes – ainda que juntas à ANPD (Autoridade Nacional de Proteção de Dados) – para criar soluções e atender a demandas específicas de práticas algorítmicas discriminatórias a indivíduos, grupos ou comunidades.

No contexto da discussão sobre discriminação algorítmica, o STF tem um papel importante a desempenhar na definição de parâmetros de controle para o uso de algoritmos em decisões importantes que afetam os direitos dos cidadãos, pela via de acesso da autodeterminação, da igualdade e da não discriminação. Ele pode estabelecer diretrizes normativas importantes para orientar as decisões judiciais em casos futuros envolvendo discriminação algorítmica em outros tribunais, bem como no cenário latino-americano.

O STF pode ser chamado a julgar casos específicos relacionados à discriminação algorítmica, tanto advindos de indivíduos ou grupos prejudicados por decisões automatizadas tomadas com base em algoritmos discriminatórios ou enviesados, ou ainda chamado a decidir sobre omissões legislativas ou mesmo acerca da (in)constitucionalidade de novas leis e marcos regulatórios a serem propostos. O STF, assim, é importante *lócus* multiportas à defesa institucional dos direitos fundamentais dos cidadãos.

Para cumprir seu papel na proteção dos direitos fundamentais no contexto da discriminação algorítmica, o STF precisa estar atualizado sobre as últimas tendências tecnológicas e ter uma compreensão sólida das implicações éticas e regulatórias do uso de algoritmos. Além disso, é importante que o STF conte com corpo técnico sólido e atualizado, disposto a demonstrar os nuances e etapas do procedimento de criação, programação, divulgação e uso das ferramentas de Inteligência Artificial, bem como de que modo se dá a coleta, tratamento e uso em treino de dados para a atuação dos algoritmos e do *machine learning* dos mesmos. A colaboração com especialistas em tecnologia é necessária para garantir que as decisões tomadas sejam informadas e baseadas em evidências científicas.

Além disso, o STF também pode desempenhar um papel importante na definição de políticas públicas relacionadas ao uso de algoritmos. Isso pode incluir a criação de regulamentações para garantir a transparência e responsabilidade no desenvolvimento e uso de ferramentas algorítmicas, bem como a promoção da diversidade nas equipes de desenvolvimento destas tecnologias para evitar vieses e discriminações. O Supremo pode trabalhar em conjunto com outras instituições governamentais, organizações da sociedade civil e empresas para desenvolver políticas públicas que promovam o uso ético e responsável de algoritmos, tanto em âmbito público quanto em âmbito privado.

Além disso, não se trata de inovação institucional: o STF há muito induz políticas públicas por meio de decisões estruturantes e de coordenação interinstitucional. No sistema prisional, ao reconhecer o “estado de coisas inconstitucional” na ADPF 347. No campo climático, na ADPF 708, a Corte impôs a reativação e execução do Fundo Clima, com exigências de governança e planejamento, reordenando a política pública setorial. Na proteção de povos indígenas, a ADPF 709 determinou a elaboração e monitoramento de um plano de

enfrentamento à Covid-19, com sala de situação envolvendo governo e sociedade civil — um modelo explícito de governança dialogada. De modo análogo, na ADPF 828, que tratou das remoções forçadas durante a pandemia, a Corte modulou a atuação administrativa em contexto de vulnerabilidade habitacional, proibindo despejos coletivos e determinando a formulação de protocolos participativos e de mediação social, revelando a potencialidade do controle jurisdicional estrutural para a proteção de grupos vulneráveis (Leal; Paulo, 2024). Em matéria algorítmica, o raciocínio lógico é análogo: o Supremo Tribunal Federal pode fixar parâmetros mínimos e instaurar um arranjo cooperativo entre órgãos públicos, setor privado e sociedade civil, garantindo, por via dialógica, que os deveres de não discriminação se convertam em políticas operacionais efetivas, sobretudo quando direitos de pessoas e grupos em situação de vulnerabilidade estrutural estejam sendo amealhados.

É importante destacar que o papel do Tribunal Constitucional brasileiro na proteção dos direitos fundamentais no contexto da discriminação algorítmica não se limita ao Brasil. Como uma Corte de referência na América Latina, o STF pode desempenhar um papel importante na definição de padrões regionais para o uso de algoritmos em decisões importantes que afetam os direitos humanos. Isso pode incluir a colaboração com outras Cortes Constitucionais – diálogo judicial ou *cross-judicial fertilization* (Jacobs, 2003) – da região para desenvolver abordagens comuns para lidar com a discriminação algorítmica, ou mesmo a referências a *standards* protetivos já implementados pela Corte Interamericana de Direitos Humanos, em uma clara aplicação de uma conjectural aplicação das teorias de *Ius Constituticionale Commune* latino-americano (ICCAL), que se fazem especialmente necessárias para riscos que não são nacionalizados, mas que estão internacionalmente considerados e que merecem um olhar transnacional e transconstitucional atento.

A discriminação injustificada é um mal humano, o qual se pensava, com raras exceções de tristes episódios, já ter sido superado. A discriminação humana pode ser refletida para os algoritmos, mas não é somente essa que se deve combater. O grande desafio está no combate ao preconceito enviesado invisível, que apenas se concretiza enquanto discriminação via algoritmo na entrega de um resultado discriminatório pelas máquinas, como explorado. Suas raízes, no entanto, não estão no disco rígido de qualquer *software* ou *hardware*, mas no rastro de uma condição humana ainda presa

a visões injuriosas das diferenças; uma que não enxerga na pluralidade e na diversidade a real essência de ser humano.

Em suma, a jurisdição que exige explicações verificáveis e delimita a abertura por graus produz segurança jurídica para desenvolvedores e tutela efetiva e diferenciada para a proteção de vulneráveis. Com essa moldura, passa-se à aplicação escalonada (A1–A3) e aos critérios práticos da ponderação.

O dever de motivação constitucional ganha arquitetura operativa: i) distribuição dinâmica do ônus informacional; ii) critérios de auditoria e documentação proporcional ao risco; iii) supervisão humana significativa; iv) correções técnicas e acessos graduados para a verificação de falhas. Falta, porém, a régua decisória que converta essas razões em graus verificáveis de abertura. É precisamente o que o próximo tópico entrega: a lei material e epistêmica da ponderação acoplada à matriz A1–A3 como regra de decisão.

5.3 A ponderação material como baliza interpretativa para uma gradação correlacional entre restrição e fruição de direitos fundamentais: o valor interpretativo-balizador de casos acerca da tutela de direitos de jurisdicionados pertencentes a grupos vulneráveis

Consolidado o diagnóstico e delimitadas as tensões entre transparência e segredo algorítmico, cumpre agora explicitar o arquétipo decisório que orienta a solução racional de colisões – é aqui que a teoria alexyana e a evolução da lei do sopesamento entram em cena. O que se exige na prática é saber quanto abrir da caixa-preta. O escrutínio em tese não é complexo: precisa da medição exata entre o motivo algorítmico de determinada decisão automatizada injustificadamente discriminatória e o percentual de adentramento na complexidade do algoritmo e da IA para encontrar, ou a justificação de tal discriminação (a verificar posteriormente se lícita ou ilícita), ou a mera ocorrência de desvios dado o descuido e o risco da atividade. Nesse sentido, equaciona-se a transparência algorítmica com o segredo algorítmico no máximo compatível entre ambos, aportado em cada caso, em busca de uma razão justificadora da decisão automatizada.

Tomados esses pressupostos e situado o debate no horizonte da teoria da argumentação jurídica alexyana, cumpre explicitar o propósito que anima o autor.

Alexy busca um “direito adequado” – um direito que observe a legalidade positiva e, por meio de correções racionais, maximize sua eficácia social (Alexy, 2016). Essa busca decorre da dupla natureza do direito: uma dimensão real, que expressa a positividade e se preocupa com a eficácia jurídico-social; e uma dimensão ideal, voltada à correção de conteúdos e procedimentos – isto é, a exigências de moralidade e justiça (Alexy, 2016). Daí o não-positivismo alexyano e a centralidade da argumentação em um sistema dinâmico (e não estático) de regras e princípios. Nesse sistema, a proporcionalidade consolidou-se, ao longo de décadas, como técnica decisória de referência nas cortes constitucionais, notadamente para a tutela de direitos fundamentais em hard cases (Borowski, 2022a).

O exame da proporcionalidade – derivado da teoria dos princípios – tornou-se o “critério material mais importante para se avaliar reivindicações decorrentes de direitos fundamentais” (Borowski, 2022a). Compreender sua arquitetura requer acompanhar as fases de desenvolvimento da lei do sopesamento. Três grandes momentos se destacam (Trivisonno; Borowski, 2022): a cada etapa, a fórmula incorpora elementos antes desconsiderados, em diálogo com críticas que pressionam por maior controlabilidade racional. Enquanto a lei de colisão permanece estável – princípios colidem, não contradizem –, a lei da ponderação sofreu ajustes relevantes (Alexy, 2016). Como a finalidade primeira dos direitos fundamentais – inclusive quando veiculam bens coletivos – é a proteção do indivíduo (Sarlet, 2018), a estrutura de decisão deve minimizar “custos evitáveis” segundo a otimalidade de Pareto: preferir o estado A ao B sempre que a passagem não piorar a situação de nenhum participante e melhorar a de ao menos um (Tovar, 2022, p. 391). Daí a formulação segundo a qual, “quanto maior for o grau de não-satisfação ou de afetação de um princípio, tanto maior terá que ser a importância da satisfação do outro” – regra de decisão, e não procedimento generalizante (Sieckmann, 2011).

É precisamente para enfrentar o risco de discricionariedade epistêmica que a evolução do modelo incorpora, em metanível, uma dimensão de certeza das premissas. O acréscimo do componente epistêmico não altera o núcleo da lei da ponderação, mas explicita seu ônus de conhecimento: quanto mais intensa a restrição, maior deve ser a firmeza das bases empíricas e normativas que a sustentam (Alexy, 2022; Borowski, 2022b). O ponto nevrálgico está no suporte das razões: a

proporcionalidade só é tão boa quanto as premissas que a alimentam, e daí surge a necessidade de entendê-la verdadeiramente (Alexy, 2022, p. 115):

$$PeR_{i,j} = \frac{I_i \cdot PeA_i \cdot C_i^e \cdot C_i^n}{I_j \cdot PeA_j \cdot C_j^e \cdot C_j^n}$$

Esse componente epistêmico não deve ser compreendido apenas como exigência abstrata de rigor argumentativo, mas como condição de possibilidade da própria gradação decisória A1–A3. Sem a demonstração mínima de premissas fáticas confiáveis — seja por documentação técnica, relatórios de validação, registros de *logs* ou amostras de *datasets* —, não há como calibrar proporcionalmente a abertura da caixa-preta. A ponderação epistêmica, nesse sentido, evita que a decisão recaia em arbítrio ou em deferência acrítica à parte detentora do sistema, impondo que cada incremento de intensidade na intervenção corresponda a uma base probatória mais sólida. Trata-se, portanto, do ponto de ancoragem que reconduz a confiança social a critérios verificáveis e impede que a sindicabilidade constitucional se dissolva em retórica de princípios sem densidade empírica. A afetação deve ser, portanto, feita em graus, apoiadas na necessidade estrita de relativização.

Nesse quadro, o papel do perito digital/algorítmico adquire relevância especial: é ele quem fornece as bases técnicas que permitem ao julgador atribuir grau de certeza às premissas fáticas. Descrições das variáveis de programação e de algoritmização não são apenas insumos de análise, mas a matéria-prima sobre a qual se constrói a ponderação epistêmica. Sem essa mediação técnica, a gradação pretendida não passaria de abstração retórica sem base epistêmica. Com ela, converte-se em dado verificável e sindicável, capaz de sustentar a abertura proporcional da caixa-preta e evitar arbitrariedades.

Adota-se, assim, diretamente, a fórmula do peso completa refinada de Alexy, que agrega à intensidade das interferências (I) e ao peso abstrato dos princípios (PeA) a variável “C”, responsável por graduar a certeza das premissas relevantes do caso em dois planos – empírico e normativo – segundo a “escala da certeza”: “seguro ou certo (s), plausível (p) e não evidentemente falso (e)” (Alexy, 2022, p. 116). Sob a lei da ponderação epistêmica – “quanto mais grave for uma restrição a um direito fundamental, maior deve ser a certeza das premissas que apoiam a restrição” –, válida ceteris paribus, a preferência entre princípios deixa de depender apenas de I e PeA para exigir, adicionalmente, que o suporte fático e inferencial seja suficientemente

firme para sustentar a medida restritiva (Alexy, 2022; Borowski, 2022b). O resultado é uma regra de decisão que, no metanível da ponderação, calibra o ônus justificatório: a gravidade da restrição impõe elevação correlata do grau de certeza (C), tanto quanto permite que incertezas residuais – quando devidamente qualificadas – refreiem soluções maximalistas. As premissas empíricas são estruturantes no todo do modelo e – de modo ainda mais visível – na versão refinada: elas alimentam C no seu componente fático, condicionando o próprio espaço de precedência material. Se os fatos que sustentam a intervenção são apenas "plausíveis" ou "não evidentemente falsos", a fórmula exige contenção decisória; se são "seguros", autoriza-se maior compressão, sempre em correlação com I e PeA (Alexy, 2022; Klatt, 2021). Assim – e apenas assim – a ponderação reduz o risco de discricionariedade epistêmica, torna transparente o papel da prova e das inferências causais, e reconstrói, com racionalidade controlável, porque um princípio pode preponderar em face de outro em contextos de alta ou baixa certeza (Borowski, 2022b).

Para ligar a teoria à prática, cada subteste da proporcionalidade se apoia em parâmetros técnicos verificáveis. Primeiro a adequação, em que se pergunta: o sistema realmente cumpre o fim legítimo para os grupos afetados? Para tal verificação exige-se documentação clara do funcionamento da aplicação, indicadores de desempenho desagregados por grupo (*clusters*) e validação por fonte externa (*feedback* humano preferencialmente). Quando esses elementos mostram desempenho consistente e replicável, o grau de certeza sobe; quando faltam, ele cai. Se a medida de restrição (abertura do algoritmo) for adequada para o fim de dar transparência algorítmica (nota-se que *a priori* já temos essa clarificação; o que não possibilita deixar de realizar uma análise tópica aos casos), avalia-se a segunda etapa do teste, a necessidade, em que se pergunta se há meio menos gravoso com eficácia equivalente para privilegiar a transparência algorítmica. Dessa forma, impõe-se examinar e, quando possível, substituir variáveis indiretas (como bairro de residência ou intervalos de carreira), realizar testes de sensibilidade e comparar arranjos (*clusters*) de variáveis menos intrusivos. O grau de certeza aumenta quando essas alternativas são exploradas e documentadas; diminui quando a escolha de variáveis é opaca ou não testada. Se a medida eleita (abertura do algoritmo) se demonstrar adequada e necessária para o caso em análise – por hipótese –, parte-se para a última etapa do teste da proporcionalidade, a proporcionalidade em sentido estrito ou

ponderação, no qual se pergunta se o custo jusfundamental, social, individual, coletivo e empresarial (de efetivar ou não a medida de abertura dos algoritmos), imposto pelos critérios do sistema, compensa o benefício alegado? Requer-se o maior número de provas do itinerário lógico da tomada de decisão da aplicação, o que habilita o uso de técnicas de engenharia reversa e o uso de perícia. O grau de certeza sobre a necessidade da abertura se eleva quando o itinerário lógico não puder ser verificado, diante de inconsistências, lacunas ou segredo injustificados; regride diante de trilhas da decisão discriminatória. Nessa hipótese, não há necessidade de abertura do algoritmo, mas de reparação e responsabilização imediata, ainda que não se descarte a possibilidade de densificação das investigações da aplicação.

Quando o benefício de proteção de direitos – especialmente de pessoas e grupos vulneráveis – supera o custo da intrusão, não existirem meios menos gravosos e o itinerário lógico permanecer opaco, a abertura é devida no grau necessário, que se verificará *in casu* e poderá ser aplicado de modo escalonado, desde o A1 até o A3. Se, ao contrário, as trilhas forem suficientes para recompor o caminho da decisão e alternativas menos intrusivas entregarem o mesmo resultado protetivo, a abertura se limita a essas medidas, com monitoramento e correções proporcionais.

Postas as bases e explicitada a centralidade epistêmica, passa-se à formalização operativa da ponderação, na qual a ‘fórmula do peso completa refinada’ é aplicada como regra de decisão sensível às premissas empíricas e normativas do caso, basta agora entender finalmente a alocação dos algoritmos nesse cenário ponderativo.

Algoritmos são procedimentos passo a passo que transformam entradas em saídas; na inteligência artificial, especialmente no aprendizado de máquina, modelos inferem regularidades a partir de dados históricos para classificar, prever ou decidir. Como todo dado carrega escolhas e contextos, as saídas refletem essa historicidade. O problema jurídico emerge quando, sob aparência de neutralidade e eficiência, decisões automatizadas replicam desigualdades preexistentes. A partir daqui, importa menos “o que é um algoritmo” e mais “como ele decide”: com quais dados, sob qual documentação e com que capacidade de ser auditado e explicado.

Daí derivam critérios objetivos para graduar a intervenção, isto é, o impacto no direito afetado, a vulnerabilidade do atingido – podendo ser ainda mais preponderante diante de uma situação de interseccional discriminação –, a opacidade

residual após esclarecimentos voluntários – ou total diante de nenhuma resposta –, alternativas técnicas menos intrusivas, e, por fim, a posição de mercado que ocupa a desenvolvedora, dada a sua assimetria informacional – *big techs* devem ter maior responsabilidade em se desincumbir do ônus *probandi* de provar a higidez de suas aplicações. Esses critérios justificam, caso a caso, a gradação A1, A2 ou A3.

Neste ponto, a ponderação deve ser compreendida não apenas como técnica, mas como norma jurídica (postulado normativo) – status conferido pelo art. 489, § 2º, do Código de Processo Civil. Este dispositivo impõe ao julgador que justifique o objeto e os critérios da ponderação em caso de colisão de normas, explicitando as razões que autorizam a interferência e as premissas fáticas que a fundamentam. Assim, a ponderação assume a condição de postulado aplicativo que estrutura a interpretação e aplicação de regras e princípios com racionalidade e justificação interna (Alexy, 2020, p. 196). Ela atua como instrumento de equidade, buscando a solução mais justa para o caso concreto quando a *textura aberta* da lei não abarca todas as hipóteses e exigindo um raciocínio prudencial do intérprete (Nedel, 2006, p. 2019-211). Nesse quadro, a ponderação permite sopesar bens, valores e razões não apenas entre princípios (proporcionalidade em sentido estrito), mas também em relação à derrotabilidade das regras – reconhecendo que o raciocínio jurídico é não monotônico e que regras podem admitir exceções implícitas quando razões superiores (axiológicas ou teleológicas) justificam sua superação, sem invalidá-las (Prevedello, 2019, p. 188-189). Aplicadas as leis da ponderação de Alexy – a lei material (“quanto mais alto o grau de satisfação ou de prejuízo a um princípio, tanto maior a importância da satisfação do outro”) e a lei epistêmica (“quanto mais intensa a intervenção em um direito fundamental, tanto maior deve ser a certeza das premissas que a sustentam”) – o modelo decisório proposto calibra a intervenção judicial. No contexto da discriminação algorítmica, isso significa que a ponderação deve incidir sobre as premissas empíricas (dados, modelos, funcionamento) e normativas que sustentam as decisões automatizadas, garantindo que o grau de abertura da “caixa-preta” (A1-A3) seja proporcional ao risco e à vulnerabilidade e que a decisão seja solidamente justificada e controlável.

A lei epistêmica impõe que quanto maior a intervenção no segredo algorítmico, maior deve ser a certeza das premissas fáticas – documentação, *logs*, rastreabilidade – que justificam a abertura. Em contextos de alta assimetria

informacional, sobretudo em mercados dominados por grandes operadores, há deslocamento progressivo do ônus informacional para quem detém o sistema, a fim de prevenir discricionariedade epistêmica (Alexy, 2015).

Diante da realidade da discriminação algorítmica, surge a necessidade de desenvolver uma abordagem jurídica que permita efetivar a proteção e a tutela a direitos fundamentais com as urgentes demandas trazidas e ocasionadas pela inovação tecnológica, constituindo o melhor exemplo do atraso conceitualmente necessário do direito para com as realidades do mundo. Neste último ponto, serão analisados os principais direitos fundamentais em colisão, evocando-se direitos implícitos e trazendo uma correlação e um nivelamento às categorias jusfundamentais que dizem respeito aos interesses de ambos os polos das situações de discriminação algorítmica. A partir daí, munidos de categorias jusfundamentais, será proposta uma verificação racional, dentre limites e possibilidades fáticas e jurídicas, a respeito do embate ponderativo dos direitos em colisão, que deverá ser resolvido através de uma interpretação racional que se valha dos princípios e das ferramentas da ponderação material, submetido posteriormente ao escrutínio da proporcionalidade. Ver-se-á um protagonismo da explicabilidade – fogo de Prometeu que ilumina ocorrências e desalinhamentos programáticos que resultaram em agressões indevidas a direitos. Aposta-se na explicabilidade enquanto direito (para dimensionar a origem e a extensão do dano e corrigir a aplicação) e enquanto dever (cláusula fiduciária de desenho que assegura *outcomes* explanáveis e sindicáveis, na medida – em grau – necessária ao caso concreto, alcançando inclusive múltiplos *inputs* da aplicação).

A proteção aos direitos individuais é tema necessário e urgente em contexto algorítmico. A proteção deve vir sempre e em todo o caso em que identificada lesão ou ameaça de lesão a direito fundamental pessoal. Isso não significa, porém, que em qualquer caso, seja qualquer o grau de afetação à direito fundamental, se legitime a total abertura de códigos-fonte, chaves algorítmicas, programação, caminho decisório automatizado ou demonstração do funcionamento da rede neural.

Veja-se, na hipótese, uma causa imediata para cada grau de abertura, entre A1 e A3. Quando a decisão automatizada produz efeitos patrimoniais moderados e reversíveis – por exemplo, a negativa de crédito que recai sobre consumidor pertencente a um grupo vulnerável –, a causa imediata para abertura no grau A1 está no próprio dever de esclarecimento proporcional: o operador deve fornecer trilhas

mínimas e verificáveis do processamento (*logs*), documentação do modelo aplicável ao caso (finalidade, principais variáveis de entrada, parâmetros ou regras decisórias e métricas de desempenho), relatório de impacto correspondente e justificativa individualizada em linguagem inteligível ao jurisdicionado. Se, não obstante, persistirem inconsistências ou opacidade residual quanto ao encadeamento fático-técnico, a resposta não se exaure – ela escala para uma abertura de nível inicial – o que deve ser atestado em perícia técnica especializada, sempre determinada por decisão judicial justificada; inclusive sobre se permanece como A1 ou sobe para A2.

Em chave análoga, a 3ª Turma do STJ, no REsp nº 2.124.424/SP, reconheceu, como visto, a figura do chamado "inventariante digital", responsável por acessar sob sigilo os bens digitais do falecido e reportar apenas os elementos patrimoniais relevantes, vedada a exposição de dados pessoais. O mesmo raciocínio pode ser transposto ao exame de discriminações algorítmicas: o perito digital/algorítmico deve atuar como filtro técnico, acessando o sistema de forma controlada e devolvendo ao juízo apenas os achados estritamente indispensáveis ao contraditório e à atribuição de responsabilidade, sem desnudar integralmente o código ou a base de dados.

Quando a decisão automatizada atinge bens de subsistência – pense na suspensão de benefício assistencial que compromete renda alimentar ou cuidados de saúde –, a causa imediata para o grau A2 reside no risco qualificado ao mínimo existencial. Mantêm-se as exigências do nível anterior e soma-se perícia, sob sigilo e controle judicial, com acesso a amostras representativas do conjunto de dados pertinente, às variáveis com maior contribuição para o resultado, aos limiares utilizados e às regras de decisão efetivamente aplicadas no caso. Aqui se preserva o segredo algorítmico, e a abertura se dá até o ponto eventual que o perito entender identificado o motivo da discriminação. Se ainda houver incerteza material sobre o nexo causal do resultado – ou se o risco se alçar a patamar de ameaça imediata a direito de grupos vulneráveis –, a proporcionalidade conduz ao passo seguinte.

Nos domínios em que a decisão automatizada incide sobre bens máximos – vida e integridade física ou direitos fundamentais essenciais, como em evento com sistema de direção autônoma que fere pedestres –, a causa imediata para o grau A3 (cabível apenas em instância judicial) está na combinação entre gravidade do impacto, opacidade persistente após A1/A2 (possíveis em instâncias judiciais e

administrativamente) e inexistência de meio menos intrusivo para recompor o nexo causal dada a dificuldade (opacidade) do sistema. Admite-se, então, a abertura estritamente necessária das camadas relevantes do sistema – arquitetura da aplicação, critérios codificados treinados, código operacional, base de treinamento, *feedbacks* de treinamento e, por último, se e somente se indispensável trechos do código-fonte, até ultimar, na hipótese mais gravosa, no código-fonte como um todo – como uma intervenção máxima dos sistemas algorítmicos em virtude da gravidade dos resultados da decisão automatizada. Tudo se faz sob protocolo reforçado de confidencialidade, com objeto e duração delimitados objetivando na maior medida do possível o segredo. O objetivo não é desnudar a tecnologia, mas viabilizar justificação pública e atribuição de responsabilidade com base em evidência verificável, o que pode vir a colapsar o núcleo jusfundamental do segredo algorítmico. Ainda assim, a decisão demonstra-se adequadamente justificada pelo patamar de abertura A3.

Acerca especificamente de casos de discriminação algorítmica por decisão automatizadas há que se entender aplicada a mesma escala A1-A3, que orientará a abertura segundo a causa imediata verificada, ou em verificação, no caso em concreto. Para uma abertura em grau A1, bastam indícios consistentes de tratamento desigual com efeitos patrimoniais moderados e reversíveis, com identificação do impacto e mínimo de justificativa individualizada, com ônus do operador em linguagem natural acessível. Para a abertura em grau A2, exige-se situação em que o tratamento injustificadamente discriminatório atinja direitos mais sensíveis e essenciais ou posições jurídicas sensíveis de grupos vulneráveis quaisquer que sejam. Nesse grau de abertura já se justifica o acesso controlado por amostra de dados, com salvaguardas de confidencialidade exposição detalhada de perito e desenvolvedora acerca de justificativas individualizadas acerca da origem dos desvios cometidos pela aplicação. No grau final de abertura, A3, ficarão as hipóteses de práticas discriminatórias graves, que recaem sobre direitos fundamentais estruturantes e essenciais, ou quando persistir a opacidade material após a abertura em graus A1 e/ou A2. Nesse último ponto, justifica-se a utilização de engenharia reversa e/ou a abertura tanto quanto for sendo estritamente necessária das camadas relevantes do sistema, sob protocolo que tente manejar a abertura com o segredo algorítmico, evitando ao máximo a exposição total da engenharia codificada, mas a permitindo em caso de necessidade. Não se trata de expor à público toda a tecnologia, mas de

propiciar condições de justificação pública e de verificação acerca da atribuição de responsabilidades sobre as injustificadas discriminações algorítmicas.

Verificando em exemplos: um sistema automatizado de avaliação de risco apresenta impacto racial desproporcional e, mesmo após explicações iniciais, permanece com pontos sem clareza. Inicia-se já no nível intermediário de abertura (A2), dado o impacto relevante e os indícios de viés. O primeiro passo é verificar se o modelo funciona de modo adequado para os grupos atingidos. Em seguida, examina-se a necessidade: é preciso estudar alternativas que possam operar como substitutas de atributos sensíveis (por exemplo, bairro de residência ou escolaridade) e explorar conjuntos de alternativas menos intrusivas. Na etapa final, a decisão deve equilibrar impacto, pesos e grau de certeza, o que exige registros de execução, perícia técnica independente e plano de correção. Aqui, o grau de certeza é moderado: a documentação é parcial e não abarca todas as informações necessárias. Dois fatores impõem elevar o nível: a persistência do impacto desigual e a resistência do operador em fornecer registros, ou a inexistência deles. Nessa hipótese, passa-se ao nível aprofundado (A3), sob segredo de justiça, com entrega de documentação completa, acesso controlado aos registros e reexecução supervisionada do sistema com perito algorítmico, buscando preservar o código-fonte quando não for estritamente necessário expô-lo. Por último podemos voltar ao caso da Cambridge Analytica e verificar, dada a lesão projetável de danos à autodeterminação e às esferas públicas democráticas, diante do alcance coletivo sensível, a abertura em nível mínimo A2 ou A3 estaria justificada, pelo potencial dano resultante.

Quando há indícios de impacto díspar por raça, gênero ou território, a intervenção tende a iniciar em A2. Persistindo opacidade residual relevante – sobretudo diante de bairros segregados ou correlações espúrias com grupos vulneráveis –, a abertura escala a A3 para permitir a verificação sob sigilo de critérios de decisão. A reprodução de padrões históricos discriminatórios recomenda A2 desde o início, dada a assimetria entre candidatos e plataformas. Se a filtragem impacta direitos de subsistência de forma massiva ou incide sobre interseccionalidades severas, a intensidade do risco e o déficit de explicabilidade autorizam A3, com perícia assistida focada nos parâmetros efetivamente determinantes dos resultados, preservado o segredo sob regime de confidencialidade judicial.

Em relação ao grau de certeza epistêmica (Ce) Quando o impacto e o peso do direito afetado são altos e as premissas estão bem demonstradas (certeza segura), a abertura tende a iniciar já em A3, com perícia técnica nos moldes delineados. Com impacto médio e certeza plausível, A2 é a medida adequada, com testes e plano de mitigação. Em impactos baixos e certeza frágil, inicia-se por A1, advertindo o operador de que a persistência de opacidade eleva o nível e desloca o ônus informacional para quem detém o controle do sistema. Em todos os cenários, a abertura cresce à medida que aumentam o risco, a vulnerabilidade do grupo atingido e a falta de explicabilidade; o que o juiz determina, de ofício ou a requerimento das partes a qualquer momento no processo de cognoscibilidade.

Em benefícios sociais e serviços públicos essenciais (saúde, educação, assistência), a natureza do bem jurídico e a posição de vulnerabilidade dos titulares inclinam o padrão para A3. A abertura profunda – código e parâmetros relevantes ao resultado impugnado, acompanhados de trilhas de auditoria – é proporcional porque sem ela não há como conferir confiabilidade às premissas que restringem direitos fundamentais de grupos já expostos a desvantagens estruturais. O sigilo judicial e a compartimentalização do acesso operam como salvaguardas adequadas ao segredo legítimo remanescente. Que será relativizado com critérios adequados e previsíveis, protegendo o que merece proteção, e levantando a eficácia jusfundamental de modo justificado, à benefício da não-discriminação.

O ser humano tende a confiar mais em outros humanos do que em máquinas. Naturalmente. Tal paradigma talvez esteja a ponto de ser substituído, com uma maior confiança depositada nas máquinas, dada a (aparente) qualidade decisório. Há assim, como visto, que passar por uma necessária opção acerca de um ponto de vista de mundo, se será um mundo de pacífica convivência com a automação de máquinas, haja vista uma menor exigência de sua efetividade, já que “pilotadas” e programadas anteriormente por seres humanos, ou se exigiremos mais segurança epistêmica, mais demonstrativos fáticos para a tomada de decisão. É precisamente para isso que serve a gradação A1–A3: reconduzir a confiança a critérios verificáveis.

Aliado ao entendimento acerca dos contornos e possibilidades do mundo, há também que se realizar a opção acerca de um objetivo claro, como o de exigir explicabilidade na medida das possibilidades técnicas. A busca por uma XAI/xAI (*Explainable Artificial Intelligence*) promove um equilíbrio entre a total desaparição de

interesses privados no desenvolvimento de Inteligências Artificiais, por conta do segredo comercial, e de uma total clausura das caixas-pretas (Bonat; Peixoto, 2020, p. 54-56). E há maneiras distintas de se exigir tais explicações, visto que muitas são as formas de XAI. Se pode ter capacidade de **i)** explicação para apenas justificar a decisão – a partir de resultados específicos ou da lógica do raciocínio interno do funcionamento, ou **ii)** para conhecer os caminhos tomados pela aplicação e, com isso, promover aprendizado de máquina para outras aplicações ou para que seja de conhecimento geral, ou **iii)** ainda para incrementar a lógica da aplicação ou deixa-la mais segura, ou ainda, e por fim, para **iv)** controlar de maneira mais eficientes os resultados, tornando mais aprimorado o controle das vulnerabilidades (Adadi; Berrada, 2018). A medida da adequação dessas possibilidades deverá ser levada em conta na tomada de decisão na colisão de direitos fundamentais.

À vista do exposto, a matriz A1–A3 não opera apenas como marco para a engenharia de dados, mas como critério jurídico de densidade do ônus argumentativo e de abertura proporcional do algoritmo. Quanto mais alto o risco materializado sobre bens de subsistência e sobre sujeitos atravessados por interseccionalidades severas, maior deve ser a exigência de segurança epistêmica e de reconstrução pública do itinerário lógico do modelo, sob pena de inviabilizar a própria lei material da ponderação. É precisamente nesse ponto que a explicabilidade deixa de ser desiderato tecnológico e se converte em conteúdo jurídico exigível – implícito à autodeterminação informativa e à igualdade material – com eficácia horizontal qualificada entre titulares e plataformas e com reforço do dever estatal de proteção. Em termos operacionais, a gradação A1–A3 reconduz a confiança a critérios verificáveis, preservando o segredo comercial legítimo remanescente, mas nunca à custa da verificação de premissas que comprimem direitos fundamentais. É nesse giro – da calibragem técnico-jurídica à positividade normativa – que se insere o próximo passo: o reconhecimento, já em curso, de deveres legais de transparência e de revisão em decisões automatizadas em experiências comparadas e no direito brasileiro, como se examina a seguir.

A proteção contra a discriminação automatizada permite inferir, pelo contexto jurídico atual, o reconhecimento de um direito implícito à explicabilidade das decisões automatizadas. Esse direito decorre da autodeterminação individual na proteção de dados pessoais e privacidade. A aplicação aqui não é da eficácia vertical, contra o

Estado, mas da eficácia horizontal, entre usuários e empresas de tecnologias, ou então do reconhecimento de uma nova eficácia vertical contra agentes privados (Teubner, 2020, p. 273-274) que detém o domínio sobre a (auto)regulação e o uso dos instrumentos de tecnologia pela população, eis que dominam grande fatia mercadológica das tecnologias de rede (Celeste, 2021, p. 66), funcionando como verdadeiros e novos núcleos paraestatais de poder institucional do desterritorializado ambiente virtual (De Gregorio, 2021, p. 57-58).

A explicabilidade em Inteligência Artificial já é prevista alguns países. Na França, decretos obrigam que cidadãos sejam informados sempre que decisões individuais forem tomadas por algoritmos, devendo a administração esclarecer como o algoritmo contribuiu para a decisão, incluindo dados utilizados e metodologias. Na Itália, exige-se pleno conhecimento do funcionamento e critérios dos algoritmos usados em decisões administrativas. No Brasil, a LGPD prevê revisão de decisões automatizadas, mas não torna obrigatória a explicabilidade.

A transparência algorítmica busca coibir práticas discriminatórias e aperfeiçoar os sistemas, alinhando-os ao Estado Democrático de Direito. Entretanto, desafios persistem devido à complexidade dos modelos (*black box*), reforçando a necessidade de explicabilidade e divulgação de dados de treinamento. O equilíbrio entre ética, direito e proteção ao segredo comercial é fundamental, sendo necessárias justificativas sólidas para eventuais restrições ao acesso.

O método para solucionar qual direito fundamental deverá sobressair fortalecido (argumentativo-racionalmente) da colisão é o chamado teste da proporcionalidade, contudo ele não pode vir desconectado da noção da proibição do excesso (*Übermassverbot*) e da vedação da proteção insuficiente (*Untermassverbot*), do âmbito de proteção, ou núcleo duro[14], de um direito fundamental e dos limites aos limites (*Schranken-Schranken*) dos direitos fundamentais, tampouco pode se olvidar do dever de proteção estatal (*Schutzpflicht*) a ser otimizado quando da interpretação e aplicação do direito e das normas jusfundamentais (Leal; Maas, 2020, p. 50).

Para a efetivação dos deveres estatais de proteção aos direitos fundamentais, o Estado não pode violar (em excesso) o âmbito de proteção do direito fundamental precedido, tampouco pode proteger insuficientemente, isto é, ficando aquém dos

[14] Segundo a teoria externa dos limites dos direitos fundamentais, existe inicialmente um direito em si, ilimitado, *prima facie*, que, mediante a imposição de restrições, vai limitando o seu campo de atuação, numa aplicação definitiva do âmbito de proteção (Sarlet, 2013, p. 199).

níveis mínimos de proteção constitucionalmente exigidos. É nessas perspectivas que se pode falar que o teste da proporcionalidade possui uma dupla face, atuando simultaneamente como controle de constitucionalidade das medidas que restringem direitos fundamentais, bem como para o controle das omissões (ou proteção insuficiente) do Estado no cumprimento de um dever de proteção (Sarlet, 2013, p. 202-205).

Sarlet (2013, p. 202) explica que uma constituição não se restringe a regulamentar formalmente uma série de competências, mas estabelece, conjuntamente, uma ordem substancial de princípios. Desse modo, o controle de constitucionalidade material, concernente na observância da proteção do núcleo essencial de determinado direito fundamental, visa atender aos requisitos da proporcionalidade (e da razoabilidade), bem como respeitar a proibição do retrocesso. Esses seriam os limites aos limites dos direitos fundamentais, isto é, instrumentos que funcionam como barreiras à restringibilidade de determinado direito, garantindo sua eficácia.

Dessa forma, o teste da proporcionalidade, funcionando enquanto instrumento lógico-argumentativo para o controle da legitimidade da solução do conflito interpretativo em caso de colisões de direitos fundamentais constitucionalmente estabelecidos, deve ser analisado em três partes ou subtestes[15-16], quais sejam, a adequação (ou idoneidade técnica) (*Geeignetheit*), a necessidade (*Erforderlichkeit)* e a proporcionalidade em sentido estrito (ou ponderação) (*Verhältnismäßigkeit im engeren Sinne)*. Por meio destas etapas de verificação da validade jurídica e da legitimidade constitucional do meio adotado para a resolução de conflitos entre princípios, poderá ser avaliado, escalonadamente, se o meio empregado é adequado, necessário e em qual medida de restrição e satisfação ele deverá ser otimizado, sempre conforme as possibilidades fáticas e jurídicas, ainda que para tanto seja necessário o diálogo multidisciplinar pelos julgadores (Gavião Filho, 2011, p. 239).

Nesse ponto tem-se a primeira lei da ponderação (material), no sentido de que quanto maior for o grau da intensidade da intervenção, ou restrição, de um direito

[15] Fala-se também em uma quarta etapa do teste da proporcionalidade, que viria antes do subteste da adequação, da análise da legitimidade (jurídica) do meio empregado. Ela examina se o ato persegue um fim legitimamente jurídico (Klatt; Meister, 2012, p. 8).

[16] Além disso, como estes subtestes são satisfeitos ou não, e não podem ser ponderados frente a outros testes, são considerados como regras, de modo que sua não satisfação caracterizará a ilegalidade e a inconstitucionalidade do resultado da colisão de princípios (Gavião Filho, 2011, p. 239).

fundamental, tão maior deverá ser o grau de importância, de satisfação, ou, de realização, do outro (Gavião Filho, 2011, p. 249). Em termos operativos, isso significa que quanto mais se exigir abertura do segredo comercial-algorítmico, maior deve ser a demonstração – empírica e normativa – de que a igualdade material, a não discriminação e a autodeterminação informativa reclamam tutela prevalente naquele caso concreto[17].

Ao final da ponderação verifica-se a relação de concretização de um princípio e o prejuízo do preterido. A ponderação pode ser expressa em uma lei material: "quanto mais alto é o grau de não cumprimento ou prejuízo de um princípio, tanto maior deve ser a importância do cumprimento do outro" (Gavião Filho, 2011, p. 249).

De modo algum o escrutínio da proporcionalidade poderá ser operado em partes, constituído aqui um verdadeiro caso de eficácia jurídico com natureza de regra – com aplicabilidade na modalidade "tudo ou nada" – visto que cada etapa e postulado do escrutínio desenvolvem papel ímpar e substancial para coibir uma irrefreada discricionariedade arbitrária; típica do uso isolado da ponderação, caracterizada pela presença de elementos retóricos complementares *ethos* e *pathos*, com pouco ou nenhum apego à dimensão do *logos* (preferencialmente acompanhado complementarmente apenas do *pathos*) (Von Schlieffen, 2022, p. 185).

Em síntese, a ponderação material, aplicada a casos de discriminação algorítmica, não admite atalhos: exige a explicitação dos bens em jogo, a qualificação do risco, a verificação da confiabilidade das premissas e a demonstração – passo a passo – de porque a medida escolhida supera alternativas menos restritivas. A regra é simples, ainda que exigente: quanto maior o impacto e a vulnerabilidade – e quanto mais opaco o processo decisório –, maior a abertura informacional e a densidade da motivação.

A proporcionalidade, surgida em 1794 no artigo 10(2) do Código Civil Prussiano (*Allgemeines Landrecht für die Preußishen Staaten)*, em que a lei autorizava o governo a exercer o poder de polícia para assegurar a paz pública, a segurança e a ordem, tomando as medidas necessárias para tanto, isto é, as medidas

[17] A matriz A1–A3 dá corpo a esse mandamento: em A1, a explicabilidade documental satisfaz a exigência mínima de controle; em A2, a perícia sob sigilo e o acesso controlado elevam a justificabilidade da intervenção; em A3, a abertura profunda estritamente necessária viabiliza a recomposição do itinerário lógico-decisório quando o impacto é grave/irreversível – sempre com salvaguardas de confidencialidade, compartimentalização e objeto delimitado. Assim, a abertura não é curiosidade técnica – é correlação proporcional entre grau de dano e vulnerabilidade de um lado e ônus de verificação pública das premissas de outro.

minimamente essenciais para tanto (Grimm, 2016, p. 171). O Supremo Tribunal Administrativo da Prússia (*Preußishes Oberverwaltungsgericht*) (1882-1914), usou intensamente a proporcionalidade para examinar a legitimidade da intervenção governamental na economia e na proteção à vida social, também em questões atinentes ao direito de propriedade, quanto ao direito de liberdade de reunião e de expressão. A proteção dos direitos individuais, portanto, se baseava na legalidade dos atos administrativos, bem como em sua proporcionalidade. Assim, o executivo se consagrou como maior infrator de direitos individuais, o judiciário se concretizou como garantidor deles e o legislativo como órgão representativamente legítimo para a escolha de valores da comunidade política (Barak, 2012, p. 177-178).

O direito constitucional sempre foi substantivo, centrado não apenas na interpretação, mas na doutrina da proporcionalidade que gradualmente se desenvolveu para ser o centro da jurisprudência constitucional alemã, herdeira da tradição do desenvolvimento dos direitos pela via administrativa alemã e prussiana (Cohen-Eliya; Porat, 2017, p. 116-120). Característica marcante do Tribunal Constitucional Federal alemão é o conceito mais abrangente dos direitos, garantindo e reconhecendo o direito, p.e., de se andar de cavalo em parques e de alimentar os pombos[18]. Essa concepção aberta dos direitos é herdeira da origem administrativa dos direitos se baseando em uma ideia geral de direitos de liberdade, pelo Estado de direito e pela proporcionalidade, que impõem aos governos uma obrigação negativa de respeito e deferência aos direitos e respeito aos cidadãos (Cohen-Eliya; Porat, 2017, p. 117-119). Möller (2012, p. 208) faz uma reconstrução moral do modelo alemão de direitos constitucionais e, ao final, conclui que não há limitações aos direitos nesse modelo, podendo se adjudicar constitucionalmente de todas as questões, e tendo sempre alguma margem de reconhecimento de novos direitos específicos. Ou, ainda, novas relações de eficácia destes direitos fundamentais.

No constitucionalismo digital, a assimetria informacional e o poder estrutural de plataformas ensejam reconhecer uma eficácia jusfundamental qualificada nas relações com grandes operadores: desloca-se o ônus argumentativo/informacional para quem detém o sistema e o mercado, justamente porque a opacidade não permite a satisfação do mínimo constitucional de igualdade e devido processo. Esse

[18] A Corte Europeia dos Direitos Humanos, seguindo essa tendência, reconheceu o direito de ser dormir uma noite tranquila e quieta, no caso do *Heathrow airport* (Hatton v. United Kingdom, (2003) (37 EHRR 28), e, também o direito a se viver em uma sociedade livre (Cohen-Eliya; Porat, 2017, p. 118).

reconhecimento, porém, dignifica ambos os direitos: protege-se o núcleo essencial do segredo comercial-algorítmico – vantagem competitiva e inovação – na exata medida em que não inviabilize a verificação das premissas que restringem direitos fundamentais. O equilíbrio se extrai da aplicação disciplinada da primeira lei da ponderação (lei material da ponderação – *materielles Gesetz der Abwägung*): se o impacto sobre bens de subsistência e grupos em interseccionalidade é alto e a opacidade residual persiste, então cresce a importância de satisfação do direito antidiscriminatório e se legitima a migração a A2/A3 sob sigilo; se o impacto é moderado e reversível e há explicações adequadas, então A1 basta. Desse modo, a tese enfrenta o dilema pela graduação proporcional do ângulo de abertura – A1–A3 – , reconduzindo a confiança social a critérios verificáveis e convertendo a ponderação material em regra de decisão apta a compatibilizar proteção de vulneráveis e integridade concorrencial, sem absolutos e sem abdicar do núcleo essencial de nenhum dos lados.

À luz desse quadro, a lei material da ponderação deixa de ser uma cláusula retórica para operar como regra de decisão em constitucionalismo digital: direitos não se abatem por absolutos, mas por correlação proporcional entre intensidade do impacto e densidade do ônus de verificação. A eficácia jusfundamental qualificada nas relações com grandes plataformas desloca, em razão da assimetria informacional, o encargo de documentação e demonstrabilidade. Dignifica-se, assim, ambos os polos, e o faz na exata medida em que não inviabilize a recomposição pública do itinerário lógico-decisório quando há risco elevado a bens de subsistência e a grupos sob interseccionalidades severas. Claro, se a resposta for, *in casu*, justificável a ponto de se ver uma total e abrupta abertura de todo o código-algorítmico, o julgador poderá fazê-lo, desde que fundamente a decisão com base nesse núcleo de valores comuns que são compartilhados no presente esforço.

A partir dessa âncora, o argumento desloca-se do porquê para o como decidir: a ponderação material não exime método – exige-o. É nesse ponto que o escrutínio em três tempos da proporcionalidade – adequação (*Geeignetheit*), necessidade (*Erforderlichkeit*) e proporcionalidade em sentido estrito (*Verhältnismäßigkeit im engeren Sinne*) – passa a calibrar, caso a caso, o ângulo proporcional de abertura (A1–A3) e a distribuição dinâmica do ônus informacional, segundo o binômio impacto/vulnerabilidade e a opacidade residual. O passo seguinte – e apenas assim

– é explicitar a arquitetura argumentativa que legitima a decisão: delimitar o objeto do conflito, qualificar as premissas fático-técnicas, explicitar alternativas menos gravosas e justificar a precedência obtida, sob a dupla exigência de justificação interna e justificação externa do *decisum*.

Assim, quando enuncia-se a fundamentação da preocupação acerca da discriminação algorítmica na lei material da ponderação não se está autorizando a verificar a racionalidade das decisões apenas e unicamente a partir desta, mas descrevendo que necessariamente é a lei material da ponderação a legitimar que se verifique a racionalidade da justificação das decisões que amplifiquem a eficácia material, ou intensidade, dos direitos fundamentais tutelares dos princípios da transparência e igualdade – exemplos dentre outros possíveis – tanto quanto esses sofreram antes da intervenção e tanto quanto o direito fundamental ao segredo (comercial, algorítmico ou industrial) das desenvolvedoras deverá suportar a partir da medida judicial. Aqui considerando tanto a dimensão tópica, no modelo de Viehweg – e decorrentes da escola de Mainz –, como na universazibilidade de MacCormick – e decorrentes de sua cátedra em Edimburgo (Paulo, 2022, p. 136-137).

Para tanto, é necessário examinar as etapas do escrutínio. Nesse sentido, os subtestes da adequação e da necessidade tratam das possibilidades fáticas, expressando o mandamento a ser otimizado, visando excluir restrições evitáveis aos direitos fundamentais porventura precedidos. A adequação estabelece uma verificação da possibilidade da medida interventiva em um direito fundamental, promotora da realização de outro direito fundamental, a ser tomada é idônea ou não, isto é, se é apta a realizar um fim constitucionalmente legítimo (não proibido). No âmbito da proporcionalidade na atividade legislativa, sempre que várias medidas forem consideradas aptas, será do espaço estrutural de decisão do legislador optar por uma delas, levando-se em conta a ideia de Pareto-ótimo[19].

Para tanto, necessário verificar e consolidar o entendimento acerca do fenômeno da opacidade algorítmica. Só assim se poderá verificar a correção de medida que a enfrente. A opacidade algorítmica — a “caixa-preta” — tem três fontes principais: **i)** complexidade técnica dos modelos; **ii)** segredo algorítmico; **iii)** dados confidenciais e volumosos. O antídoto jurídico exige, pelo menos, a rastreabilidade

[19] O princípio da adequação é uma expressão da ideia de Pareto-ótimo e exclui a adoção de meios que possam obstruir pelo menos um direito sem promover nenhum outro direito ou interesse (Klatt; Meister, 2012, p. 9).

(*logs* de dados/modelos e documentação de *datasets*) *by design*, a auditabilidade independente proporcional ao risco (o que se está empreendendo esforço para vislumbrar, a explicabilidade *by design* adequada ao público afetado e, em uma dimensão ligeiramente mais ética, a supervisão humana significativa nas decisões com impacto em direitos. Quanto maior o impacto, maior o grau exigível desses fatores de controle.

Já a necessidade diz que quando duas, ou mais, medidas forem consideradas igualmente idôneas deve-se optar pela medida que intervenha menos intensamente no outro direito fundamental. Aqui, em uma esfera de atuação legislativa, a escolha deverá ser fundamentadamente sustentada pelos conhecimentos gerais e científicos existentes no momento da intervenção, que deverá ser, na hipótese de ainda se sustentarem duas ou mais medidas, a mais eficaz, rápida, provável e com menor afetação de direitos fundamentais (Gavião Filho, 2011, p. 241-248).

Nesse ponto, a Avaliação de Impacto Algorítmico (AIA) surge como instrumento imprescindível: ao documentar preventivamente os riscos, as alternativas disponíveis e os efeitos diferenciados sobre grupos vulneráveis, fornece ao intérprete parâmetros objetivos para aplicar o subteste da necessidade e para fundamentar, de forma epistêmica, a escolha pela medida menos gravosa. A ausência de AIA, quando exigida, ou sua realização meramente formal, acarreta o agravamento do ônus argumentativo da empresa ou órgão público que opera o sistema, legitimando, em contrapartida, maior grau de abertura proporcional e de escrutínio judicial.

O subteste da proporcionalidade em sentido estrito expressa um mandamento de otimização frente às possibilidades jurídicas. Esse teste é consubstanciado pela noção de ponderação[20], a qual estabelecerá uma relação de precedência

[20] Para um maior aproveitamento da ponderação Alexy (2015) formulou, e Klatt e Meister (2012, p. 10-11) ampliaram, a fórmula peso, um modelo matemático em que analisam que o peso absoluto, ou a correta interferência do princípio alvo de preferência, terá precedência sobre o outro colidente, é fruto de uma contraposição entre o peso relativo de cada princípio (i e j), os seus graus de interferência ao caso, e a confiabilidade empírica e normativa a cada um deles no caso, conforme segue, em que W é o peso abstrato dos princípios, I é a intensidade de interferência de um sobre o outro, R^e é a confiabilidade da premissa empírica e R^n da premissa normativa:

$$W_{i;j} = \frac{W_i.I_i.R^e_i.R^n_i}{W_j.I_j.R^e_j.R^n_j}$$

Resultado desse teste será analisado na escala triádica, na qual os resultados da ponderação poderão ser i) leve, ii) moderado ou iii) sério, ou ainda, no modelo triádico duplo, i) leve-leve, ii) leve-médio, iii) leve-sério, iv) médio-leve, v) médio-médio, vi) médio-sério, vii) sério-leve, viii) sério-médio, ix) sério-sério. Assim, pode-se descrever a escala triádica básica em uma escala exponencial, em que as respostas poderão ser 2^0 (igual a 1), 2^1 (igual a 2) e 2^2 (igual a 4), de modo que quando Wi;j for maior

condicionada entre os direitos fundamentais ou os bens jurídicos coletivos constitucionalmente protegidos, com base na justificação racional entre o grau de intensidade da intervenção no direito precedido pelo grau de importância de realização do direito que ganhou precedência. Essas decisões no âmbito da atividade legislativa não contam com maiores questionamentos, contudo, no âmbito judicial, emerge o problema da legitimidade judicial na decisão sobre os graus de intensidade de intervenção e de importância de realização desses direitos fundamentais, porquanto muito facilmente se rompem os princípios formais da competência decisória do legislador democraticamente eleito[21] (Gavião Filho, 2011, p. 249-251).

Assim, é com base na ponderação material, acerca dos efetivos núcleos de significações dos direitos fundamentais em colisão, especificamente a transparência algorítmica e o segredo algorítmico, que deverá se vislumbrar, **i)** qual está com seu núcleo essencial mais desprivilegiado em certa situação fática, **ii)** a correlação (e o veículo dessa) entre a intensidade da intervenção (na maioria dos casos) do segredo algorítmico (do fato de o algoritmo não estar sendo transparente por proteção à propriedade intelectual) na identificação das motivações pelas quais ocorreu discriminação automatizada (identificação de maior risco injustificado, ou justificado apenas por dados que denotem ser o usuário de grupos vulneráveis ou estar geograficamente em área menos nobre), **iii)** a necessidade do estabelecimento gradativo da intensidade da transparência sobre o segredo, de modo a compatibilizar a melhor relação entre intervenção e satisfação, o que demandará experimentação por níveis, autorizando-se judicialmente e gradativamente a abertura de cada vez mais

---

do que 1, então o princípio que se pretendeu como preferido tem precedência ao princípio que se objetivou preterir, mas quando for igual a 1, não há resposta segura através da fórmula, e a questão ficará para a discricionariedade judicial quando do resultado da ponderação.

[21] Streck (2013, p. 20-24 e 68-69) traça críticas ao modelo proposto por Alexy, uma vez que a discricionariedade judicial, fruto do maior protagonismo reconhecido ao judiciário para a concretização dos direitos fundamentais, leva, indubitavelmente, a um déficit democrático de legitimidade da atuação judicial e, como a teoria da argumentação depende desta discricionariedade, tece crítica a ela de igual forma. Ao que Trevisan (2017, p. 237-251) responde que Alexy, seguindo a linha kantiana sobre os direitos do homem, defende que as regras do discurso devem ter validez universal, isto é, que toda a argumentação tenha pretensão de veracidade e de correção (fundamentabilidade), que todos os participantes são sujeitos iguais em relação aos direitos, e que o direito deve postular a universazibilidade. De modo que três são os argumentos necessário para a fundamentação teórico-discursiva dos direitos dos homens, o da autonomia, o do consenso e o da democracia; em destaque o último dirá que a democracia somente é possível nos moldes de um Estado Constitucional Democrático de Direito no qual os direitos do homem estão, na forma de direitos fundamentais, indissociavelmente unidos à própria concepção de democracia. De modo que a jurisdição constitucional, cujos atores legítimos são os juízes, possa cumprir a ordenação-quadro (constituição) tanto quanto na atividade legislativa.

etapas do segredo acerca da programação, da base de dados, do treinamento, até se chegar ao código-fonte, ou então o conhecimento técnico específico a indicar ao magistrado onde possivelmente houve o enviesamento.

A todos os casos, deverá ser sopesado, desde logo porque implícito no texto constitucional, ou a partir da intervenção legislativa, a necessidade de sindicabilidade das decisões automatizadas tomadas pelas Inteligências Artificiais. Caso os modelos de XAI já estejam adotados ou exigidos pela legislação ou pela jurisprudência, deve-se avaliar, igualmente, a extensão do conhecimento acerca das possibilidades discriminatórias para permitir a apuração de eventuais perdas e danos, responsabilizações e penalizações. De modo que a busca pela explicabilidade deve ser, em primeira instância adotada pelos desenvolvedores e, após, exigida pela regulação e pela jurisprudência e, posteriormente, seja tanto exigida quanto utilizada como parâmetro na identificação das respostas adequadas a casos por juízes e tribunais. O que se pode e deve ainda realizar é o que a explicabilidade significa em termos jurídicos, em grau e extensão para com interpretação e transparência (Colaner, 2021); isso importa, contudo, no presente momento, menos do que sua efetiva aplicação e exigência.

A ponderação material é preocupação chave na busca pela gradativa explicabilidade a partir da necessidade de quais elementos e em qual nível da aplicação se está a fundamentalidade da compreensão acerca dos motivos pelos quais a discriminação algorítmica ocorreu automaticamente. Fica caracterizada, portanto, quais os pesos-padrão (explicabilidade e segredo) a serem contrastados, colididos e reequilibrados, sendo que necessitam conservar seu núcleo fundamental irredutível, sendo relativizados em sua proteção a partir da necessidade de satisfação do outro direito fundamental reclamado, a que se deverá identificar o grau de exigibilidade, importância e relevância sócio-protetivo-jurídica em casos. A preocupação jurídica para com o paradigma da fruição e das limitações dos direitos fundamentais indica a ponderação como um conceito guia não isolado na construção de um modelo decisório e interpretativo juridicamente e constitucionalmente apto a lidar com a complexa questão da discriminação algorítmica, conservando direitos e interesses das partes envolvidas, e equacionando criteriosamente os termos e a intensidade das afetações a direitos fundamentais.

Em termos normativos, a gradação A1 garante o direito de saber o porquê houve discriminação, a gradação A2 garante o direito de testar e verificar nova ocorrência da discriminação, já a gradação A3 garante o direito de verificar fortemente os motivos, ainda que isso signifique uma abertura possivelmente pública de componentes protegidos por segredos empresariais. Todos os níveis são aptos a corrigir a aplicação para a não repetição da ocorrência da discriminação. A transparência não é fim em si – é a condição de possibilidade do contraditório efetivo e da responsabilização em ambiente automatizado. Com isso, a confiança social não se apoia em promessas tecnocráticas, mas em critérios verificáveis, aptos a sustentar o juízo acerca da repercussão a partir de episódios discriminatórios. Essa é a orientação que juízes e tribunais devem seguir, salvaguardando a máxima dos direitos fundamentais, tanto quanto possível.

Em síntese, a ponderação material não se apresenta como retórica de equilíbrio, mas como verdadeiro postulado aplicativo de decisão em tempos de constitucionalismo digital. O modelo A1–A3, ao correlacionar intensidade do impacto, vulnerabilidade e opacidade residual, positiva um dever de abertura proporcional, convertendo em critério jurídico verificável o que antes se dissipava na retórica da neutralidade técnica.

Nessa chave, a confiança social não repousa em promessas de eficiência, mas em razões públicas reconstruíveis: quanto maior o risco à igualdade material e à autodeterminação informativa, tanto mais denso o ônus argumentativo da parte que detém o sistema, e tanto mais ampla a exigência de abertura. É aqui que a lei material da ponderação cumpre sua função: não autoriza violar o núcleo essencial de nenhum dos polos, mas impõe gradações que preservem, a um só tempo, a integridade da inovação e a tutela dos vulneráveis.

O resultado é um paradigma interpretativo-operativo: i) direitos fundamentais não se resolvem em absolutos, mas em graus de efetividade calibrados pela proporcionalidade; ii) a opacidade algorítmica é fator jurídico de risco que desloca dinamicamente o ônus informacional; iii) a explicabilidade deixa de ser desiderato tecnológico e se converte em conteúdo jurídico exigível, inscrito na cultura da justificação constitucional. O grau de abertura das informações e a exigência de justificativa pública também variam conforme quem opera o sistema. Quando o Estado usa modelos que afetam liberdades ou acesso a direitos, a abertura sobe um degrau

– a transparência é a regra e o segredo é a exceção, sempre limitado ao estritamente necessário. Em empresas privadas de médio porte, preservam-se os direitos das pessoas atingidas sem inviabilizar o negócio: o juiz pode modular a forma de mostrar as informações, mas não pode reduzir o nível de proteção devido. Em grandes plataformas, dado o alcance e o poder de mercado, a exigência de documentação, registros e verificação independente é mais intensa – a abertura cresce e o controle é mais rigoroso, com proteção apenas do que não for indispensável para verificar a justiça da decisão.

Ao cabo, a tese sustenta que a ponderação material, aplicada a colisões entre transparência e segredo algorítmico, não apenas equaciona o conflito, mas inaugura um *iter* decisório racional para casos de discriminação automatizada: a abertura A1–A3 torna-se a régua do contraditório efetivo, a salvaguarda de grupos em múltiplas vulnerabilidades e a condição de legitimidade democrática da decisão. É esse giro da ponderação à praxis da explicabilidade que se leva adiante, iniciando com o estabelecimento de revisões de decisões automatizadas (art. 20 da LGPD), passando ao exame de solicitações em caráter administrativo das decisões automatizadas (A1 a A2), ao necessário apreço jurisdicional da questão de maiores violações substanciais, ou de direitos fundamentais com especial destaque protetivo, como o de grupos vulneráveis – sobretudo diante de *big techs* –, com o escalonamento judicante (A2 a A3), que demanda e justifica uma intensificação no escrutínio para a abertura proporcional baseada em certezas epistêmicas controladas e asseveradas – em juízo – por perito digital/algorítmico que reportará os achadas (*stricto sensu*) ao juízo, e não um *report* completo, ao menos inicialmente, tendo em vista o segredo judicial que é repercutido e, *in casu*, deferido pelo magistrado, à luz do direito fundamental ao segredo comercial-intelectual-algorítmico.

Essa lógica encontra paralelo na jurisprudência do Superior Tribunal de Justiça, que no REsp nº 2.124.424/SP, como se viu, admitiu a figura do chamado *inventariante digital*: pessoa de confiança, designada judicialmente, que acessa bens digitais do falecido sob sigilo e reporta ao juízo apenas os elementos patrimoniais relevantes, preservando aqueles de cunho pessoal. Analogamente, no contexto algorítmico, o perito digital/algorítmico deve atuar como filtro técnico: examina os sistemas sob protocolo de confidencialidade e devolve ao processo apenas os achados estritamente necessários à solução do caso, sem desnudar integralmente a

aplicação. Dessa forma, compatibiliza-se a sindicabilidade constitucional com a proteção residual do segredo comercial, equilibrando inovação e tutela de direitos.

Em um escalonado prático, progressivo e objetivo: **i)** inicia-se com o direito de revisão de decisões automatizadas (art. 20 da LGPD); **ii)** avança ao exame de solicitações em sede administrativa (A1 a A2), com entrega documental e registros verificáveis; **iii)** culmina no necessário controle jurisdicional dos casos de maior gravidade, quando violações substanciais a direitos fundamentais – sobretudo em face de *big techs* e de grupos vulneráveis – reclamam a intensificação do escrutínio (A2 a A3). Nesse último nível, a abertura proporcional deve ser justificada por certezas epistêmicas controladas, asseguradas mediante perícia digital/algorítmica. O perito não entrega um "desnudamento integral" do sistema, mas achados estritamente relevantes ao caso concreto, reportados diretamente ao juízo sob regime de segredo judicial. Assim, compatibiliza-se o direito fundamental ao contraditório e à não discriminação com a tutela residual do segredo comercial e da propriedade intelectual.

O resultado é um modelo graduado de sindicabilidade substancial e constitucionalmente direcionada, que distribui o ônus informacional conforme o risco e a vulnerabilidade, preserva o núcleo essencial do segredo e, ao mesmo tempo, garante que a explicabilidade demonstre conteúdo jurídico exigível e verificável para transformar práticas lesivas, remendar desigualdades estruturais, responsabilizar operadores e reparar vítimas. Quando houver incidência sobre grupos em situação de vulnerabilidade, o modelo aciona gatilhos de tutela reforçada: presunção prática de assimetria informacional, elevação do grau de certeza exigível (C), métricas estratificadas por subgrupos e abertura informacional mais densa naquilo que for estritamente necessário para aferir impacto díspar e suas causas, o que se amplifica ao passo que se amplificam as vulnerabilidades ou a opacidade, entre interseccionalidades e manutenção de segredos resistidos pelas desenvolvedoras. Esse modelo evita tanto a arbitrariedade tecnocrática quanto a retórica vazia de princípios, ao exigir que toda intervenção seja calibrada pela proporcionalidade material e pela certeza epistêmica mínima das premissas.

No fundo, a equação da ponderação material revela-se menos como cálculo e mais como ato de resistência hermenêutica: recusar que a técnica dite sozinha o destino humano e insistir que o Direito permaneça como espaço de justificação pública. Cada abertura proporcional — seja em A1, A2 ou A3 — é uma forma de

reacender o lume de Prometeu contra a escuridão da caixa-preta; é afirmar que, mesmo no coração da automação, a razão jurídica ainda pode impor limites, restaurar equilíbrios e operar um princípio de prioridade protetiva quando estejam, em jogo sujeitos histórica e estruturalmente subalternizados, protegendo os frágeis e vulneráveis contra os donos do poder estabelecido (econômico na nova dinâmica de mundo).

A lição é simples e importante: se o algoritmo tende a repetir o passado como destino, a ponderação devolve ao futuro sua condição de escolha. Não há neutralidade onde persiste a desigualdade, nem segredo intangível quando a vida concreta de pessoas em vulnerabilidade está em jogo. A racionalidade constitucional não é apenas ferramenta técnica, mas memória de justiça ao recordar que a dignidade humana é inegociável e que nenhuma inovação pode consumi-la em silêncio. É papel do jurista, em um Estado Democrático-Constitucional de Direito, o aperfeiçoamento de sua técnica e de seu escrutínio hodierno para repercutir transformações sociais, ainda que a partir do direito, efetivando o dever de proteção estatal e os direitos fundamentais de todos os seres. Ao criar um ambiente de respeito e dignidade, há um favorecimento geral, estrutural e histórico.

Assim, o capítulo se encerra como se inicia: com a promessa de que, entre as cinzas da opacidade e o fogo da explicação, é sempre possível fazer do Direito a chama que vigia e que guarda — não para deter o avanço da máquina, mas para assegurar que ela jamais suplante a condição humana de ser livre e igual, de modo digno e substancial.

6 CONCLUSÃO

> *Se tu próprio as quiseres ouvir, que eles te amarrem de pés e mãos, de pé na carlinga do barco veloz, e que as pontas das cordas prendam fora do teu alcance (Homero, 2006, p. 142)*

Odisseu sabe, por aviso de Circe, que nenhum humano resiste ao canto das Sereias. Ordena então que os remadores tampem os ouvidos com cera e impõe a si mesmo uma limitação: ser amarrado ao mastro, com a instrução de ser atado ainda mais forte se suplicar para ser solto. A travessia dá-se, assim, sem naufrágio: ele ouve, conhece e sobrevive. A lição é simples e decisiva para o nosso tempo: certas seduções só se tornam seguras quando a liberdade escolhe limites públicos e verificáveis.

O canto moderno é a promessa tecnossolucionista da modernidade sistemas com mais eficiência, mais precisão, mais assertividade, mas que escondem algo por debaixo da água turva, opaca. O mastro é a Constituição e suas razões públicas; as cordas, os deveres de explicabilidade, rastreabilidade, auditabilidade e supervisão humana significativa; a cera nos ouvidos, os filtros institucionais que impedem que narrativas de performance substituam prova e fundamento. Entre Cila (a anomia da caixa-preta) e Caríbdis (a mora regulatória que apenas posterga decisões), a travessia exige vínculos autoimpostos que preservem agência e responsabilidade na busca pelo entendimento acerca das decisões automatizadas para uma correta tomada de decisão acerca de um posicionamento claro.

Tudo que se defende aqui se reconcilia: quem decide deve justificar; quem concentra dados e modelo assume ônus informacional proporcional ao risco e ao impacto; a abertura do sistema segue níveis graduados (A1–A3) definidos por proporcionalidade; o segredo concorrencial subsiste, mas cede no estritamente necessário ao contraditório efetivo. Retomando Josef K., a conclusão é um antídoto contra o “processo sem autos”: ouvir o canto sem perder o leme.

Em *Os Miseráveis*, Jean Valjean reduzido ao número e ao passado — “um dado” que apaga o sujeito — espelha o risco do determinismo algorítmico: uma variável *proxy* (o rótulo penal) que captura toda a identidade e bloqueia qualquer reclassificação justa. O inspetor Javert encarna a legalidade sem ponderação; o

Direito, porém, exige proporcionalidade. É essa a lição para a era dos escores: proibir que um único atributo, histórico ou dívida, se torne o destino a replicar o ciclo da miséria.

Em *Frankenstein*, não é o “monstro” que nasce culpado, é o criador que abdica do dever de cuidado. A criatura pede para ser escutada, pede para ser ensinada o seu contexto e clama por seu reconhecimento enquanto ser — exatamente o que sistemas opacos recusam aos afetados. O paralelo é direto: quem projeta e opera modelos não pode largá-los no mundo. Deve documentar tudo, desde o desenho, o treinamento com curadoria da base de dados, monitorar seus resultados, conduzir seu aprendizado por *feedbacks,* e, também, ao final, quando da implementação e utilização, explicar os resultados quando interpelado. O verdadeiro monstro não é a criatura, mas a recusa em reconhecê-la como obra de nossas próprias mãos.

Ao final do percurso, retomando em ciclo o vestíbulo inicial, a advertência de Eliot (1934) adquire densidade conclusiva: uma civilização que toma a acumulação informacional por sabedoria repete circularmente a mesma perda do fundamento. O mito do progresso tecnológico não absolve o Direito de sua tarefa prudencial; ao contrário, impõe humildade epistêmica e disciplina institucional para que a decisão automatizada se submeta a critérios de justificação e de responsabilidade, evitando uma simplificação tecnocrática da vida. Nessa reordenação – vida, sabedoria, conhecimento, informação – reside a promessa de uma jurisdição que recusa o ruído e as alucinações. A tarefa do Estado-juiz é reafirmar o telos jurisdicional para as aplicações algorítmicas.

Dentro da era digital, emergem desafios críticos relacionados à discriminação algorítmica e à proteção dos direitos fundamentais, demandando uma abordagem jurídico-constitucional para equacionar o conflito entre ambos. A igualdade, em sua forma material, é o princípio central na discussão da vulnerabilidade, merecendo atenção especial para grupos que demonstram fragilidades não comuns na pessoa média, como consumidores, idosos e jovens. A vulnerabilidade, agravada pela discriminação estrutural, sugere a necessidade de uma tutela jurídica diferenciada, considerando as múltiplas e interseccionais facetas da injustiça.

Essa nova forma de discriminação, que na verdade é apenas uma roupagem automatizada dos próprios estigmas humanos ainda presentes, potencializados pela automatização da internet, está renovando tratamentos explícitos que já deveriam ter

sido erradicados, o que passa também por um distanciamento existente entre a ponta de criação das programações e os usuários, que em regra se dão na origem destas Big Techs e outras empresas de tecnologias do meio norte do mundo. A colaboração entre empresas, governos, organizações e sociedade civil é fundamental para enfrentar os desafios da discriminação algorítmica, desde a criação de espaços de diálogo e cooperação para discutir e debater os efeitos dos algoritmos até a implementação de fóruns permanentes e de autoridades competentes para criar soluções e atender a demandas específicas de práticas algorítmicas discriminatórias a indivíduos, grupos ou comunidades.

O direito pode ser visto como uma tecnologia: não é exatamente uma técnica nova, mas sim criativa e engenhosa na resolução prática de problemas jurídicos oriundos do conflito humano, restaurando a dignidade tanto dos envolvidos quanto da sociedade. Isso não significa que seja algo impossível de entender; pelo contrário, sua principal qualidade está em explicar racionalmente suas premissas e conclusões, com impactos jurídicos e práticos, detalhando os caminhos de sua fundamentação – baseados em fatos ou normas, estas podendo ser mais ou menos específicas.

Em sendo falhas nessa resposta, o Poder Judiciário, no entanto, servirá como última barreira, a efetivar o dever de proteção estatal (*Schutzpflicht*) do direito à não discriminação. Os magistrados estão autorizados a otimizar e compatibilizar os direitos fundamentais e, portanto, o descumprimento de preceitos fundamentais, como a não-discriminação, admitem a relativização de direitos de segredo algorítmico, de propriedade das empresas, para a salvaguarda de direitos fundamentais mais relevantes em casos; movimento esse que ganha forças a partir do movimento da startup chinesa DeepSeek, que lançou sua inovação tecnológica em código aberto (auditável), além de uma maior sustentabilidade, pela maior eficiência energética, dois movimentos que podem alterar significativamente toda a estrutura do setor de tecnologia e dos algoritmos.

A inteligência artificial, por mais sofisticada que seja, permanece derivativa do saber humano: amplia alcance e velocidade, mas não substitui a prudência nem corrige, por si, as insuficiências do conhecimento. Não é entidade autônoma e abstrata, e sim um arranjo técnico erigido sobre experiências e dados do mundo – um recorte, quando não apenas uma sombra, do real. Por isso, sua presença no decidir deve ser instrumental e subsidiária, jamais finalística. Convertida enfim, a técnica

tende a aprisionar a deliberação em representações opacas do social, projetadas por modelos que refletem – e podem distorcer – a realidade. O avanço tecnológico não elimina o risco de empobrecimento explicativo; ao contrário, impõe vigilância redobrada para que a razão prática continue à frente da automação. É com os pés firmes no presente concreto que o direito deve ordenar o uso de sistemas algorítmicos, para que sirvam como meios de aperfeiçoar a vida comum – e não como novas autoridades do decidir.

A crescente e recente preocupação com a proteção de dados tem justificativa. O temor pela coisificação é uma realidade, atinge a todos e é lucrativo para empresas de certos ramos apoiados na inovação tecnológica. A proteção de dados, assim, assume papel de escudo aos direitos e garantias individuais. A concretização do direito fundamental específico à proteção dos dados pessoais viria como decorrência lógica das movimentações normativo-constitucionais brasileiras, e mundiais, acabou sendo concretizada via Emenda Constitucional. Na verdade, bastava a constatação, por parte do Supremo Tribunal Federal, uma vez que a organização protetiva já era de estrutura jusfundamental.

Alicerçou-se, ao longo do texto, que a LGPD não inovou em proteger dados pessoais, sensíveis ou não, em terras brasileiras. O seu mérito é o da consolidação de uma ordem jurídica de proteção de dados, bem como de um retorno da preocupação acerca desse tema, fruto dos incidentes que assolavam a Internet e seus usuários na última década. Viu-se também que a LGPD, além de organizar a tutela protetiva dos dados, lança bases gerais à nova legislação que deverá vir à efetividade para regular o uso e a existência da Inteligência Artificial no Brasil.

No específico, delineou-se os usuários-consumidores, ou usuários-titulares, como o grupo majoritário visado pela lei. Desse modo, nota-se a importância da discussão acerca da temática da proteção de dados na tutela da dignidade do usuário, que se torna verdadeiro fornecedor de dados, que posteriormente são utilizados para influenciá-lo, sem transparência.

Foi possível notar, também, mediante o emprego de alguns casos práticos, que a realidade de vazamento de dados e desrespeito da legislação vigente existe, acarreta danos e consolida situações. Além disso, e de modo mais acentuado, viu-se como vícios na coleta, armazenamento e uso dos dados pode ferir gravemente direitos fundamentais como o princípio da igualdade, em sua vertente de vedação à

discriminação. A discriminação algorítmica, que merece estudo detido em outras oportunidades, não pode ser um sintoma apenas lateral das práticas comerciais. O custo social do enviesamento algorítmico é alto, e seu nascedouro está no modo como se tutelam os dados pessoais. O Estado Democrático Brasileiro, tanto por seu corpo político, como por sua jurisdição constitucional, tem o dever de tutelar o consumidor em rede, prevenindo-o de ser tratado como um barril de petróleo, como uma mercadoria.

É exatamente nesse ponto que o percurso normativo se reencontra com o diagnóstico fático: a proteção de dados fornece o alicerce, mas a resolução dos riscos algorítmicos exige governança ativa — deveres positivos que transformem direitos em práticas verificáveis de desenho, validação e controle, sob parâmetros de proporcionalidade e de tutela preferencial de vulneráveis.

A consolidação da LGPD como regime protetivo não esgota o problema: a governança de riscos algorítmicos exige que o direito operacionalize deveres positivos de prevenção, transparência e revisão humana significativa quando houver impactos relevantes a direitos fundamentais. Nesse ponto, a tensão entre autodeterminação informativa/explicabilidade e segredo comercial deve ser resolvida por controle de proporcionalidade material, com resposta calibrada ao grau de afetação de grupos vulneráveis, evitando tanto a opacidade tecnocrática quanto a devassa desnecessária de ativos empresariais.

À luz dessas premissas, enunciam-se, em chave sintético-operativa, as conclusões a que se chega ao longo dos capítulos; organizando o diagnóstico, a moldura normativa e o procedimento decisório proposto. Nesse cenário, chega-se às seguintes conclusões objetivamente pontuadas:

1. A desigualdade é estrutural e migra potencializada ao digital, ganhando novas engrenagens. A vulnerabilidade no mundo digital não é acidente, mas arquitetura social que, no ambiente automatizado, reaparece por variáveis-proxy, bases enviesadas e ciclos de retroalimentação. Ela se reconfigura por meio de processos de datafication e de inferência a partir de padrões pretéritos. O efeito é cumulativo: o que era desvantagem histórica converte-se em tendência decisória discriminatória. Nesse cenário, o dever estatal de proteção não pode ser meramente reativo; ele se adianta às formas

contemporâneas de exclusão, garantindo que a igualdade material e a não discriminação moldem o emprego de sistemas algorítmicos quando incidirem sobre sujeitos em condição múltipla e interseccionalmente vulnerável.

2. Igualdade material demanda sindicabilidade reforçada e não aparente neutralidade. Quando o impacto recai sobre grupos vulneráveis, a jurisdição abandona a ficção de uma neutralidade autossuficiente e assume a tarefa de correção estrutural: intensifica o escrutínio sobre as razões, aprofundando o contraditório e o dever de fundamentação. O padrão de controle cresce para impedir que a técnica cristalize o passado desigual como destino. O que se exige é justificação pública compatível com a gravidade da interferência e com o peso dos direitos em jogo, de modo que o resultado não seja fruto de performance, mas de razões verificáveis.

3. Opacidade é governança, não fatalidade técnico-programada. As zonas de sombra decorrem de decisões de desenho e de gestão do ciclo de vida dos modelos; não são imposições naturais da tecnologia. Por isso, a resposta jurídica incide sobre a governança – isto é, sobre como se concebe, valida, aplica e revê o uso de sistemas – porque tais escolhas produzem opacidades que não podem ser racionalmente toleradas. A intervenção normativa deve recair tanto sobre regras de governança (seguidas por regulação ou por economia processual) quanto – e não apenas – sobre o produto final da decisão automatizada, ajustando a intensidade do controle ao risco projetado e ao lugar do vulnerável na situação concreta; quando a decisão se efetivar, suas repercussões cabíveis ocorrerão nas esferas administrativas (da empresa ou do Estado, se por ele comandada ou utilizada) e/ou judiciais.

4. Transparência é condição de validade e de continuidade do uso de IAs e Algoritmos no Brasil; explicabilidade é o padrão. A permanência legítima de sistemas algorítmicos depende de transparência; a sua aceitabilidade constitucional exige explicabilidade proporcional — a reconstrução, em linguagem pública, do itinerário lógico-decisório que liga dados, critérios e resultado. Se essa justificação falhar, a atuação jurisdicional intervém na exata medida do risco para determinar esclarecimentos e permitir controle efetivo.

5. LGPD e CPC operam como ponte normativa. Os princípios da LGPD (art. 6º) e o direito de revisão/explicação (art. 20), articulados ao dever de motivação adequada (CPC, art. 489, § 2º), convertem o discurso da proteção de dados e da não discriminação em standards de decisão: orientam a prevenção, estruturam a prestação de contas e delimitam a revisão das escolhas automatizadas. Essa ponte constitucionalmente ancorada pelo dever de proteção assegura que a justificabilidade das decisões não seja um adorno retórico, mas uma exigência operacional que vincula Administração, jurisdicionados e particulares em relações assimétricas de poder.

6. A aplicação da proposta interpretativa é irredutivelmente dependente do caso concreto – dos dados disponíveis, da intensidade do impacto e da configuração dos direitos em colisão. Por isso, não há – nem deve haver – regra *ex ante* que resolva a tensão entre explicabilidade e segredo industrial antes da emergência de um conflito normativo específico. O que se oferece, aqui, é um procedimento graduado e controlável que organiza o juízo de proporcionalidade com correspondência probatória exigível e padrão de certeza (C) calibrado pelas evidências do processo. Assim – e apenas assim – evita-se tanto o decisionismo quanto a retórica da transparência. Ainda que se reconheça que mesmo *a priori* parece claro que haverá um peso maior para a transparência algorítmica e um ônus maior para o segredo algorítmico.

7. A matriz A1–A3 não é mero rótulo classificatório, mas o mecanismo operativo que ordena a intensidade do controle segundo quatro vetores: impacto sobre direitos fundamentais, situação de poder do controlador (se *big tech*, para verificar a eficácia jusfundamental), opacidade técnico-organizacional e incidência sobre grupos em situação de vulnerabilidade. A partir deles, calibra a abertura informacional, a densidade documental, a supervisão humana significativa e a perícia sob controle judicial.

8. Em A1, onde o impacto é limitado, a opacidade é moderada, basta uma explicabilidade documental robusta, com descrição dos critérios decisórios e métricas mínimas que foram utilizadas que demonstram de ali decorrer o viés discriminatório. Com isso, já se eleva o contraditório da esfera retórica à esfera

técnica, fixando-se um C mínimo plausível e exigindo-se demonstração de alternativas menos gravosas já consideradas.

9. Em A2, quando o impacto se adensa, a opacidade cresce, sobretudo com indícios de impacto mais gravoso sobre grupos vulneráveis. Com isso, a abertura informacional se aprofunda: além da documentação e das métricas por subgrupos, já se permite, caso já haja processo, perícia algorítmica independente, sob sigilo judicial, com acesso controlado a artefatos técnicos, elevação do C a patamar seguro nos pontos críticos e ônus probatório dinâmico deslocado ao operador para suprir assimetria informacional. Aqui, se não judicializado, na instância administrativa poderá ser exigido Relatórios de Impacto de Dados e Algorítmicos, e indicação inicial de AIA – o que ainda precisa ser definido pelo campo da Regulação (que não é alvo do presente esforço).

10. Em A3, diante de alto impacto e atingimento de bens essenciais, sobretudo se sobre grupos vulneráveis, a abertura é profunda e estritamente necessária: documentação exaustiva, métricas e testes reprodutíveis por subgrupos, auditoria metodológica completa com o perito como eixo epistêmico e, quando indispensável, exposição controlada da lógica decisória, sob sigilo judicial e com abertura escalonada e *reports* específicos. Maximizar a confiança epistêmica (C) do que ocorre na aplicação e permitir a reconstrução pública do itinerário lógico-decisório é o alvo nesse grau.

11. Quando o Estado está envolvido na aplicação com potencial discriminatório, recomenda-se aumentar o nível de transparência, especialmente em casos que envolvem liberdade, segurança, saúde, educação, benefícios ou políticas públicas específicas. Para empresas privadas de médio porte, sugere-se manter o nível de transparência adequado ao contexto, mas facilitar a forma de apresentação das informações, como por meio de exemplos ou explicações detalhadas, resguardando a confidencialidade quando necessário. No caso de plataformas operadas por grandes empresas de tecnologia, indica-se ampliar ainda mais a transparência, exigir registros completos, explicações acessíveis e verificação

independente sempre que houver sinais de desigualdade sistemática. Para casos conduzidos pelo poder público e por grandes plataformas, o grau de confiança exigido nas informações e verificações segue a régua: nível 1: baixo; nível 2: baixo a médio; nível 3: a partir de médio, ajustado para cima quando o impacto sobre liberdades ou a escala dos efeitos assim o impuser.

12. A régua do escalonamento se traduz em medidas proporcionais e verificáveis: quanto mais alto o degrau, mais ampla a abertura, mais densa a documentação e mais intensa a supervisão e a perícia, com correspondência necessária entre I (intensidade da interferência), PeA (peso abstrato do princípio que cede) e o C exigido pelas premissas que sustentam a decisão.

13. Ônus probatórios dinâmicos. No ecossistema algorítmico, a assimetria informacional é constitutiva; por isso, o ônus da prova não permanece estático, mas se desloca segundo a capacidade de produção de prova técnica. À luz do constitucionalismo digital, os direitos fundamentais projetam eficácia horizontal – vinculando particulares entre si – e eficácia vertical – destinada à relação entre Estado e indivíduo, por meio do dever estatal de proteção e da regulação que densifica deveres públicos. Há, no entanto, que se pensar em nova eficácia, talvez uma a ser nomeada de “verticalizada”, decorrendo das novas forças cibernéticas – grandes empresas e corporações que controlam o meio digital, a produção de dados e o desenvolvimento de IAs e algoritmos.

14. Perito algorítmico – eixo epistêmico da sindicabilidade. O perito algorítmico é peça de articulação entre a exigência constitucional de justificação pública e a preservação do segredo remanescente: atua sob sigilo judicial, conduz testes reprodutíveis, valida métricas estratificadas por subgrupos, reconstrói o a decisão a partir dos dados, pelo modelo de tratamento e processamento destes, até o modelo da aplicação algorítmica e, por fim, a decisão automatizada, registrando incertezas e limites inferenciais. Reporta apenas aquele achado estritamente necessário para a verificação dos resultados discriminatórios. Em A1, sua intervenção é, em regra,

desnecessária; em A2, apenas se judicializado, converte-se em instrumento de equalização informacional; Em A3, a perícia é regra.

15. Segredo comercial-algorítmico como bem jurídico – com segredo remanescente. O segredo comercial-algorítmico é protegido pelo ordenamento e integra a livre iniciativa e a concorrência, mas, no constitucionalismo digital, sofre conformação pelos direitos fundamentais com eficácia horizontal (desenvolvedoras não hegemônicas, vertical (IAs controladas pelo Estado) e verticalizada (*Big techs*) – esta última incidindo, de modo reforçado, sobre plataformas dominantes. A solução é proporcional e escalonada: abre-se o suficiente para viabilizar controle de discriminação e motivação pública, preservando tudo o que não seja estritamente necessário à verificação das razões.

16. Modelo decisório proposto. O procedimento de tomada de decisão racional proposto se baseia em uma arquitetura decisória viva, sem se ater a um catálogo de ferramentas estanques. Buscando sempre reduzir assimetrias informacionais, fortalecendo a igualdade material e consolidando uma cultura de justificação pública das escolhas automatizadas, sem sacrificar a inovação nem o ambiente concorrencial. O modelo integra (i) diagnóstico de vulnerabilidades estruturais; (ii) enquadramento constitucional e legal orientado pelo dever de proteção e pela igualdade material; e (iii) um procedimento racional que depende da verificação, *in casu*, de classe de risco (A1–A3), da intensidade da interferência (I) e do peso abstrato (PeA) dos princípios em colisão, do grau de certeza exigível (C) para legitimar a preferência, da alocação dinâmica dos ônus de justificação conforme a assimetria informacional, do lugar do segredo comercial-algorítmico como segredo remanescente e da necessidade de intervenção pericial enquanto tradução técnica em razões públicas. A decisão que emerge do modelo não deve ser uma aposta, mas uma reconstrução justificável do itinerário lógico que conecta fatos, normas, dados, *inputs, outputs* e consequências sobre pessoas concretas.

A partir do uso de Inteligências Artificiais (*disclaimer* ético de uso), foram montadas tabelas de operação do modelo criado, que seguem em quatro situações – análise de crédito (empresa privada), policiamento ostensivo (Estado), uma revisão do case do COMPAS (Estado-juiz), e a seleção de entrega de anúncio para vagas de trabalho (*big tech*):

Dimensão	Nível 1 (impacto menor; transparência razoável)	Nível 2 (impacto médio; sinais de desigualdade)	Nível 3 (alto impacto; risco grave de injustiça)
Quem decide	Empresa privada média	Empresa privada média	Empresa privada média
Documentos mínimos	Objetivo do sistema; como funciona em linhas gerais; quais informações são consideradas; de onde vêm os dados; registros de cada decisão (data e hora, quais informações foram usadas, nota atribuída, ponto de corte aplicado)	Tudo do nível 1 + indicar quais informações pesaram mais no caso; regras internas que orientam a aprovação/negativa; amostra significativa de decisões para verificação independente	Tudo do nível 2 + trilha completa de como a decisão foi formada; descrição do processo de verificação feito pelo responsável pelo sistema
Como verificar justiça e acerto	Comparar taxas de aprovação entre grupos comparáveis; observar erros que liberam e erros que negam; verificar se as notas condizem com os resultados ao longo do tempo	Tudo do nível 1 + ver se, com o mesmo padrão de risco, grupos diferentes recebem tratamentos diferentes; testar pontos de corte diferentes e ver quem passa a ser aprovado ou reprovado em cada grupo	Tudo do nível 2 + examinar, caso a caso, o que muda quando se altera apenas uma informação, para detectar decisões injustificadamente desfavoráveis
Exame por especialista indicado pelo juízo	Em regra, dispensável	Pode ser necessário, com acesso controlado a	Necessário, com revisão completa do modo de funcionamento e

Dimensão	Nível 1 (impacto menor; transparência razoável)	Nível 2 (impacto médio; sinais de desigualdade)	Nível 3 (alto impacto; risco grave de injustiça)
(com confidencialidade)		documentos e parte do funcionamento	repetição dos testes
Grau de confiança exigido	Satisfatório	Elevado	Muito elevado
Abertura de informações	Documentos e registros	Documentos e registros + acesso controlado a elementos essenciais	Exposição controlada da lógica da decisão, somente o indispensável
Proteção do que é segredo legítimo	Guardar em sigilo o que não for necessário para verificar a justiça	Idem	Idem

2

Dimensão	Nível 1 (impacto menor)	Nível 2 (impacto médio)	Nível 3 (alto impacto sobre liberdades)
Quem decide	Estado	Estado	Estado
Documentos mínimos	Objetivo do sistema; período considerado; fontes de dados; aviso sobre possíveis faltas de registro	Tudo do nível 1 + descrição de efeitos de reforço (quando mais presença gera mais registros e retroalimenta o sistema); justificativa das informações usadas; mapas das áreas priorizadas	Tudo do nível 2 + relato completo das verificações feitas ao longo do tempo, comparando previsões passadas com o que de fato ocorreu
Como verificar justiça e acerto	Comparar acertos e erros entre bairros e entre grupos; medir abordagens que nada encontram e onde ocorrem; avaliar se as áreas indicadas como “quentes” realmente o são	Tudo do nível 1 + comparar diferenças nas abordagens e nos erros entre áreas com perfis sociais distintos; medir erro de localização	Tudo do nível 2 + realizar testes de “e se” (alterar apenas um fator e ver se o resultado muda); verificar concentração injusta em certos territórios
Exame por especialista	Admissível quando houver indícios	Recomendado	Obrigatório, com revisão completa

Dimensão	Nível 1 (impacto menor)	Nível 2 (impacto médio)	Nível 3 (alto impacto sobre liberdades)
indicado pelo juízo (com confidencialidade)			e repetição de testes
Grau de confiança exigido	Baixo	Baixo a médio	A partir de médio
Abertura de informações	Documental ampla	Documental + acesso controlado a elementos técnicos essenciais	Acesso técnico mais amplo, com explicação controlada de como as prioridades são definidas
Proteção do que é segredo legítimo	Guardar apenas o que não for necessário para verificar a justiça	Idem	Idem

3

Dimensão	Nível 1	Nível 2	Nível 3 (caso típico, pois afeta liberdade)
Quem decide	Estado (Justiça/Execução penal)	Estado (Justiça/Execução penal)	Estado (Justiça/Execução penal)
Documentos mínimos	Finalidade (o que se pretende estimar); quais informações entram; origem dos dados	Tudo do nível 1 + manual claro de funcionamento; critérios de ajuste para que a pontuação reflita a realidade	Tudo do nível 2 + documentos completos sobre verificações feitas; pontos de corte adotados e razões de sua escolha
Como verificar justiça e acerto	Comparar, para cada grupo, onde o sistema erra para mais e para menos; comparar quem recebe rótulo de alto risco em condições semelhantes	Tudo do nível 1 + medir se a chance de acerto é parecida entre grupos; verificar se a pontuação corresponde, em cada grupo, ao que acontece depois	Tudo do nível 2 + analisar erros graves que levam a decisões mais duras; testar como pequenas mudanças de informação afetam o resultado em cada grupo

Dimensão	Nível 1	Nível 2	Nível 3 (caso típico, pois afeta liberdade)
Exame por especialista indicado pelo juízo (com confidencialidade)	—	Recomendado	Obrigatório, com testes repetidos em subgrupos
Grau de confiança exigido	Baixo	Baixo a médio	A partir de médio
Abertura de informações	—	Documental + acesso técnico controlado	Explicação controlada da lógica da pontuação, somente o indispensável para verificar a justiça do caso
Proteção do que é segredo legítimo	—	Manter o que não for necessário para o controle	Idem

4

Dimensão	Nível 1 (impacto menor)	Nível 2 (impacto médio; indícios de desvantagem)	Nível 3 (alto impacto; risco sistêmico)
Quem decide	Grande plataforma	Grande plataforma	Grande plataforma
Documentos mínimos	Objetivo do sistema; como as vagas são mostradas; regras gerais de segmentação; fontes de dados dos perfis	Tudo do nível 1 + registros por período do que foi exibido para diferentes públicos; critérios para ordenar e ocultar; explicação clara sobre limitações e filtros	Tudo do nível 2 + relatórios completos de verificação, incluindo testes de diferenças de alcance entre públicos e correções aplicadas
Como verificar justiça e acerto	Comparar quem recebeu e quem não recebeu anúncios de trabalho com perfis semelhantes; medir diferenças entre públicos com características parecidas	Tudo do nível 1 + verificar se há bloqueio indireto de grupos para certas vagas; testar mudanças de regra e observar quem passa a ver as ofertas	Tudo do nível 2 + avaliar efeito concentrado de exclusão em larga escala; impor planos de correção com metas e prazos

Dimensão	Nível 1 (impacto menor)	Nível 2 (impacto médio; indícios de desvantagem)	Nível 3 (alto impacto; risco sistêmico)
Exame por especialista indicado pelo juízo (com confidencialidade)	Admissível	Recomendado	Obrigatório, com verificação independente e acompanhamento por período
Grau de confiança exigido	Baixo	Baixo a médio	A partir de médio
Abertura de informações	Documental robusta	Documental + acesso controlado a painéis e relatórios	Acesso mais amplo aos relatórios, com explicação controlada das regras de entrega e priorização
Proteção do que é segredo legítimo	Resguardar apenas o que não é necessário para verificar a justiça	Idem	Idem
Documentos mínimos	Objetivo do sistema; como as vagas são mostradas; regras gerais de segmentação; fontes de dados dos perfis	Tudo do nível 1 + registros por período do que foi exibido para diferentes públicos; critérios para ordenar e ocultar; explicação clara sobre limitações e filtros	Tudo do nível 2 + relatórios completos de verificação, incluindo testes de diferenças de alcance entre públicos e correções aplicadas
Como verificar justiça e acerto	Comparar quem recebeu e quem não recebeu anúncios de trabalho com perfis semelhantes; medir diferenças entre públicos com características parecidas	Tudo do nível 1 + verificar se há bloqueio indireto de grupos para certas vagas; testar mudanças de regra e observar quem passa a ver as ofertas	Tudo do nível 2 + avaliar efeito concentrado de exclusão em larga escala; impor planos de correção com metas e prazos
Exame por especialista indicado pelo juízo (com confidencialidade)	Admissível	Recomendado	Obrigatório, com verificação independente e acompanhamento por período
Grau de confiança exigido	Baixo	Baixo a médio	A partir de médio

Dimensão	Nível 1 (impacto menor)	Nível 2 (impacto médio; indícios de desvantagem)	Nível 3 (alto impacto; risco sistêmico)
Abertura de informações	Documental robusta	Documental + acesso controlado a painéis e relatórios	Acesso mais amplo aos relatórios, com explicação controlada das regras de entrega e priorização
Proteção do que é segredo legítimo	Resguardar apenas o que não é necessário para verificar a justiça	Idem	Idem

Concluída a cartografia prática nos quadros – que, por domínio e por operador, ordenam o quanto abrir, o que registrar e como verificar –, passa-se do mapa ao motor do método: a engrenagem epistêmica que dá unidade ao escalonamento A1–A3 e impede que a proporcionalidade se dissolva em retórica.

Antes de enunciar definitivamente a versão epistêmica da ponderação, cumpre explicitar a engrenagem de racionalidade que sustenta o modelo: sem uma régua de certeza a proporcionalidade arrisca-se a virar retórica. É precisamente aqui que se insere a dimensão epistêmica da lei material da ponderação: a preferência entre direitos só se legitima quando o convencimento sobre as premissas fáticas e normativas acompanha a gravidade da interferência. A escala A1–A3 dá corpo a essa exigência ao transformar risco, reversibilidade, opacidade e incidência sobre vulneráveis em degraus de exigência justificativa, de modo que o quanto abrir da caixa-preta, o quantum de explicação exigível e a própria carga argumentativa não resultem de intuição, mas de um padrão público e verificável de decisão.

A versão epistêmica da lei da ponderação eleva a certeza das premissas (C) ao patamar de condição para a compressão de direitos. Em ambiente algorítmico, C é função de três eixos: documentação rastreável (dados, modelo, logs), auditabilidade (testes reprodutíveis, métricas por subgrupos) e coerência causal do nexo entre critérios e resultados. A matriz A1–A3 é o corpo dessa exigência: em A1, a explicabilidade documental torna “plausíveis” as premissas e viabiliza contraditório; em A2, a perícia (quando justificada e em juízo) sob sigilo eleva C a patamar “seguro” para os pontos críticos; em A3, a abertura profunda – sempre estritamente necessária – maximiza C quando o impacto é grave/irreversível sobre bens essenciais e grupos

vulneráveis, na medida do necessário (em um escalonamento fático e registrado pelo perito). Quanto maior I (intensidade da intervenção) e PeA (peso abstrato do princípio afetado), maior deve ser C; e é A1–A3 que entrega C, em graus de verificação posteriormente ao escrutínio de uma decisão (ainda não efetivada).

Dessas premissas epistêmicas e do escalonamento A1–A3 decorre um desdobramento institucional imediato: padrões regulatórios e deveres de governança devem traduzir o modelo em obrigações concretas de desenho, validação e auditoria – sem os quais a cultura de justificação permaneceria retórica.

Nesse contexto, a criação de regulamentações específicas para o uso de algoritmos em decisões importantes e o estabelecimento de critérios claros para o desenvolvimento desses sistemas são medidas essenciais para garantir a transparência e a equidade nas decisões automatizadas, enquanto isso não ocorre, tal abertura material para o reconhecimento do direito fundamental à explicabilidade será vital para a tutela constitucional dos direitos afetados. Afinal, é fundamental considerar as dimensões humanas em que a privacidade e o mal ou mau processamento de dados pode ocasionar, daí a importância da tutela à autodeterminação informativa, como escudo e controle pelos indivíduos perante as decisões automatizadas que se valem de seus dados.

A autodeterminação informativa é explorada para compreender em quais dimensões humanas a privacidade e os dados podem ir de encontro à dignidade humana, ao se imiscuir de demonstrar seu adequado uso, em prejuízo, ou mesmo benefício, do cidadão. Em seguida, é explorada a dimensão metafísico-constitucional do direito à participação do ser humano no mundo, autonomizando-o enquanto um ser que guarda dignidade em si e que, só por isso, já é merecedor de uma tutela que lhe garanta a visualização dos processos que lhe dizem respeito. Um paralelo entre a obnubilação do ser em regimes autoritários serve de paralelo para se analisar o que é a nova realidade algorítmica.

A abordagem metafísico-constitucional do direito à participação do ser humano no mundo é relevante para a compreensão das questões éticas envolvidas na automatização de processos de tomada de decisão. A transparência e a explicabilidade são essenciais para garantir a justiça e a equidade na sociedade atual, protegendo-se a dignidade dos indivíduos afetados.

A transparência e a explicabilidade desses sistemas são fundamentais para garantir a compreensão e contestação das decisões tomadas, especialmente em casos em que elas podem ser discriminatórias ou violar os direitos dos indivíduos, no que poderão desenvolvedores, conselhos, juízes e tribunais se debruçar para encontrar os defeitos e avaliar a extensão dos danos, com seu nexo causal, e mesmo a culpabilidade dos desenvolvedores e, a partir de uma regulação, dos acreditadores públicos. Nesse sentido, a jurisdição constitucional, que tutelará tais direitos, tanto nesse primeiro momento quanto em última instância, deverá ser capaz de encontrar as bases para exigir, ainda que minimamente, a proteção a direitos fundamentais prejudicados pelas aplicações, com especial relevo para a descoberta das razões acerca do cometimento, ou mero acontecimento, de discriminações automatizadas. Há um dever jurídico intrinsicamente motivado por substratos morais que determina a proteção e a defesa de usuários em uma nova dinâmica de putativa soberania cibernética de empresas e desenvolvedoras de aplicações algorítmicas.

Aliás, o Direito é também uma tecnologia. Uma técnica nada inovadora, mas bem inovativa e inventiva (*Ius est Ars inveniendi*) de solucionar conflitos humanos. E nem por isso deixa de ser explicável. Aliás, está aí a sua maior virtude: a capacidade de racionalmente demonstrar as premissas e suas conclusões com efeitos jurídicos e práticos, e os caminhos de sua fundamentação, seja em premissas fáticas ou normativas, sendo essas últimas mais ou menos concretas.

Assim a preferência em colisões que envolvam alegação de discriminação algorítmica será legítima se: (i) classificado o caso em A1–A3; (ii) identificados a intensidade da interferência (I) e o peso abstrato (PeA) dos princípios em jogo; (iii) fixado o grau de certeza (C) compatível com a gravidade da medida; (iv) calibrada a abertura informacional no exato grau exigido por A1–A3; (v) alocados os ônus probatórios de forma dinâmica, impondo ao operador a produção da base técnico-justificativa mínima; e (vi) motivada a decisão com razões públicas suficientes. Em termos normativos: quanto maior I e PeA do princípio interferido, tanto maior deve ser C e mais ampla a abertura informacional exigida, preservado o segredo remanescente. E quando o caso envolver grupos vulneráveis, a régua se desloca um degrau para cima a partir de justificação casuística que demonstre a necessidade desse deslocamento *a posteriori*. Aqui, o operador assume a carga de demonstrar que o resultado não reproduz desigualdades estruturais, assegurando um benefício da

dúvida em prol da pessoa discriminada e do grupo vulnerável. Nessa sorte, é preciso garantir que a técnica sirva à dignidade e à igualdade material, buscando a consolidação cada vez maior da não-discriminação com o mínimo de interferência que se demonstra racionalmente necessária no segredo comercial, e não o inverso.

É justamente aqui que o argumento alcança o seu ponto mais delicado: quanto maior a promessa de eficiência, mais sutil tende a ser a captura do juízo, porque a racionalidade se disfarça de mera técnica – e a técnica, por sua vez, naturaliza-se como destino. Se a discriminação algorítmica se alimenta de opacidades – epistêmicas, institucionais e materiais –, a resposta jurídica não pode limitar-se a reafirmar valores: ela precisa organizar condições de possibilidade para que esses valores sejam exigíveis no mundo, isto é, para que a decisão permaneça imputável, contestável e controlável. Em outras palavras, o problema não é apenas o conteúdo do que se decide, mas a arquitetura do decidir: quais garantias públicas impedem que a conveniência do sistema substitua a justificabilidade perante o sujeito atingido.

A metáfora de Odisseu, contudo, ganha densidade adicional quando lida pelo itinerário crítico de Jon Elster (1979), para quem o pré-compromisso – o ato de impor amarras a si próprio – não é a romantização da autorrestrição virtuosa, mas o reconhecimento de que a racionalidade, individual ou institucional, precisa de arranjos para sobreviver ao próprio impulso. Mais do que isso, Elster problematiza a transposição direta do indivíduo para o corpo político, advertindo que, no mundo institucional, muitas vezes não se trata de o agente limitar-se a si próprio, mas de limitar terceiros – e é nesse deslocamento que a questão retorna ao eixo do poder e da legitimidade. Nessa chave, a travessia jurídica da era algorítmica exige um cuidado duplo: entre a Sila, deve-se construir amarras públicas verificáveis contra o canto tecnocrático, e entre Caríbdis, impedir que essas mesmas amarras sejam capturadas como técnica de governo sobre os vulneráveis, isto é, como dispositivo de heterocontenção travestido de neutralidade. A conclusão é que o “mastro” constitucional não serve para sufocar a agência – serve para torná-la possível, devolvendo ao afetado a condição de sujeito de razões e impondo ao Estado o dever de transformar estatísticas em justificações públicas controláveis.

Contra o labirinto sem informação (sem autos), contra a criatura sem tutor e contra o número que sufoca a pessoa, em busca de conhecimento, o gesto de Odisseu

– amarrar-se ao mastro – torna-se método inicial do universo jurídico para atravessar a tormenta da era algorítmica. Ver, até desejar, mas não sucumbir.

REFERÊNCIAS

ADADI, Amina; BERRADA, Mohammed. Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). **IEEE Access.** v. 6, p. 52138-52160, 2018.

ALEXY, Robert. **La doble natureza del derecho.** Madrid: Trotta, 2016.

ALEXY, Robert. Princípios formais. 2022. *In*: TRIVISONNO, Alexandre Travessoni Gomes; BOROWSKI, Martin. **O debate sobre a teoria dos princípios fundamentais de Robert Alexy.** São Paulo: Marcial Pons, 2022.

ALEXY, Robert. **Teoria da argumentação jurídica:** a teoria do discurso racional como teoria da fundamentação jurídica. Trad. Zilda Hutchinson Schild Silva. 5. Ed. São Paulo: Forense, 2020.

ALEXY, Robert. **Teoria dos Direitos Fundamentais**. Trad. Virgílio Afonso da Silva. 2. Ed. São Paulo: Malheiros, 2015.

ALMEIDA, Silvio Luiz de. **Racismo Estrutural**. São Paulo: Pólen, 2019.

ÁLVAREZ, José Luis Domínguez. El paulatino reconocimiento de Los derechos digitales como precisa para garantizar la dignidad de la persona ante el avance del mundo digital. In: DÍEZ, José María Soberanes; DOMÍNGUEZ, Gustavo Garduño (Coords). **La interacción de Las redes sociais, la tecnología y Los derechos humanos.** Pamplona: Ediciones Universidad de Navarro, p. 101-117, 2023.

ÁLVAREZ, José Luis Domínguez. **IUSALGORITMIA: Las reglas de desenho y la inteligencia artificial**. Valencia: Tirant lo Blanch, 2024.

ANARTE, Laura Flores. Sesgos de Género En La Inteligencia Artificial: El Estado de Derecho Frente a La Discriminación Algorítmica Por Razón de Sexo. **Revista Internacional de Pensamiento Político = International Journal of Political Thought**, v. 18, 2023, p. 95–120. Acesso em: https://dialnet.unirioja.es/servlet/articulo?codigo=9273666. Disponível em: 18 set. 2024.

ANDRÉS, Maisés Barrio. ¿Como deben ser regulados Los algoritmos?: la regulación europea y española de la inteligencia artificial. In: NOVOA, César García; RODRÍGUEZ, Francisco Hernández (Dirs.) **Estúdios sobre inteligência artificial y economía digital.** Madrid: Aranzadi, p. 47-69, 2024.

ARENDT, Hannah. **As origens do totalitarismo.** Trad. Roberto Raposo. São Paulo: Companhia das Letras, 2012.

ARENDT, Hannah. **Eichmann em Jerusalém: um relato sobre a banalidade do mal.** São Paulo: Companhia das Letras, 1999.

ARENDT, Hannah. **Liberdade para ser livre.** Trad. Pedro Duarte. Rio de Janeiro: Bazar do Tempo, 2019.

ARISTÓTELES. **A Política.** 3. Ed. São Paulo: Martins Fontes, 2006.

ARISTÓTELES. **Ética a Nicômaco.** 5. Ed. São Paulo: Martin Claret, 2001.

ASTOBIZA, Aníbal Monasterio. Ética algorítmica: Implicaciones éticas de una sociedad cada vez más gobernada por algoritmos. **Dilemata**, n. 24, p. 185-217, 2017.

ÁVILA, Humberto. **Teoria dos Princípios**: da definição à aplicação dos princípios jurídicos. 17. Ed. rev. e atual. São Paulo: Malheiros, 2016.

BARAK, Aharon. **Proportionality.** Constitutional rights and their limitation. Cambridge: Cambridge University Press, 2012.

BAROCAS, Solon; HARDT, Moritz; NARAYANAN, Arvind. **Fairness and Machine Learning: Limitations and Opportunities**. Cambridge: MIT Press, 2019.

BIONI, Bruno Ricardo. **Proteção de dados pessoais**: A função e os limites do consentimento. São Paulo: Forense, 2019.

BONAT, Débora; PEIXOTO, Fabiano Hartmann. **Racionalidade no direito: inteligência artificial e precedentes.** Curitiba: Alteridade, 2020.

BONAVIDES, Paulo. **Curso de direito constitucional.** 15. ed. São Paulo: Malheiros, 2004.

BORGES, Caio Afonso. Regimes de visibilidade e poder das plataformas digitais nas relações de trabalho: uma análise a partir do trabalho imaterial de influenciadores digitais. Revista **Direito e Práxis**, [S. l.], v. 16, n. 2, p. 1–31, 2025. Disponível em: https://www.e-publicacoes.uerj.br/revistaceaju/article/view/85606. Acesso em: 21 jul. 2025.

BORGES, G. S.; FILÓ, M. da C. S. Inteligência artificial, gênero e direitos humanos: o caso Amazon. **Revista Justiça do Direito**, Passo Fundo, v. 35, n. 3, p. 218-243, 2021. Disponível em: https://seer.upf.br/index.php/rjd/article/view/12259. Acesso em: 28 set. 2024.

BOROWSKI, Martin. Discricionariedade na ponderação na Alemanha e na Convenção Europeia. *In*: TRIVISONNO, Alexandre Travessoni Gomes; BOROWSKI, Martin. **O debate sobre a teoria dos princípios fundamentais de Robert Alexy.** São Paulo: Marcial Pons, 2022a.

BOROWSKI, Martin. Os princípios formais e a fórmula do peso. *In*: TRIVISONNO, Alexandre Travessoni Gomes; BOROWSKI, Martin. **O debate sobre a teoria dos princípios fundamentais de Robert Alexy.** São Paulo: Marcial Pons, 2022b.

BRADBURY, Ray. **Farenheit 451 – a temperatura na qual o papel do livro pega fogo e queima.** Trad. Cid Knipel. São Paulo: Globo, 2009.

BRASIL. Supremo Tribunal Federal. **Ação Direta de Inconstitucionalidade 6387**. Relatora: Ministra Rosa Weber, 06 maio 2020a.

BRASIL. Supremo Tribunal Federal. Referendo na Medida Cautelar na **Ação Direta de Inconstitucionalidade 6.389**. Relatora: Ministra Rosa Weber, 07 maio 2020b.

BRUGGER, Winfried; LEAL, Mônia Clarissa Hennig. Os direitos fundamentais nas modernas constituições: análise comparativa entre as Constituições Alemã, Norte-americana e Brasileira. **Revista do Direito**, n. 28, p. 123-142, jul. 2007.

BUITEN, Miriam C. Towards intelligent regulation of artificial intelligence. **European Journal of Risk Regulation**, n. 10, p. 41–59, 2019.

BURRELL, Jenna. How the machine 'thinks': Understanding opacity in machine learning algorithms. **Big Data & Society**, v. 1, n. 12, 2016. Disponível em: <https://journals.sagepub.com/doi/epub/10.1177/2053951715622512>. Acesso em: 08 ago. 2025.

CAMBI, Eduardo Augusto Salomão; AMARAL, Maria Eduarda Toledo Pennacchi Tibiriçá. Inteligência artificial no Poder Judiciário, discriminação algorítmica e direitos humanos-fundamentais. **Suprema-Revista de Estudos Constitucionais**, v. 3, n. 2, p. 189-218, 2023.

CANTARINI, Paola. Entropia dos algoritmos: é possível se falar em equidade algorítmica? In: AVANCI, Thiago Felipe. **O futuro do direito**: o que esperar do direito e das tecnologias. Londrina: Thoth, 2022.

CARBONELL, Miguel. Constitucionalismo, minorías y derecho. **Isonomía Revista de Teoría y Filosofía del Derecho**, n. 12, v. 98, 2000.

CARUGATI, Christophe. The 2017 Facebook Saga: A Competition, Consumer and Data Protection Story. **European Competition and Regulatory L. Review**, v. 2, n; 1, p. 4-10, 2018.

CASTELLS, Manuel. **A Era da Informação:** economia, sociedade e cultura, v. 1: A sociedade em Rede, 4. ed., Lisboa: Fundação Calouste Gulbenkian, 2011.

CELESTE, Edoardo. Constitucionalismo digital: mapeando a resposta constitucional aos desafios da tecnologia digital. **Direitos Fundamentais & Justiça**, a. 15, n. 45, p. 63-91, jul./dez., 2021

CELESTE, Edoardo. Digital constitutionalism: a new systematic theorisation. In: Taylor & Francis. **International Review of Law, Computers & Technology** 33:1, p. 76-99, jan. 2019.

CELESTE, Edoardo. Constituições digitais: origem, modelos e papel de um movimento inovador. In: GAVIÃO FILHO, Anizio Pires; PAULO, Lucas Moreschi. (Org.). **Os direitos fundamentais no constitucionalismo digital**. São Paulo: Dialética, 2024, p. 13-44.

CHALA, Bárbara Guerra; PAULO, Lucas Moreschi. Limites da liberdade de expressão à luz da proporcionalidade: o Inquérito das fake news. In: GAVIÃO FILHO, Anizio Pires; PAULO, Lucas Moreschi. (Org.). **Constitucionalismo, direitos fundamentais, proporcionalidade e argumentação.** São Paulo: Dialética, 2022, p. 129-149.

CHONDROS, Thomas G. **Natural philosophy and the development of mechanics and engineering from the 5th century B.C. to Middle-Ages**. FME Transactions, Belgrade, v. 45, n. 4, p. 603-619, 2017.

CLARKE, Roger. Profiling: A hidden challenge to the regulation of data surveillance. **Journal of Law & Information Science**, v. 4, p. 403-419, 1993.

CODED BIAS; Direção: Shalini Kantayya. Produção: Shalini Kantayya. Reino Unido: Netflix, 2020.

COLLINS, Hugh; KHAITAN, Tarunabh. **Foundations of Indirect Discrimination Law**. Oxford: Hart, 2017.

COLLINS, Patrícia Hill; BILGE, Sirma. **Interseccionalidade**. Trad. Rane Souza. São Paulo: Boitempo, 2021.

COHEN-ELIYA, Moshe; PORAT, Iddo. The Administrative origins of constitutional rights and global constitutionalism. In: JACKSON; TUSHNET (ed.). **Proportionality.** New frontiers, new challenges. Cambridge: Cambridge University Press, 2017.

COLANER, Nathan. Is explainable artificial intelligence intrinsically valuable?. **AI & Society**, n. 37, p. 231–238, 2022.

CORTE IDH. Corte Interamericana De Direitos Humanos. Caso **Fazenda Brasil Verde vs. Brasil.** Sentença de Exceções Preliminares, Méritos, Reparações e Custas de 20 de outubro de 2016. San José, Costa Rica, 2016.

CORTE IDH. Corte Interamericana De Direitos Humanos. **Caso empregados da fábrica de fogos de Santo Antônio de Jesus e seus familiares vs. Brasil**. Sentença de mérito, reparações e custas de 15 de julho de 2020. San José, Costa Rica, 2020.

COZMAN, Fábio G.; NERI, Hugo. O que, afinal, é Inteligência Artificial? In: COZMAN, Fabio G.; PLONSKI, Guilherme Ary; NERI, Hugo (org.). **Inteligência artificial: avanços e tendências**. São Paulo: Instituto de Estudos Avançados, 2021.

CRESTANE, Dérique S. **Discriminação algorítmica: a aplicabilidade dos standards protetivos fixados pela Corte Interamericana De Direitos Humanos e pelo Supremo Tribunal Federal em relação ao direito de igualdade e não discriminação a partir das noções de Ius Constitutionale Commune Latino-Americano e dever de proteção estatal**. Dissertação (Mestrado em Direito) – Faculdade de Direito, Universidade de Santa Cruz do Sul, p. 180, 2022.

CRESTANE, Dérique Soares; LEAL, Mônia Clarissa Hennig. **Discriminação algorítmica e Discriminação estrutural.** São Paulo: Tirant lo Blanch, 2024.

CRENSHAW, Kimberlé. Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics. **University of Chicago Legal Forum**, Chicago, v. 1989, n. 1, p. 139-167, 1989.

CRENSHAW, Kimberlé. Documento para o Encontro de Especialistas em Aspectos da Discriminação Racial Relativos ao Gênero. **Revista Estudos Feministas**. a. 10, n. 1, p. 171-188. Florianópolis, 2002.

CREUZ, Derek Assenço. Por que descolonizar o Direito da Antidiscriminação? : Uma crítica à racionalidade liberal-colonial do direito à igualdade. **Revista Direito e Práxis**, *[S. l.]*, v. 16, n. 1, p. 1–27, 2024. Disponível em: https://www.e-publicacoes.uerj.br/revistaceaju/article/view/83410. Acesso em: 23 jul. 2025.

CRIADO-PEREZ, Caroline. **Invisible Women: Data Bias in a World Designed for Men**. Londres: Penguim, 2020

DAURA, Samir Alves. Behavioral Economics and Consumer Law: New Perspectives to Confront the Overindebtedness. **RBPP**, v. 8, n.2, p. 568-599, ago., 2018.

DE GREGORIO, Giovanni. The rise of digital constitutionalism in the European Union. **International Journal of Constitutional Law**, v. 19, n. 1, p. 41-70, 2021.

DEBRUSK, Chris. **The risk of machine-learning bias (and how to prevent it)**. *MIT Sloan Blogs*, 2018.

DONEDA, Danilo Cesar Maganhoto et al. Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. **Pensar-Revista de Ciências Jurídicas**, Fortaleza, v. 23, n. 4, p. 1-17, 2018. Disponível em: https://ojs.unifor.br/rpen/article/view/8257. Acesso em: 28 set. 2024.

DONEDA, Danilo. **Da Privacidade à Proteção de Dados Pessoais.** 2. ed. São Paulo: Revista dos Tribunais, 2019.

DONEDA, Danilo. Panorama Histórico da Proteção de Dados Pessoais. In MENDES, Laura Schertel; DONEDA, Danilo; SARLET, Ingo Wolfgang; e RODRIGUES JÚNIOR, Otavio. **Tratado de Proteção de Dados Pessoais**. Rio de Janeiro: Forense, 2021.

DONEDA, Danilo. Registro da sustentação oral no julgamento da ADI 6389, sobre a inconstitucionalidade do art. 2º, *caput* e §§1º e 3º da MP 954/2020. **Civilistica.com**. Rio de Janeiro, v. 9, n. 1, p. 1-9, 2020

DUARTE, A.; NEGÓCIO, R. de V. Todos são iguais perante o algoritmo? Uma resposta cultural do direito à discriminação algorítmica. **Direito Público**, Brasília, v. 18, n. 100, 2022. DOI: 10.11117/rdp.v18i100.5869. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/5869. Acesso em: 30 set. 2024.

EHRHARDT JR, Marcos; SILVA, Gabriela Buarque Pereira. Contratos e Algoritmos: Alocação de Riscos Discriminação e Necessidade de Supervisão por Humanos. **JURISMAT**, n. 13, p. 33-33, 2021.

ELIOT, Thomas Stearns, **Choruses from The Rock**. Nova Iorque: Harcourt, Brace and Company, 1934.

ERNST, Christian. Artificial Intelligence and Autonomy: Self-Determination in the Age of Automated Systems. In: WISCHMEYER, Thomas; RADEMACHER, Timo (editors). **Regulating Artificial Intelligence.** Cham: Springer, 2020, p. 53-74.

EUBANKS, Virginia. **Automating inequality**: how high-tech tools profile, police, and punish the poor. Nova Iorque: St. Martin's Press, 2017.

EYSENCK, Michael W.; EYSENCK, Christine. **Inteligência artificial x humanos: o que a ciência cognitiva nos ensina ao colocar frente a frente a mente humana e a IA.** Trad: Gisele Klein. Porto Alegre: Artmed, 2023.

FERRARI, Isabela. **Justiça Digital**. 2. ed. São Paulo: RT, 2021.

FERRARI, Isabela; BECKER, Daniel; WOLKART, Erik Navarro. Arbitrium ex machina: panorama, riscos e a necessidade de regulação das decisões informadas por algoritmos. **Revista dos Tribunais**, v. 995, p. 635-655, set./2018.

FINEMAN, Martha Albertson. O sujeito vulnerável: ancorando a igualdade na condição humana. Revista **Direito e Práxis,** [S. l.], v. 14, n. 2, p. 1371–1401, 2023. Disponível em: https://www.e-publicacoes.uerj.br/revistaceaju/article/view/66023. Acesso em: 21 jul. 2025.

FISCHER, Emma; VIOLLIER, Pablo. La intervención humana como resguardo antes la toma automatizada de decisiones: Implicancias Éticas y Jurídicas. in: PIRELA, Michelle Azuaje. **Introducción a la Ética y el Derechos de la Inteligencia Artificial**. Madrid: La Ley, 2023,

FISS, Owen M. Groups and the Equal Protection Clause. **Philosophy and Public Affairs**, v. 5, n. 2, 1976.

FOUCAULT, Michel. **Vigiar e punir: nascimento das prisões**. Petrópolis: Vozes, 2014.

FRANÇA, Eduarda Peixoto da Cunha; NÓBREGA, Flavianne Fernanda Bitencourt. A discriminação estrutural e interseccional no Brasil: uma análise do Caso "Empregados da Fábrica de Fogos de Santo Antônio de Jesus e seus Familiares vs. Brasil" . **Revista Direito e Práxis**, [S. l.], v. 16, n. 1, p. 1–27, 2024. Disponível em: https://www.e-publicacoes.uerj.br/revistaceaju/article/view/80270. Acesso em: 21 jul. 2025.

FRANÇA. **Décret n° 2017-330 du 14 mars 2017 relatif aux droits des personnes faisant l'objet de décisions individuelles prises sur le fondement d'un traitement algorithmique**. Légiferance. République Française, 2017. Disponível em:

https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000034194929. Acesso em: 19 out. 2023.

FRANÇA. **LOI n° 2016-1321 du 7 octobre 2016 pour une République numérique (1)** Légiferance. République Française, 2016. Disponível em: https://www.legifrance.gouv.fr/loda/id/JORFTEXT000033202746?init=true&page=1&query=2016-1321&searchField=ALL&tab_selection=all. Acesso em: 19 out. 2023.

FRASER, Nancy**. Reconhecimento sem ética?** Trad. Ana Carolina Freitas e Mariana Prandini. São Paulo: Lua nova, 2007.

FRAZÃO, Ana. Quais devem ser os parâmetros éticos e jurídicos para a utilização da inteligência artificial?. **Revista eletrônica [do] Tribunal Regional do Trabalho da 9ª Região**, Curitiba, v. 9, n. 86, p. 119-124, mar. 2020.

FREDMAN, Sandra. **Direct and Indirect Discrimination: Is There Still a Divide?** In: COLLINS, Hugh; KHAITAN, Tarunabh. **Foundations of Indirect Discrimination Law**. Oxford: Hart, p. 31-56, 2017.

FRICKER, Miranda. **Epistemic Injustice: Power and the Ethics of Knowing**. New York: Oxford University Press, 2007.

FULLER, Lon L. **The Morality of Law**. New Haven: Yale University Press, 1963.

GAVIÃO FILHO, Anizio Pires. **Colisão de Direitos Fundamentais, Argumentação e Ponderação.** Porto Alegre: Livraria do Advogado, 2011.

GAVIÃO FILHO, Anizio Pires. Elementos do Constitucionalismo Digital. In: GAVIÃO Filho, Anizio Pires; MOTTA, Francisco José Borges; PAULO, Lucas Moreschi (Org.). **O Constitucionalismo Digital e a Crise das Democracias Liberais**. São Paulo: Editora Dialética, 2023.

GAVIÃO FILHO, Anizio Pires; FREITAS, Luiz Fernando Calil de. Direitos fundamentais estatuídos não diretamente ou implícitos. **Revista Direitos Fundamentais & Democracia**, Curitiba, v. 25, n. 3, p. 232-257, set./out. 2020.

GAVIÃO FILHO, Anizio Pires. **Discriminação Algorítmica Indireta e Proporcionalidade**. In: GAVIÃO FILHO, Anizio Pires; PAULO, Lucas Moreschi (Org.). Constitucionalismo Contemporâneo e Direitos Fundamentais. São Paulo: Tirant lo Blanch, 2025.

GRIMM, Dieter. **Constitutionalism**: past, present and future. Oxford: Oxford University Press, 2016.

HAN, Byung-Chul. **No enxame**: **perspectivas do digital**. Trad. Lucas Machado. Petrópolis: Vozes, 2018.

HAN, Byung-Chul. **Sociedade Paliativa: a dor hoje**. Trad. Lucas Machado. Petrópolis: Vozes, 2021.

HAO, Karen. AI's white guy problem isn't going away. **MIT Technology Review.** Artificial Intelligence. abr. 2019. Disponível em: < https://www.technologyreview.com/2019/04/17/136072/ais-white-guy-problem-isnt-going-away/>. Acesso em 13 ago. 2025.

HARARI, Yuval Noah. **Homo Deus: uma breve história do amanhã**. Trad. Paulo Geiger. São Paulo: Companhia das Letras, 2016.

HEIKKILÄ, Melissa. The viral AI avatar app Lensa undressed me–without my consent. **MIT Technology Review.** Artificial Intelligence. dez. 2022. Disponível em: <https://www.technologyreview.com/2022/12/12/1064751/the-viral-ai-avatar-app-lensa-undressed-me-without-my-consent/>. Acesso em 23 mar. 2023.

HEIKKILÄ, Melissa. These new tools let you see for yourself how biased AI image models are. **MIT Technology Review**. Artificial Intelligence. mar. 2023a. Disponível em: <https://www.technologyreview.com/2023/03/22/1070167/these-news-tool-let-you-see-for-yourself-how-biased-ai-image-models-are/>. Acesso em 23 mar. 2024.

HEIKKILÄ, Melissa. Why it'll be hard to tell if AI ever becomes conscious. **MIT Technology Review.** Artificial Intelligence. out. 2023b. Disponível em: < https://www.technologyreview.com/2023/10/17/1081818/why-itll-be-hard-to-tell-if-ai-ever-becomes-conscious/>. Acesso em 29 mar. 2024.

HOBBES, Thomas. **Do Cidadão**. São Paulo: Martins Fontes, 2002.

HOBBES, Thomas. **Leviatã**. São Paulo: Martins Fontes, 2008.

HOFFMANN-RIEM, Wolfgang. **Teoria geral do direito digital***:* transformação digital: desafios para o direito. Tradução: Ítalo Fuhrmann. 2ª. ed. Rio de Janeiro: Forense, 2022.

HOLLANEK, Tomasz. AI transparency: a matter of reconciling design with critique. **AI & Society**, n. 38, p. 2071–2079, 2023.

HOMERO. **Odisseia**. São Paulo: Cultrix, 2006.

HOMERO. **Ilíada**. Londres: Penguim, 2013.

HOSANAGAR, Kartik. **A Human's guide to machine intelligence**: How algorithmics are shaping our lives and how we can stay in control. New York, Viking, 2019.

HUCKINS, Grace. Minds of machines: The great AI consciousness conundrum. **MIT Technology Review**. Artificial Intelligence. out. 2023. Disponível em: <https://www.technologyreview.com/2023/10/16/1081149/ai-consciousness-conundrum/>. Acesso em 19 abr. 2024.

HUGO, Victor. **Os Miseráveis**. Dois Irmãos: Clube de Literatura Clássica, 2025.

IBGE. Instituto Brasileiro de Geografia e Estatística. **Síntese de indicadores sociais: uma análise das condições de vida da população brasileira: 2023**. Disponível em:

https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2102052. Acesso em: 28 set. 2024.

ITÁLIA. **Provvedimento nº 112 del 30 marzo 2023 [9870832]**. Garante Per La Protezione Dei Dati Personali. Repubblica **Italiana**, 2023. Disponível em: <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9870832>. Acesso em: 19 out. 2023.

JACOBS, Francis G. Judicial Dialogue and the Cross-Fertilization of Legal Systems: The European Court of Justice. **Texas International Law Journal**, v. 38, 2003.

JARDIM, Augusto Tanger. Reflexões sobre inteligência artificial e jurisdição: Potenciais problemas resultantes do uso da ia à luz dos elementos definidores da jurisdição. In: GAVIÃO FILHO, Anizio Pires; PAULO, Lucas Moreschi (Org.). **Direitos Fundamentais e Constitucionalismo Digital**. São Paulo: Dialética, 2024

JOHNSTON, David. **Breve história da justiça.** São Paulo: Martins Fontes, 2018.

KANT, Immanuel. **Crítica da razão prática**. Rio de Janeiro: Nova Fronteira, 2017.

KANT, Immanuel. **Fundamentação da metafísica dos costumes**. Tradução de Edson Bini. São Paulo: Martin Claret, 2009.

KLATT, Matthias. Ponderando direitos e interesses: objeções à tese da assimetria de Frederick Schauer. **Revista Brasileira de Direito**, Passo Fundo, v. 17, n. 1, p. 1-28, mai. 2021. Disponível em: https://seer.atitus.edu.br/index.php/revistadedireito/article/view/4495/2942. Acesso em: 15 abr. 2025.

KLATT, Matthias; MEISTER, Moritz. **The constitutional structure of proportionality**. Oxford: Oxford University Press, 2012.

KLEINBERG, Jon; LUDWIG, Jens; MULLAINATHAN, Sendhil; SUNSTEIN, Cass R. **Discrimination in the age of algorithms**. Yale Law Journal, v. 118, n. 8, p. 1400-1472, 2019

KONDER, Carlos Nelson; KONDER, Cíntia Muniz de Souza. O conceito jurídico de hipervulnerabilidade é necessário para o direito?. In: TEIXEIRA, Ana Carolina Brochado; MENEZES, Joyceane Bezerra. **Gênero, Autonomia e Vulnerabilidade: repercussões jurídicas.** Indaiatuba, SP: Editora Foco Jurídico, 2020.

LEAL, Mônia Clarissa Hennig. **A Constituição como Princípio: os limites da jurisdição constitucional brasileira**. Barueri: Manole, 2003.

LEAL, Mônia Clarissa Hennig; ALVES, Fernando Roberto Schnorr. O Supremo Tribunal Federal e as sentenças estruturantes: análise crítica da ADPF 709 (enfrentamento da pandemia entre os povos indígenas enquanto grupo vulnerável). **Revista Culturas Jurídicas**, Ahead of Print, 2023.

LEAL, Mônia Clarissa Hennig; CRESTANE, Dérique Soares. Algorithmic

discrimination as a form of structural discrimination: Standards of the Inter-American Court of Human Rights related to vulnerable groups and the challenges to judicial review related to structural injunctions. **UNIO – EU Law Journal**, v. 9, n. 1, p. 29–44, 2023.

LEAL, Mônia Clarissa Hennig; PAULO, Lucas Moreschi. Algoritmos discriminatórios e jurisdição constitucional: os riscos jurídicos e sociais do impacto dos vieses nas plataformas de inteligência artificial de amplo acesso. **Revista de Direitos e Garantias Fundamentais**, v. 24, n. 3, p. 165–187, 2023.

LEAL, Mônia Clarissa Hennig; PAULO, Lucas Moreschi. Moradia vs. Propriedade: a ADPF 828/DF e a proteção contramajoritária em tempos de pandemia à luz da proporcionalidade. **Revista da Faculdade de Direito do Sul de Minas**, v. 40, n. 1, 2024.

LEAL; Mônia Clarissa Hennig; MAAS, Rosana Helena. **"Dever de proteção estatal", "proibição de proteção insuficiente" e controle jurisdicional de Políticas Públicas**. Rio de Janeiro: Lúmen Juris, 2020.

LEAL; Mônia Clarissa Hennig; MAAS, Rosana Helena. "Dever de proteção estatal", "proibição de proteção insuficiente" e "proibição de ex cesso": espectro de sua conformação e desenvolvimento pela teoria constitucional alemã. **Revista Brasileira de Estudos Políticos**, n. 125, p. 397-438, jul./dez. 2022.

LIMBERGER, Têmis. Da evolução do direito a ser deixado em paz à proteção dos dados pessoais. **Revista do Direito**, Santa Cruz do Sul, p. 138-160, jul. 2008.

LONG, Clarissa. Privacy and Pandemics. In PISTOR, Katharina. **Law in the time of COVID-19**. Columbia Law School Books, 2020.

LOPES, José Reinaldo, **As palavras e a lei**: Direito, ordem e justiça na história do pensamento jurídico moderno. 34 ed. São Paulo: FGV Editora, 2004.

LÓPEZ, José Miguel Hernández. **Reglamento de inteligência artificial anotado.** Barcelona: Bosch Editor, 2024.

LORDELO, João Paulo. Vieses implícitos e técnicas de automação decisória: riscos e benefícios. **Revista da ANNEP de Direito Processual**, v. 1, n. 2, art. 44, 2020.

LUCCIONI, Alexandra Sasha; AKIKI, Christopher; MITCHELL, Margaret; JERNITE, Yacine. Stable bias: analyzing societal representations in diffusion models. **Arxiv. Computer Science**. Cornell University, mar. 2023.

LUNA SALAS, Fernando; PERONA, Riccardo; CARRILLO DE LA ROSA, Yezid. Impacto y límites de la inteligencia artificial en la práctica jurídica. **Via Inveniendi Et Iudicandi**, *[S. l.]*, v. 17, n. 2, 2023. Disponível em: https://revistas.usantotomas.edu.co/index.php/viei/article/view/8773.. Acesso em: 3 dec. 2024.

MACHADO DE ASSIS, Joaquim Maria. **Teoria do Medalhão.** Rio de Janeiro: Pop Stories, 2022.

MACIEL, Rafael. **Manual prático sobre a Lei Geral de Proteção de Dados Pessoais:** Atualizado com a Medida Provisória nº 869/18. Goiânia: RM Digital, 2019.

MAGGIOLINO, Mariateresa. EU trade secrets law and algorithmic transparency. **Bocconi Legal Studies**, Research Paper n. 3363178, 2019.

MANJIKIAN, Mary. **Cybersecurity ethics: an introduction**. 2. ed. Londres e Nova Iorque: Routledge, 2023.

MARQUES, Claudia Lima; MIRAGEM, Bruno. **O novo direito privado e proteção dos vulneráveis**. São Paulo: Revista dos Tribunais, 2012.

MARTINS, Fernando Rodrigues. Sociedade da informação e promoção à pessoa. **Revista de Direito do Consumidor**, São Paulo, Revista dos Tribunais, v. 96, p. 225-257, nov./dez. 2014.

MASCARO, Alysson Leandro. **Filosofia do Direito**. 5 ed., rev., atual., e ampl. São Paulo: Atlas, 2016.

MELLO, Celso Antônio Bandeira De. **Conteúdo Jurídico do Princípio da Igualdade**. 3. ed. São Paulo: Malheiros, 1993.

MENDES, Laura S.; BIONI, Bruno R. O Regulamento Europeu de Proteção de Dados Pessoais e a Lei Geral de Proteção de Dados Brasileira: Mapeando Convergências na Direção de um Nível de Equivalência. **Revista de Direito do Consumidor**. São Paulo, Revista dos Tribunais, v. 124, p. 157-180, jul./ago. 2019.

MENDES, Laura Schertel. O Direito Fundamental à proteção de dados pessoais. **Revista de Direito do Consumidor**. São Paulo, Revista dos Tribunais, v. 79, p. 45-81, jul./set. 2011.

MENDES, Laura Schertel. **Privacidade, proteção de dados e defesa do consumidor:** linhas gerais de um novo direito fundamental. São Paulo: Saraiva, 2014.

MENDES, Laura Schertel; DONEDA, Danilo. Reflexões iniciais sobre a nova lei geral de proteção de dados. **RDC**, São Paulo, v. 120, p. 469-483, nov./dez., 2018.

MENDES, Laura Schertel; FONSECA, Gabriel Campos Soares da. STF reconhece direito fundamental à proteção de dados: comentários sobre o referendo da Medida Cautelar nas ADIs 6387, 6388, 6389, 6390 e 6393. **Revista de Direito do Consumidor**. São Paulo, Revista dos Tribunais, v. 130, p. 471-478, jul./ago., 2020.

MENDES, Laura Schertel; MATTIUZZO, Marcela. Discriminação Algorítmica: Conceito, Fundamento Legal E Tipologia. **Direito Público**, v. 16, n. 90, 2019.

MILL, John Stuart**. Utilitarismo.** Tradução de Pedro Galvão. Lisboa: Edições 70, 2005.

MIRAGEM, Bruno. A Lei Geral de Proteção de Dados (Lei 13.709/2018) e o direito do consumidor. **Revista dos Tribunais**, São Paulo, v. 1009, n.p., nov. 2019.

MÖLLER, Kai. **The global model of constitutional rights.** Oxford: Oxford University Press, 2012.

MOREIRA, Adilson José. **Racismo Recreativo.** São Paulo: Pólen, 2019.

MOREIRA, Adilson José; MARTINELLI, Gabriela Doll; HELENA DE ARAÚJO BENTO, Helena de Araújo Bento; PALMIERI, Rafaella Pavanello. Discriminação Estética / Aesthetic Discrimination. **Revista Direito e Práxis**, *[S. l.]*, v. 14, n. 3, p. 1934–1959, 2023. Disponível em: https://www.e-publicacoes.uerj.br/revistaceaju/article/view/78114. Acesso em: 23 jul. 2025.

MOROZOV, Evgeny. **Big tech: a ascensão dos dados e a morte da política**. Trad. Claúdio Marcondes. São Paulo: Ubu, 2018.

NEDEL, Antônio Carlos. **Uma tópica jurídica:** clareira para a emergência do Direito. Porto Alegre: Livraria do Advogado, 2006.

NICOLELIS, Miguel. **O verdadeiro criador de tudo: como o cérebro humano esculpiu o universo como nós o conhecemos.** São Paulo: Planeta, 2020.

O'DONNELL, James. Meet the early-adopter judges using AI. **MIT Technology Review**, 11 ago. 2025.

O'NEIL, Cathy. **Algoritmos de destruição em massa**: como o big data aumenta a desigualdade e ameaça a democracia. Trad. Rafael Abraham. Santo André: Editora Rua do Sabão, 2020.

ORWELL, George. **1984**. Trad. Alexandre Hubner e Heloisa Jahn. 38. ed. São Paulo: Companhia das Letras, 2009 [1949].

PARANHOS, Mário C. Oliveira. **Viés algorítmico: uma análise sobre discriminações automatizadas**. Rio de Janeiro: Lúmens Juris, 2022.

PASQUALE, Frank. **The black box society: the secret algorithms that control money and information**. Cambridge e Londres: Harvard University Press, 2015.

PAULO, Lucas Moreschi. Discriminação algorítmica, grupos vulneráveis e a lei material da ponderação: um enfoque jurídico-constitucional para as sobreposições de direitos fundamentais na era digital. In: GAVIÃO FILHO, Anizio Pires; PAULO, Lucas Moreschi. (Orgs.). **Os Direitos Fundamentais no Constitucionalismo Digital.** São Paulo: Dialética, 2024.

PAULO, Lucas Moreschi. **Fundamentos para uma teoria da coerência normativa no raciocínio jurídico.** São Paulo: Dialética, 2022.

PAULO, Lucas Moreschi. Opacidade dos algoritmos e a necessidade de transparência: Garantindo explicabilidade. In: **Anais do XIX Seminário Internacional Demandas Sociais e Políticas Públicas na Sociedade Contemporânea e XV Mostra Internacional de Trabalhos Científicos**. Santa Cruz do Sul: UNISC, 2023.

PEIXOTO, Fábio Hartmann. **Inteligência Artificial e Direito: Convergência ética e estratégica**. Coleção Direito, Racionalidade e Inteligência Artificial, v. 5. Curitiba: Alteridade, 2020.

PÉREZ, Hugo Neves. Inteligência artificial y derechos humanos: ¿oportunidad o amenaza?. In: JUEZ, Marra Méndez (Coord.). **Derechos humanos e inteligência artificial: amenazas y desafíos.** Madrid: Aranzadi, p. 125-139, 2024.

PINHEIRO, Patricia Peck. **Proteção de dados pessoais: comentários à Lei n.13.709/2018 (LGPD)**. São Paulo: Saraiva, 2018.

POLONSKI, Vyacheslav. **Mitigating algorithmic bias in predictive justice.** 4 AI design principles for AI fairness, 2018.

POSTEMA, Gerald. **Law's Rule: the nature, value, and viability of the rule of law**. Oxford: Oxford University Press, 2022.

PREVEDELLO, Alexandre. **Teoria da Derrotabilidade. As exceções não previstas nas regras jurídicas.** Curitiba: Juruá, 2019.

RAWLS, John. **A Theory of Justice**. Harvard University Press, 1971.

RAWLS, John. **Justiça como equidade: uma reformulação**. Tradução de Álvaro de Vita. São Paulo: Martins Fontes, 2003.

RAWLS, John. Social Unity and primary goods. In: SEN, Amartya; WILLIAMS, Bernard (Orgs.). **Utilitarianism and beyond**. Cambridge: Cambridge University Press, p. 159-185, 1982.

REIS, Paulo Victor Alfeo. **Algoritmos e o Direito.** São Paulo: Almedina, 2020.

REQUIÃO, Maurício; COSTA, Diego. Discriminação algorítmica: ações afirmativas como estratégia de combate. **Civilistica.com**, Rio de Janeiro, v. 11, n. 3, p. 1–24, 2022. Disponível em: https://civilistica.emnuvens.com.br/redc/article/view/804. Acesso em: 10 out. 2024.

RICHARDSON, Anamaria. **Biased data lead to biased algorithms**. CMAJ, v. 194, n. 9, p. E341-E341, 2022.

RODOTÀ, Stefano. **A vida na sociedade de vigilância: privacidade hoje**. Rio de Janeiro: Renovar, 2008.

ROUSSEAU, Jean-Jacques. **O Contrato Social ou Princípios do Direito Político.** Trad. Edson Bini. São Paulo: Edipro, 2017.

RUSSELL, Stuart. **Inteligência Artificial a nosso favor: como manter o controle sobre a tecnologia**. Trad. Berilo Vargas. São Paulo: Companhia das Letras, 2021.

SABA, Roberto. (Des)igualdad estructural. **Derecho y Humanidades**, n. 11, 2005.

SAGÜÉS, María Sofía. Discriminación estructural, inclusión y litigio estratégico. In: FERRER MAC-GREGOR, Eduardo; MORALES ANTONIAZZI, Mariela; FLORES PANTOJA, Rogelio. **Inclusión, Ius Commune y justiciabilidad de los DESCA en la jurisprudencia interamericana. El caso Lagos del Campo y los nuevos desafíos. Colección Constitución y Derechos**. Querétaro México: Instituto de Estudios Constitucionales del Estado de Querétaro, 2018. p. 129-178.

SANTOMIL, Pablo, et al. Análisis de Los riesgos empresariales derivados de la inteligência artificial. In: NOVOA, César García; RODRÍGUEZ, Francisco Hernández (Dirs.) **Estúdios sobre inteligência artificial y economía digital**. Madrid: Aranzadi, p. 71-94, 2024.

SANTOS, Ana Luiza L. Lei Geral de Proteção de Dados: os caminhos percorridos até sua vigência no contexto da sociedade da informação. In VASCONCELOS, Adaylson W. Souza de (Org.). **A (não)efetividade das ciências jurídicas no Brasil 3.** Ponta Grossa: Atena, 2021.

SANTOS, Ana Luiza L. Lei Geral de Proteção de Dados: um estudo comparativo em relação à efetividade dos direitos fundamentais. **Revista dos Tribunais**, São Paulo, v. 1013, p. 105-126, mar., 2020.

SARLET, Ingo Wolfgang. Dos direitos e garantias fundamentais. In: CANOTILHO, J. J. Gomes; MENDES, Gilmar F.; SARLET, Ingo W.; STRECK, Lenio. L. (Coords.). **Comentários à constituição do Brasil.** p. 183-212. São Paulo: Saraiva/Almedina, 2013.

SARLET, Ingo Wolfgang; MARINONI, Luiz Guilherme; MITIDIERO, Daniel. **Curso de Direito Constitucional**. 9. ed. São Paulo: Saraiva Educação, 2020.

SARTORI, Ellen Carina Mattias; BAHIA, Cláudio José Amaral. Big Brother is watching you: da distopia orwelliana ao direito fundamental à proteção de dados pessoais. **Revista de Direitos e Garantias Fundamentais**, v. 20, n. 3, p. 225-248, 2019.

SCHERMER, Bart Willem; CUSTERS, Bart; VAN DER HOF, Simone. The Crisis of Consent: How Stronger Legal Protection May Lead to Weaker Consent in Data Protection. **Ethics and Information Technology, The Crisis of Consent: How Stronger Legal Protection may lead to Weaker Consent in Data Protection,** 2014.

SCHWAB, Klaus. **A quarta revolução industrial**. Trad. Daniel Moreira Miranda. São Paulo: Edipro, 2016.

SCHWABE, Jürgen. **Cinquenta anos de jurisprudência do tribunal constitucional federal alemão**. Trad. B. Henning e L. Martins. Montevideo: Fundación Konrad-Adenauer, 2005.

SHELLEY, Mary. **Frankenstein, ou, O Prometeu Moderno.** Dois Irmãos: Clube de Literatura Clássica, 2022.

SHIRER, William L. **Ascensão e queda do Terceiro Reich**: Triunfo e consolidação (1933-1939). 2. ed. Rio de Janeiro: Nova Fronteira, 2017.

SIECKMANN, Jan-R. Los derechos fundamentales como princípios. 2011. *In*: SIECKMANN, Jan-R. **La teoría principialista de los derechos fundamentales**. Estudios sobre la teoría de los derechos fundamentales de Robert Alexy. Madrid: Marcial Pons, 2011.

SILVA, Tarcízio. **Racismo algorítmico: inteligência artificial e discriminação nas redes digitais**. Edições Sesc SP, 2022.

SILVERIO, Valter Roberto. Ação afirmativa e o combate ao racismo institucional no Brasil. **Cadernos de pesquisa**, São Carlos, p. 219-246, 2002. Disponível em: https://www.scielo.br/j/cp/a/RkKqjbycXDYS93kh8bNdLLs/#. Acesso em: 22 set. 2024.

SOARES SILVA, I. M.; MENDES BARBOSA, L. Inov(ação): : discriminação algorítmica racial e as inteligências artificiais no Brasil. **Revista do CAAP**, *[S. l.]*, v. 28, n. 2, p. 1–30, 2024. DOI: 10.69881/rcaap.v28i2.49200. Disponível em: https://periodicos.ufmg.br/index.php/caap/article/view/49200. Acesso em: 10 out. 2024.

SOARES, Carolina Borges. Direito à autodeterminação informativa e proteção de dados em tempos de pandemia: análise do atual contexto jurídico brasileiro. In: FERRAZ, Miriam Olivia Knopik; VETTORAZI, Karlo Messa (org.). **Direitos fundamentais e a era tecnológica - Law Experience**. Curitiba: FAE/Bom Jesus, 2020.

STEINER, Christopher. **Automate This: how algorithms came to rule our world**. New York: Penguin Group, 2012.

STRECK, Lenio Luiz. **O que é isto – decido conforme minha consciência?** 4. ed. Porto Alegre: Livraria do Advogado, 2013.

TEFFÉ, Chiara Spadaccini de; VIOLA, Mario. Tratamento de dados pessoais na LGPD: estudo sobre as bases legais. **Civilistica.com**. Rio de Janeiro, a. 9, n. 1, 2

TEUBNER, Günther. **Fragmentos constitucionais:** constitucionalismo social na globalização. Trad. Ricardo Campos. 2. ed. São Paulo: Saraiva, 2020.

THE ECONOMIST. The world's most valuable resource is no longer oil, but data. The Economist, mai. 2017. Disponível em: < https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data>. Acesso em: 26 de out. 2025.

TISCHBIREK, Alexander. Artificial Intelligence and Discrimination: Discriminating Against Discriminatory Systems. In: WISCHMEYER, Thomas; RADEMACHER, Timo (editors). **Regulating Artificial Intelligence**. Cham: Springer, 2020.

TOVAR, Alejandro Nava. A natureza do exame da proporcionalidade e os princípios formais. 2022. *In*: TRIVISONNO, Alexandre Travessoni Gomes; BOROWSKI, Martin. **O debate sobre a teoria dos princípios fundamentais de Robert Alexy.** São Paulo: Marcial Pons, 2022.

TREVISAN, Leonardo Simchen. **Ponderação, Argumentação, Racionalidade:** Robert Alexy e seus críticos. Porto Alegre: Sergio Antonio Fabris, 2017.

TRIVISONNO, Alexandre Travessoni Gomes; BOROWSKI, Martin. Prefácio dos Organizadores. 2022. *In*: TRIVISONNO, Alexandre Travessoni Gomes; BOROWSKI, Martin. **O debate sobre a teoria dos princípios fundamentais de Robert Alexy.** São Paulo: Marcial Pons, 2022.

TURING, Alan Mathison. **Intelligent machinery**. London: National Physical Laboratory, 1948.

TURK, Victoria. How AI reduces the world to stereotypes. **Rest of World Journal.** The rise of AI. Out. 2023. Disponível em: < https://restofworld.org/2023/ai-image-stereotypes/>. Acesso em 29 mar. 2023.

ULPIANO. Digesto de Justiniano. Corpus Juris Civilis. in: GARCIA DEL CORRAL, Ildefonso L. (org.). **Cuerpo del derecho civil romano:** a doble texto, traducido al castellano del latino, t. 3. Barcelona: Lex Nova, 1897.

VIEIRA, Leonardo Marques. A problemática da inteligência artificial e dos vieses algorítmicos: caso COMPAS. In: **Brazilian Technology Symposium**. 2019.

VON SCHLIEFFEN, Katharina Gräfin. **Iluminismo retórico: contribuições para uma Teoria Retórica do Direito.** Trad. João Maurício Adeodato. Curitiba: Alteridade, 2022.

WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. **International Data Privacy Law**, v. 7, n. 2, p. 76–99, 2017

WILHELM, Salvador Leal. Toma de decisionies administrativas e inteligencia artificial. in: PIRELA, Michelle Azuaje. **Introducción a la Ética y el Derechos de la Inteligencia Artificial.** Madrid: La Ley, p. 135-150, 2023.

YOUNG, Iris Marion. **Justice and the Politics of Difference**. Princeton: Princeton University Press, 1990.

ZUBOFF, Soshana. **The age of surveillance capitalism: The fight for a human future at the new frontier of power.** Nova Iorque: Public Affairs, 2019.