

# Big Data: os desafios à regulamentação do uso de dados

---

Eleonora Mesquita Ceia  
Pedro Teixeira Gueiros

## Resumo

O artigo examina os desafios da regulamentação do uso de dados em um cenário global marcado pelo avanço exponencial das tecnologias digitais e pela centralidade informacional do *Big Data*. Parte-se da constatação de que os dados se tornaram o principal ativo econômico da era digital, impulsionando modelos de negócio baseados em vigilância algorítmica e na extração massiva de informações pessoais. O estudo estrutura-se em três partes: inicialmente, parte-se de uma análise dos riscos e potencialidades associados ao *Big Data*, destacando a natureza não neutra das decisões algorítmicas. Em seguida, passa-se a uma abordagem comparativa do panorama regulatório, com ênfase no *AI Act* da União Europeia, no Projeto de Lei nº 2.338/2023 no Brasil e na Resolução do Conselho Nacional de Justiça nº 615/2025. A seguir, faz-se uma proposição de caminhos normativos para o fortalecimento da governança democrática dos dados, com base em quatro eixos: regulação coordenada e multissetorial, transparência algorítmica, promoção da justiça de dados e educação digital. Por fim, sustenta-se que regular o *Big Data* apresenta-se, antes de tudo, um enfrentamento de uma questão política e moral.

## Abstract

This article examines the challenges of regulating data use in a global scenario characterized by the exponential advancement of digital technologies and the informational centrality of Big Data. It begins with the observation that data has become the primary economic asset of the digital age, driving business models based on algorithmic surveillance and the extensive extraction of personal information. The study is structured in three parts: first, an analysis of the risks and potential associated with Big Data, highlighting the non-neutral nature of algorithmic decisions. This is followed by a comparative approach to the regulatory landscape, with emphasis on the European Union's AI Act, Brazilian Bill of Law No. 2.338/2023, and the Resolution No. 615/2025 from the Brazilian National Council of Justice. Next, a proposal is made for regulatory paths to strengthen democratic data governance, based on four axes: coordinated and multisectoral regulation, algorithmic transparency, promotion of data justice, and digital education. Finally, it is argued that regulating Big Data is first and foremost, a political and moral issue.

## 1. Introdução

Em uma breve pesquisa sobre as marcas mais valiosas do mundo, é possível observar que empresas de tecnologia se mantêm invariavelmente hegemônicas em termos de liderança na atuação do mercado. *Apple, Microsoft, Google, Amazon, Samsung, TikTok e Facebook*, são alguns exemplos que encabeçam as dez primeiras posições do ranking atualizado em 2025.<sup>1</sup> Ainda assim, é curioso constatar como uma ampla série de serviços disponibilizados por essas empresas são efetivamente

---

1 PIO, Juliana. *As marcas mais valiosas do mundo em 2025 (duas delas são brasileiras)*. Disponível em: <https://exame.com/marketing/saiba-quais-sao-as-marcas-mais-valiosas-do-mundo-em-2025-duas-delas-sao-brasileiras/>. Acesso em: 10.05.2025.

gratuitos, como redes sociais, serviços de mensageria ou apenas provedores de buscas. Nesse cenário, passa a ser questionável como tais companhias podem ser até mesmo mais valiosas que o PIB de países inteiros,<sup>2</sup> quando não há produtos efetivamente custosos, a rigor, sendo comercializados.<sup>3</sup>

Nesse sentido, a escalada vultosa de grandes corporações tecnológicas, também conhecidas como *Big Techs*, está indiscutivelmente imbricada no uso massivo de um ativo que parece ter se tornado uma verdadeira *commodity*: os dados pessoais. Sob esse fenômeno diversas expressões foram cunhadas por estudiosos com o intuito de compreender a nova ordem econômica mundial, que se edificou de forma tão recente na história humana contemporânea: *capitalismo de vigilância*,<sup>4</sup>

---

2 A título de exemplo, o valor de mercado da *Apple* alcançou o valor de US\$ 3 trilhões em 2023, situação em que apenas seis países no mundo à época tinham o Produto Interno Bruto superior a este valor: Estados Unidos, China, Índia, Japão, Alemanha e Reino Unido. GUILHERME, Guilherme. *Apenas 6 países têm PIB maior que valor de mercado da Apple*. Disponível em: <https://exame.com/in-vest/mercados/apenas-6-paises-tem-pib-maior-que-valor-de-mercado-da-apple/amp/>. Acesso em: 11.05.2025.

3 Nesse sentido, Kris Shaffer observa: “Algorithmic recommendation engines and social media feeds have been created to help users find the most relevant content and to help media producers find the most appropriate audiences. But the ways in which media producers compete for our attention, the amount of personal data mined to make the algorithms work, and the natural way our cognitive systems function all combine to make the modern media landscape ripe for propaganda”. SHAFFER, Kris. *Data versus Democracy: How Big Data Algorithms Shape Opinions and Alter the Course of History*. Nova York: Apress, 2019, p. 17-18.

4 Acerca do capitalismo de vigilância, Shoshana Zuboff, pesquisadora de psicologia social, descreve: “Os capitalistas de vigilância fazem um grande esforço para camuflar seu propósito enquanto dominam o uso do poder instrumentário para moldar nosso comportamento e, ao mesmo tempo, escapar da nossa consciência. É por isso que o Google esconde as operações que nos transformam em objetos de busca e o Facebook nos distrai do fato de que nossas tão adoradas conexões são essenciais para o lucro e poder que fluem da ubiquidade da rede e de seu conhecimento totalista”. ZUBOFF, Shoshana. *A era do capitalismo de vigilância: a luta por um futuro humano na nova fronteira do poder*. Rio de Janeiro: Intrínseca, 2020, p. 500.

*dataísmo*,<sup>5</sup> *armas de destruição de massa*,<sup>6</sup> *infocracia*,<sup>7</sup> dentre outras nomenclaturas. Em comum, observa-se que as mais variadas atividades humanas dependem progressivamente do processamento de máquinas inteligentes que se desenvolvem a partir de um número quase incalculável de dados e informações. Sob essa lógica, a confiança tradicionalmente atribuída a algoritmos humanos, naturalmente limitados, dá lugar à confiança a algoritmos efetivamente artificiais, cuja potencialidade é exponencial e de certa forma, irrestrita, como verificado pela *Lei de Moore*.<sup>8</sup>

- 
- 5 O historiador Yuval Noah Harari observa que: “O dataísmo adota uma abordagem estritamente funcional em relação à humanidade, avaliando o valor das experiências humanas de acordo com sua função nos mecanismos de processamento de dados. Se desenvolvermos um algoritmo que realize melhor a mesma função, as experiências humanas perderão seu valor”. HARARI, Yuval Noah. *Homo Deus: uma breve história do amanhã*. São Paulo: Companhia das Letras, 2016, p. 391.
- 6 Como pondera a matemática Cathy O’Neil acerca das denominadas *Armas de Destruição em Massa* (ADMs): “Dados não vão desaparecer. Nem computadores – e muito menos a matemática. Modelos de previsão são, cada vez mais, as ferramentas com as quais contaremos para administrar nossas instituições, aplicar nossos recursos e gerenciar nossas vidas. Mas, como tentei mostrar ao longo deste livro, esses modelos são construídos não apenas de dados, mas das escolhas que fazemos sobre em quais dados prestar atenção – e quais deixar de fora. Essas escolhas não tratam apenas de logísticas, lucros e eficiência. Elas são fundamentalmente morais”. O’NEIL, Cathy. *Armas de Destruição em Massa: como o big data aumenta a desigualdade e ameaça a democracia*. São Paulo: Rua do Sabão, 2020, p. 337.
- 7 Nesse sentido, o filósofo Byung-Chul Han, “A digitalização do mundo da vida avança, implacável. Submete a uma mudança radical nossa percepção, nossa relação com o mundo, nossa convivência. Ficamos atordoados pela embriaguez de comunicação e informação. O tsunami de informação desencadeia forças destrutivas. Abrange também, nesse meio-tempo, âmbitos políticos e leva a fraturas e rupturas massivas no processo democrático. A democracia degenera em *infocracia*”. HAN, Byung-Chul. *Infocracia: digitalização e a crise da democracia*. Petrópolis: Vozes, 2022, p. 25.
- 8 Conjecturada em 1965, por Gordon Moore, cofundador da Intel, a “Lei de Moore” ficou conhecida como sendo a observação empírica que prevê que o número de transistores de um *chip* dobra aproximadamente a cada 18 meses, mantendo-se o mesmo custo. A rigor, esta tendência impulsionou o avanço exponencial da capa-

As atuais e populares Inteligências Artificiais Generativas, ilustrativamente, capazes de gerar conteúdos inéditos altamente criativos, como textos, imagens e músicas, dependem de uma quantidade colossal de dados para que os grandes modelos de linguagem sejam instrumentalizados, sob os chamados *data lakes*.<sup>9</sup> Em 2023, por exemplo, o célebre *ChatGPT* da OpenAI, em seu modelo GPT-3, utilizava entre 570 *gigabytes* e 45 *terabytes* de informações coletadas na internet.<sup>10</sup> Nesse cenário, o *Big Data*, enquanto fenômeno transversal às diversas ciências do conhecimento, apresenta uma miríade de desafios e potencialidades. Sob a perspectiva jurídica, uma das principais questões a ser endereçada diz respeito à regulação das atividades que se desenvolvem à luz do grande processamento de dados.

O presente artigo estrutura-se, nesse sentido, na compreensão inicial dos riscos e das potencialidades associadas ao *Big Data*. Em seguida, passa-se ao exame do atual panorama regulatório, por meio de uma análise comparativa entre diferentes jurisdições. Por fim, propõe-se alguns apontamentos para conferir maior legitimidade às atividades desenvolvidas no contexto do *Big Data*.

---

cidade computacional em paralelo à diminuição de dispositivos eletrônicos. JORDÃO, Fabio. *Afinal, o que é a Lei de Moore?* Disponível em: <https://www.tecmundo.com.br/curiosidade/701-o-que-e-a-lei-de-moore-.htm>. Acesso em: 12.05.2025.

- 9 Imprescindível ao funcionamento de Inteligências Artificiais, como um todo, “Um *data lake* é um repositório centralizado que ingere e armazena grandes volumes de dados em sua forma original. Os dados podem ser processados e usados como base para uma variedade de necessidades analíticas.” MICROSOFT. *O que é um data lake*. Disponível em: <https://azure.microsoft.com/pt-br/resources/cloud-computing-dictionary/what-is-a-data-lake>. Acesso em: 12.05.2025
- 10 OPENAI. *What is the size of the training set for GPT-3*. Disponível em: <https://community.openai.com/t/what-is-the-size-of-the-training-set-for-gpt-3/360896>. Acesso em: 11.05.2025.

## 2. *Big Data*: entre riscos e potencialidades

**O** *Big Data*, enquanto uma expressão em certa medida amorfa voltada à compreensão de atividades que envolvem um grande volume de dados complexos e diversificados, impõe nuances concretas para que seus riscos e potencialidades sejam examinados. Justamente em razão disso, para colocar o pingô nos is, é necessário responder a seguinte pergunta: de que tipo de “dados” está se falando?

No Brasil, dados e informações de interesse público, por exemplo, estão sujeitas à regulação da Lei de Acesso à Informação (Lei nº 12.527/2011),<sup>11</sup> o que impõe ao Poder Público o dever de publicizar, como regra, as atividades desenvolvidas. Informações podem ainda ser de caráter pessoal, o que a rigor, obriga ao cumprimento das diretrizes estabelecidas pela Lei Geral de Proteção de Dados (Lei nº 13.709/2018),<sup>12</sup> sem prejuízo de outras previsões específicas, como relações de consumo e as determinações do Código de Defesa do Consumidor (Lei nº 8.078/1990)<sup>13</sup> e da Lei do Cadastro Positivo (Lei

---

11 Como determina as diretrizes da LAI: “Art. 3º Os procedimentos previstos nesta Lei destinam-se a assegurar o direito fundamental de acesso à informação e devem ser executados em conformidade com os princípios básicos da administração pública e com as seguintes diretrizes: I – observância da publicidade como preceito geral e do sigilo como exceção; II – divulgação de informações de interesse público, independentemente de solicitações; III – utilização de meios de comunicação viabilizados pela tecnologia da informação; IV – fomento ao desenvolvimento da cultura de transparência na administração pública; V – desenvolvimento do controle social da administração pública”.

12 Como se extrai dos objetivos da LGPD: “Art. 1º Esta Lei dispõe sobre o 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”.

13 Conforme expressa determinação do CDC: “Art. 43. O consumidor, sem prejuízo do disposto no art. 86, terá acesso às informações existentes em cadastros, fichas, registros e dados pessoais e de consumo arquivados sobre ele, bem como sobre as suas respectivas fontes”.

nº 12.414/2011),<sup>14</sup> ou ainda no bojo de relações virtuais, às regras do Marco Civil da Internet (Lei nº 12.965/1014).<sup>15</sup> Outras informações podem envolver também tutelas mais específicas, e nem sempre tão óbvias, a exemplo de obras criativas, sinais distintivos, patentes, segredos comerciais, que estão sujeitos à regulação em matéria de Propriedade Intelectual, notadamente por meio da Lei de Direitos Autorais (Lei nº 9.610/1998)<sup>16</sup> e da Lei de Propriedade Industrial (Lei nº 9.279/1996).<sup>17</sup> No que se refere ao *Big Data*, a questão se coloca em termos da escala com que tantas informações, de natureza jurídica distintas, são processadas, o que impõe verdadeira governança holística em atenção aos diferentes microssistemas.

A economia do *Big Data* trouxe consigo a promessa de notáveis ganhos e oportunidades, mediante a análise de imensos volumes de dados diversos, de forma veloz, imparcial e objetiva. Porém, tal análise parte de modelos matemáticos e estatísticos desenvolvidos por seres

---

14 Como se depreende dentre os direitos facultados à pessoa cadastrada em bancos de adimplemento na LCP: “Art. 5º São direitos do cadastrado: II – acessar gratuitamente, independentemente de justificativa, as informações sobre ele existentes no banco de dados, inclusive seu histórico e sua nota ou pontuação de crédito, cabendo ao gestor manter sistemas seguros, por telefone ou por meio eletrônico, de consulta às informações pelo cadastrado”.

15 Como se observa dos deveres incumbidos aos provedores de conexão de internet e provedores de aplicações de internet no MCI: “Art. 10. A guarda e a disponibilização dos registros de conexão e de acesso a aplicações de internet de que trata esta Lei, bem como de dados pessoais e do conteúdo de comunicações privadas, devem atender à preservação da intimidade, da vida privada, da honra e da imagem das partes direta ou indiretamente envolvidas”.

16 Nos termos do art. 7º, da LDA, ao apresentar as obras protegidas por direitos autorais, para além textos, fotografias, pinturas, esculturas, filmes, músicas, o inciso XIII, menciona expressamente que também são protegidas “as coletâneas ou compilações, antologias, enciclopédias, dicionários, bases de dados e outras obras, que, por sua seleção, organização ou disposição de seu conteúdo, constituam uma criação intelectual”.

17 Paralelamente, entende-se que são protegidos mediante registro invenções relacionadas a patentes, modelos de utilidade, marcas, indicações geográficas, além de jogos eletrônicos, conforme art. 2º da LPI.

humanos imbuídos de preconceitos e vieses. O problema é agravado quando se constata que esses modelos são opacos, isto é, não compreensíveis às pessoas em geral, salvo aquelas que dominam a complexidade algorítmica neles envolvida. Além disso, os modelos são protegidos por regras legais de segredos comercial e industrial<sup>18</sup>. Isso dificulta ou até mesmo impede questionamentos contra decisões tomadas no âmbito do *Big Data*. Em síntese, o *Big Data* abarca operações não transparentes, dificilmente contestáveis e, em muitos casos, ainda não passíveis de regulamentação<sup>19</sup>.

A maior preocupação, no que diz respeito aos riscos do *Big Data*, está associada, não apenas ao risco da mera inobservância do uso de dados em sentido *lato*, mas também, às consequências negativas diante do risco de violação a direitos fundamentais. Assim sendo, importa reconhecer que o *Big Data*, enquanto expressão inerente ao atual cenário de novas tecnologias, não se traduz em um fenômeno neutro *per se*. Trata-se de uma escolha tecnológica ligada a relações e estruturas de poder, que beneficiam uns e excluem outros. Vale dizer, muito embora se apresente conjecturado à luz de critérios estáticos e analíticos baseados nas ciências exatas, sua aplicação não deixa de ser efetivamente social e co-

---

18 Frank Pasquale trata dessa questão alertando para a presença massiva de ferramentas tecnológicas em domínios vitais da nossa vida civil, como saúde, educação e trabalho, que funcionam como verdadeiras “caixas pretas”. Isso porque não são transparentes com relação a suas funções, finalidades e uso de dados, devido à alta complexidade tecnológica que apenas suas proprietárias privadas dominam, cujo sigilo é protegido pelo Estado por meio de legislação específica. PASQUALE, Frank. *The Black Box Society: The Secret Algorithms that Control Money and Information*. Cambridge/London: Harvard University Press, 2015.

19 “As aplicações matemáticas fomentando a economia dos dados eram baseadas em escolhas feitas por seres humanos falíveis. Algumas dessas escolhas sem dúvida eram feitas com as melhores intenções. Mesmo assim, muitos desses modelos programavam preconceitos, equívocos e vieses humanos nos sistemas de software que cada vez mais geriam nossas vidas. Como deuses, esses modelos matemáticos eram opacos, seus mecanismos invisíveis a todos exceto aos altos sacerdotes de seus domínios: os matemáticos e cientistas da computação. Suas decisões, mesmo quando erradas ou danosas, estavam para além de qualquer contestação”. O’NEIL, *op. cit.*, p. 12.



letiva, o que leva a que quaisquer tomadas de decisões impliquem em consequências de ordem política ou moral.

Por exemplo, no caso de uma máquina inteligente ser empregada para designar as chances de uma pessoa ser melhor qualificada para uma vaga de trabalho, ou ainda, de obter a concessão de um empréstimo bancário ou de reincidir em algum crime anteriormente praticado, implica invariavelmente em uma escolha e decisão humana dos desenvolvedores do sistema algorítmico em autorizar o seu uso dentro daquele determinado contexto.

Com efeito, a publicidade com uso intensivo de dados se tornou muito lucrativa. As empresas do ramo formam perfis a partir dos dados coletados de atividades empreendidas pelos indivíduos nas redes sociais. Trata-se de vantagem informacional que possibilita a identificação de consumidores ou eleitores ideais. Os dados e perfis são vendidos ou compartilhados ao Estado<sup>20</sup>, empresas e empregadores. Essa atividade de perfilamento acaba construindo uma reputação digital dos indivíduos, por meio de informações do que elas consomem, suas preferências, interesses, hábitos, saúde, medicamentos que usam etc. O tratamento dos dados é realizado sem qualquer transparência, devido ao problema da opacidade, anteriormente explicado. Isso levanta preocupações em termos de tratamento justo de dados, pois o perfilamento é realizado sob vieses humanos, com a possibilidade de reproduzir padrões de desigualdade e discriminação, além da violação de direitos da personalidade<sup>21</sup>.

---

20 “The mountains of data collected by private corporations make them valuable partners in ‘information sharing’. There’s plenty of room for dealing on both sides. Government agencies want data that they can’t legally or constitutionally collect for themselves; data brokers have it and want to sell it”. PASQUALE, *op. cit.*, p. 48-49.

21 “Reputation systems are creating new (and largely invisible) minorities, disfavored due to error or unfairness. Algorithms are not immune from the fundamental problem of discrimination, in which negative and baseless assumptions congeal into prejudice”. PASQUALE, *op. cit.*, p. 38.

Com relação à democracia, as ferramentas tecnológicas, como a IA, os algoritmos e o *Big Data*, operam em termos contraditórios. As pessoas utilizam as redes sociais para exercerem liberdades fundamentais, como a informação, a comunicação e a manifestação do pensamento. Assim, nossa comunicação e o debate público ocorrem cada vez mais em espaços digitais privados, que são geridos por uma tecnologia complexa e não transparente. Isso resulta no monopólio e na concentração da informação pelas grandes empresas de Internet. As redes sociais, por meio de seus algoritmos, passam a deter o poder de influenciar o processo político e a opinião pública, por meio do direcionamento de notícias e posts, incluindo desinformação, ou seja, para fins não democráticos<sup>22</sup>.

Igualmente, o *Big Data* possui uma relação ambígua com o meio ambiente. De um lado, em fóruns internacionais, parte-se da premissa de que a análise de um grande volume de informações, coletadas automaticamente, formam subsídios fundamentais para a formulação de políticas de desenvolvimento sustentável, o monitoramento das mudanças climáticas e o combate ao desmatamento. De outro, neste debate foi negligenciada a preocupante “pegada ambiental” do *Big Data*<sup>23</sup>. Tecnologias da escala do *Big Data* demandam um alto consumo de energia nas instalações de *data center*. Além disso, nessas instalações “precisam operar geradores a diesel em caso de falta de energia para manter os servidores funcionando, o que acarreta a emissão de

---

22 SANTOS, Andréia. O Impacto do Big Data e dos Algoritmos nas Campanhas Eleitorais. In: BRANCO, Sérgio; DE TEFFÉ, Chiara (orgs.). *Privacidade em perspectivas*. Rio de Janeiro: Lumen Juris, 2018, p. 1-24.

23 “High data volume requires extensive facilities for storage using natural resources such as water and non-renewable energy that require maintenance and accessibility. Furthermore, the manufacturing and disposal of devices to collect and process data emits polluting substances. The data revolution, advocated as a vehicle to achieve sustainable development, is supported by technologies that endanger sustainability and the environment by other means”. LUCIVERO, Federica. *Big Data, Big Waste? A Reflection on the Environmental Sustainability of Big Data Initiatives*. *Science and Engineering Ethics*, v. 26, 2020, p. 1013.

gases de efeito estufa com impactos significativos sobre a mudança climática”<sup>24</sup>.

Outro problema ambiental ligado ao *Big Data* são as emissões nocivas provenientes do descarte do hardware de computação. Os centros de descarte e reciclagem estão localizados, em geral, em países menos desenvolvidos, o que suscita o debate sobre a justa distribuição dos benefícios e danos da revolução digital. de modo a evitar que “as populações mais vulneráveis sofram os efeitos ambientais da crescente demanda digital dos países mais desenvolvidos”<sup>25, 26</sup>.

Diante da presença do *Big Data* em variados domínios da vida dos indivíduos e das relações entre os países, junto com os significativos impactos sobre a democracia e a justiça social e ambiental, justifica-se a emergência de iniciativas de regulação, com o propósito de garantir as devidas responsabilidades no caso de abusos e violações a direitos fundamentais.

### 3. Panorama regulatório nacional e internacional

Diante da ubiquidade do *Big Data* em sua progressiva confluência das atividades desenvolvidas por agentes públicos e privados, diversas jurisdições têm buscado estabelecer marcos normativos voltados a mitigar os riscos e promover o uso ético e responsável dos dados.

---

24 LUCIVERO, *op. cit.*, p. 1015 [tradução livre dos autores].

25 LUCIVERO, *op. cit.*, p. 1022 [tradução livre dos autores].

26 Interessante notar que usualmente o discurso adotado referente às tecnologias digitais utiliza termos (imateriais, abundantes e nuvens, por exemplo), que ocultam e minimizam o real impacto da infraestrutura física do *Big Data*. Nesse sentido, Lucivero adverte: “The digital has material implications for the physical “real” world, but the language used obscures this. Places, infrastructures, and buildings all play a role in constituting the online world, and vice versa: online behaviour also has material implications in the physical “real” world. False assumptions, created by misleading language, influence people’s understanding of the data universe and economy, and shapes their attitudes towards it”. LUCIVERO, *op. cit.*, p. 1019.

Sob a realidade hiperconectada, marcada pelo uso massivo de dados instrumentalizados por ferramentas de IA, a regulação contemporânea enfrenta o desafio de acompanhar a velocidade exponencial do desenvolvimento tecnológico, em atenção à preservação de direitos fundamentais e à governança democrática da informação.

No plano internacional, destaca-se a recente promulgação do *AI Act* pela União Europeia em meados de 2024. Trata-se do primeiro regulamento geral voltado à disciplina da IA de forma integral, estabelecendo diretrizes para seu uso em uma abordagem baseada em riscos. Sob esta perspectiva, os sistemas de IA são classificados conforme o grau de potenciais impactos a direitos e liberdades dos cidadãos, visando promover a segurança jurídica e equidade no acesso e disponibilização de soluções movidas à IA. A depender dessa categorização, o regulamento impõe obrigações mais ou menos rigorosas quanto à transparência, mitigação de vieses, supervisão humana e robustez dos dados utilizados na modelagem.<sup>27</sup> Ainda que o *AI Act* não se debruce diretamente sobre o fenômeno do *Big Data*, esta regulação apresenta-se harmonizada frente a toda árvore regulatória da UE, que sistematiza importantes diretivas na “Era dos Dados”, a exemplo do Regulamento Geral de Proteção de Dados (GDPR), Regulamento dos Serviços Digitais (DSA), Regulamento dos Mercados Digitais (DMA), dentre outras.

Em perspectiva comparativa, o Brasil tem buscado estruturar sua abordagem regulatória relativamente quanto a um “Marco Geral de

---

27 A título ilustrativo, IAs de risco inaceitável são aquelas expressamente proibidas por violar direitos fundamentais, a exemplo de técnicas subliminares que alterem a consciência humana ou que explorem a vulnerabilidade das pessoas em razão de idade, incapacidade ou situação socioeconômica (art. 5º). Por sua vez, IAs de risco elevado são soluções que, em razão da função que desempenham, podem oferecer riscos significativos, a exemplo sistemas de identificação biométrica ou ainda sistemas utilizados na educação e na formação profissional, devendo, portanto, cumprir uma série de requisitos para que sejam disponibilizadas no mercado (art. 6º e Anexo III). Tem-se ainda, IAs de risco sistêmico, isto é, IAs de propósito geral que, devido à sua escala, poder computacional e capacidade de influência, representam riscos significativos, devendo, nesse sentido, cumprir critérios específicos para que sejam igualmente disponibilizados (art. 51).

IA” por meio do Projeto de Lei nº 2.338/2023. Aprovado em texto substitutivo no Senado Federal, o PL tramita atualmente na Câmara dos Deputados e visa estabelecer, de forma muito semelhante ao modelo europeu, uma regulação de IA baseada em riscos, trazendo princípios e diretrizes voltadas à responsabilidade de agentes e fornecedores de IA e à proteção dos direitos fundamentais.<sup>28</sup> O projeto também prevê a criação de um Sistema Nacional de Governança de Inteligência Artificial capaz de integrar órgãos reguladores setoriais como a Autoridade Nacional de Proteção de Dados (ANPD).<sup>29</sup> A proposta se torna especialmente relevante em um país que enfrenta desafios históricos de desigualdade estrutural e de discriminação algorítmica, sendo essencial garantir que o uso massivo de dados não reproduza, de forma automatizada, preconceitos e exclusões já presentes no tecido social.

Observa-se, ainda, de forma setorial, a atuação do Conselho Nacional de Justiça (CNJ) ao aprovar a recente Resolução n.º 615/2025, que introduz um importante avanço ao regular diretamente o uso de sistemas de IA no âmbito das atividades judiciais. O texto normativo determina que os sistemas utilizados por tribunais devem observar os princípios da supervisão humana, da transparência, da imparcialidade,

---

28 Conforme se destaca do art. 1º: “1º Esta Lei estabelece normas gerais de caráter nacional para a governança responsável de sistemas de inteligência artificial (IA) no Brasil, com o objetivo de proteger os direitos fundamentais, estimular a inovação responsável e a competitividade e garantir a implementação de sistemas seguros e confiáveis, em benefício da pessoa humana, do regime democrático e do desenvolvimento social, científico, tecnológico e econômico”.

29 Como se extrai do § 2º do art. 1º: “A fim de promover o desenvolvimento tecnológico nacional, o Sistema Nacional de Regulação e Governança de Inteligência Artificial (SIA) regulamentará regimes simplificados, envolvendo flexibilização de obrigações regulatórias previstas nesta Lei, nos seguintes casos: I – padrões e formatos abertos e livres, com exceção daqueles considerados de alto risco; II – incentivo à inovação e à pesquisa científica e tecnológica no ambiente produtivo, com vistas à capacitação tecnológica, ao alcance da autonomia tecnológica e ao desenvolvimento do sistema produtivo nacional e regional do País; III – projetos de interesse público, os que atendam as prioridades das políticas industrial e de ciência, tecnologia e inovação e os que visem à solução dos problemas brasileiros”.

da rastreabilidade e da segurança da informação.<sup>30</sup> Nesse sentido, o CNJ reconhece que os dados, especialmente quando processados em larga escala, possuem centralidade na produção de decisões judiciais automatizadas e devem ser auditáveis, éticos e justos. A resolução também estabelece que nenhum sistema pode substituir a autonomia da magistratura ou comprometer o devido processo legal, em atenção aos riscos que o *Big Data*, operado por IA, representa quando não controlado adequadamente.<sup>31</sup> Ainda que limitada à esfera judicial, essa norma representa um esforço institucional importante de internalização de boas práticas de governança algorítmica no setor público.

Dessa forma, tanto no plano internacional como no interno, observa-se uma movimentação normativa que reconhece a complexidade do ecossistema de dados e a necessidade de abordagens regulatórias transversais. O *Big Data*, por operar sobre diversas naturezas de informações e ser insumo essencial para a modelagem de sistemas preditivos, não pode ser compreendido de forma fragmentada. A regulação eficaz exige um olhar sistêmico, que articule proteção de dados, segurança informacional, responsabilidade algorítmica e participação democrática. Os marcos apresentados, ainda que incipientes e em constante construção,

---

30 Como se destaca do art. 1º: “A presente Resolução estabelece normas para o desenvolvimento, a governança, a auditoria, o monitoramento e o uso responsável de soluções que adotam técnicas de inteligência artificial (IA) no âmbito do Poder Judiciário, com o objetivo de promover a inovação tecnológica e a eficiência dos serviços judiciários de modo seguro, transparente, isonômico e ético, em benefício dos jurisdicionados e com estrita observância de seus direitos fundamentais”.

31 Nesse sentido: “Art. 32. O sistema inteligente deverá assegurar a autonomia dos usuários internos, com o uso de modelos que: I – promovam o incremento da eficiência, precisão e qualidade das atividades, sem limitar a capacidade de atuação dos usuários; II – possibilitem a revisão detalhada do conteúdo gerado e dos dados utilizados para sua elaboração, assegurando que os usuários tenham acesso às premissas e ao método empregado pela inteligência artificial na sua formulação, sem que haja qualquer espécie de vinculação à solução apresentada pela inteligência artificial e garantindo-se a possibilidade de correções ou ajustes. Parágrafo único. Em nenhum momento o sistema de IA poderá restringir ou substituir a autoridade final dos usuários internos”.

apontam para uma tendência de normatização orientada não apenas à eficiência tecnológica, mas à proteção da pessoa humana em uma sociedade global profundamente orientada por dados. Como observa Dora Kaufman, “[o] desafio é evitar que a legislação chegue tarde demais, quando a tecnologia já estiver incorporada na sociedade”.<sup>32</sup>

#### 4. Caminhos possíveis

Diante da complexidade estrutural e dos riscos sistêmicos associados ao *Big Data*, é possível inferir que a existência de normas isoladas ou setoriais não são suficientes para endereçar a questão. A construção de uma regulação eficaz requer um arranjo normativo integrado e dinâmico, capaz de lidar com o caráter complexo das tecnologias emergentes e com a velocidade de sua evolução.<sup>33</sup> Nesse sentido, ao menos quatro caminhos podem ser delineados como vetores para o aprimoramento da governança dos dados em larga escala.

Um primeiro caminho diz respeito à consolidação de uma abordagem regulatória coordenada e multissetorial, que articule diferentes microssistemas jurídicos, como a proteção de dados, defesa do consumidor, propriedade intelectual e livre concorrência, sob uma lógica de complementaridade. Esse alinhamento requer não apenas diálogo entre órgãos reguladores, mas também a adoção de estruturas de interoperabilidade normativa, a exemplo da proposta brasileira de um Sistema Nacional de Governança de IA.

---

32 KAUFMAN, Dora. *Desmistificando a inteligência artificial*. Belo Horizonte: Autêntica, 2022, p. 163.

33 Como observam Mustafa Suleyman e Michael Bhaskar, acerca dos desafios da IA: “A regulamentação não depende somente da aprovação de uma nova lei. Ela depende também de normas, estruturas de propriedade, códigos não escritos de conformidade e honestidade, procedimentos de arbitragem, cumprimento de contratos, mecanismos de supervisão. Tudo isso precisa ser integrado, e o público precisa participar”. SULEYMAN, Mustafa; BHASKAR, Michael. *A próxima onda: inteligência artificial, poder e o maior dilema do século XXI*. Rio de Janeiro: Record, 2023, p. 287

Um segundo caminho envolve a valorização do princípio da transparência algorítmica e da auditabilidade dos sistemas que operam por meio do *Big Data*. Trata-se de assegurar que as decisões automatizadas, em especial as que afetam notadamente os direitos fundamentais, possam ser compreendidas, questionadas e eventualmente revertidas por seus titulares, a exemplo do que se verifica do art. 20 da LGPD.<sup>34</sup> Isso implica estabelecer deveres de explicabilidade com nível técnico proporcional ao grau de impacto e risco social da tecnologia, além da disponibilização de canais eficazes para a prestação de contas.

Um terceiro eixo consiste na promoção da justiça de dados (*data justice*),<sup>35</sup> conceito que reconhece a centralidade dos dados na reprodução de assimetrias históricas e busca integrar a dimensão distributiva e participativa à regulação. Para tanto, inclui-se a avaliação de impactos algorítmicos com foco em grupos vulneráveis, o incentivo à representatividade dos dados utilizados em treinamentos de IA e a criação de mecanismos de escuta social na formulação de políticas públicas orientadas por dados.<sup>36</sup>

---

34 Nos termos do referido dispositivo, assegura-se que o direito ao titular de dados de solicitar a revisão das decisões automatizadas envolvendo o tratamento de dados que afetem notadamente os interesses do titular, incluindo aquelas voltadas para a construção de perfil pessoal, profissional, de consumo e de crédito e ainda aspectos da personalidade.

35 De acordo com a doutrina, a “justiça de dados” pode ser conceituada a partir de três pilares nevrálgicos: (i) visibilidade, no sentido dos dados serem utilizados de forma representativa e em respeito à privacidade; (ii) engajamento com as tecnologias, relativamente, quanto à autonomia individual frente às tecnologias e sistemas de dados e; (iii) antidiscriminação, de modo a prevenir usos injustos ou discriminatórios dos dados. DENCİK, Lina et al. *Towards data justice? The ambiguity of anti-surveillance resistance in political activism*. *Big Data & Society*, v. 3, n. 2, nov. 2016. Disponível em: <https://journals.sagepub.com/doi/10.1177/2053951716679678>. Acesso em: 06.07.2025.

36 A exemplo do que se observa no bojo da proposta do PL 2.338/2023, que inclui o dever de realização da Avaliação de Impacto Algorítmico: “Art. 25. A avaliação de impacto algorítmico de sistemas de IA é obrigação do desenvolvedor ou do aplicador que introduzir ou colocar sistema de IA em circulação no mercado, sempre



Por fim, um quarto caminho passa pela educação digital e pelo fortalecimento da cultura de dados, tanto entre usuários quanto entre tomadores de decisão. A compreensão crítica sobre como dados são coletados, tratados e utilizados é condição essencial para a construção de uma cidadania informada e para o exercício efetivo de direitos em uma sociedade movida por informações. Essa capacitação deve envolver não apenas a alfabetização de aspectos técnicos, mas também ético e jurídico, em um esforço transversal de políticas públicas educacionais.<sup>37</sup>

A conjugação desses caminhos não se traduz em uma eliminação dos desafios, mas contribui para uma abordagem normativa mais eficaz e democrática. Regular o *Big Data* revela-se, em última instância, instrumentalizar os meios de controle sobre o poder informacional em uma sociedade global progressivamente digital. A escolha entre uma potencial opacidade de ordem tecnocrática e uma governança multissetorial participativa dependerá do enfrentamento ao tema como uma questão multifacetada, isto é, de ordem técnica e científica, mas igualmente política e moral.

## Referências bibliográficas

DENCIK, Lina et al. Towards data justice? The ambiguity of anti-surveillance resistance in political activism. **Big Data & Society**, v. 3, n. 2, nov. 2016. Disponível em: <<https://journals.sagepub.com/doi/10.1177/2053951716679678>>. Acesso em: 06.07.2025.

ECO, Umberto. **Pape Satàn aleppe**: crônicas de uma sociedade líquida. Rio de Janeiro: Record, 2019.

---

que o sistema ou o seu uso forem de alto risco, considerando o papel e a participação do agente na cadeia”.

- 37 Como reflete Umberto Eco: “a internet oferece um repertório fantástico de informações, mas não os meios para selecioná-las, e a educação não consiste apenas em transmitir informação, mas também em ensinar critérios de seleção”. ECO, Umberto. *Pape Satàn aleppe: crônicas de uma sociedade líquida*. Rio de Janeiro: Record, 2019, p. 80.

GUILHERME, Guilherme. **Apenas 6 países têm PIB maior que valor de mercado da Apple**. Disponível em: <<https://exame.com/invest/mercados/apenas-6-paises-tem-pib-maior-que-valor-de-mercado-da-apple/amp/>>. Acesso em: 11.05.2025.

HAN, Byung-Chul. **Infocracia: digitalização e a crise da democracia**. Petrópolis: Vozes, 2022.

HARARI, Yuval Noah. **Homo Deus: uma breve história do amanhã**. São Paulo: Companhia das Letras, 2016.

JORDÃO, Fabio. **Afinal, o que é a Lei de Moore?** Disponível em: <<https://www.tecmundo.com.br/curiosidade/701-o-que-e-a-lei-de-moore-.htm>>. Acesso em: 12.05.2025.

KAUFMAN, Dora. **Desmistificando a inteligência artificial**. Belo Horizonte: Autêntica, 2022.

LUCIVERO, Federica. **Big Data, Big Waste? A Reflection on the Environmental Sustainability of Big Data Initiatives**. Science and Engineering Ethics, v. 26, 2020, p. 1009-1030. Disponível em: <<https://doi.org/10.1007/s11948-019-00171-7>>. Acesso em: 10.07.2025.

MICROSOFT. **O que é um data lake**. Disponível em: <<https://azure.microsoft.com/pt-br/resources/cloud-computing-dictionary/what-is-a-data-lake>>. Acesso em: 12.05.2025

O'NEIL, Cathy. **Armas de Destruição em Massa: como o big data aumenta a desigualdade e ameaça a democracia**. São Paulo: Rua do Sabão, 2020.

OPENAI. **What is the size of the training set for GPT-3**. Disponível em: <<https://community.openai.com/t/what-is-the-size-of-the-training-set-for-gpt-3/360896>>. Acesso em: 11.05.2025.

PASQUALE, Frank. **The Black Box Society: The Secret Algorithms that Control Money and Information**. Cambridge/London: Harvard University Press, 2015.

PIO, Juliana. **As marcas mais valiosas do mundo em 2025 (duas delas são brasileiras)**. Disponível em: <<https://exame.com/marketing/saiba-quais-sao-as-marcas-mais-valiosas-do-mundo-em-2025-duas-delas-sao-brasileiras/>>. Acesso em: 10.05.2025.

SANTOS, Andréia. O Impacto do Big Data e dos Algoritmos nas Campanhas Eleitorais. In: BRANCO, Sérgio; DE TEFFÉ, Chiara (orgs.). **Privacidade em perspectivas**. Rio de Janeiro: Lumen Juris, 2018, p. 1-24.

SHAFFER, Kris. **Data versus Democracy: How Big Data Algorithms Shape Opinions and Alter the Course of History**. Nova York: Apress, 2019.

SULEYMAN, Mustafa; BHASKAR, Michael. **A próxima onda: inteligência artificial, poder e o maior dilema do século XXI**. Rio de Janeiro: Record, 2023.

ZUBOFF, Shoshana. **A era do capitalismo de vigilância: a luta por um futuro humano na nova fronteira do poder.** Rio de Janeiro: Intrínseca, 2020.

---

**Eleonora Mesquita Ceia** · Doutora em Direito pela Universidade de Saarbrücken. Professora Titular de Direito Constitucional do Ibmec-RJ e Professora Adjunta de Teoria do Estado da Faculdade Nacional de Direito da Universidade Federal do Rio de Janeiro (UFRJ). Professora Colaboradora do Programa de Pós-Graduação em Direito da UFRJ.

**Pedro Teixeira Gueiros** · Mestre em Direito Civil pela PUC-Rio. Ex-bolsista da Fundação Konrad Adenauer. Professor Substituto de Direito Civil e Internacional Privado da Faculdade Nacional de Direito, UFRJ (2024-2025). Professor Assistente e Advogado Orientador do Núcleo de Prática Jurídica do Ibmec-RJ. Professor convidado dos cursos de pós-graduação da PUC-Rio, CEPED/UERJ e EMERJ. Integrante do Núcleo Legalite da PUC-Rio. Membro Suplente do Conselho Municipal de Proteção de Dados Pessoais e da Privacidade do Rio de Janeiro (CMPDPP). Graduado em Direito pelo Ibmec-RJ.