JULIANO MARANHÃO

IA

A PROTEÇÃO DE DADOS À LUZ DA IA RESPONSÁVEL

Legal Wings *institute*

lawgorithm*}

O **Legal Wings Institute** é um instituto independente, com atuação autônoma na produção de conhecimento, pesquisa, análise regulatória e promoção de debates públicos sobre direito, tecnologia, inovação, ética e regulação. Sua atuação se desenvolve com independência intelectual e metodológica, por meio da elaboração de estudos, relatórios, eventos, conteúdos e iniciativas de interlocução com diferentes setores.

O Instituto trabalha de forma colaborativa com atores do setor público, privado, acadêmico e da sociedade civil e conta com apoios institucionais, parcerias e patrocínios, que não implicam qualquer compromisso interesses específicos dos parceiros, nem condicionam suas agendas, métodos, análises ou conclusões. Para conhecer mais sobre o Legal Wings e consultar as informações disponibilizadas em seus canais oficiais, incluindo os sponsors/apoiadores do Instituto, acesse o site oficial.

https://www.legalwings.com.br

QUEM SOMOS

A Associação Lawgorithm de Pesquisa em Inteligência Artificial é constituída como Organização da Sociedade Civil de Interesse Público (OSCIP). Seu objetivo é desenvolver pesquisas e influenciar políticas públicas nas interfaces entre direito e inteligência artificial, em prol do bem comum, do desenvolvimento tecnológico e da justiça social. Atua em duas linhas de pesquisa: *(i)* Inteligência Artificial aplicada ao Direito e *(ii)* Direito da Inteligência Artificial.

SUMÁRIO

A PROTEÇÃO DE DADOS À LUZ DA IA RESPONSÁVEL

Juliano Maranhão[1]

Faculdade de Direito da Universidade de São Paulo

1. INTRODUÇÃO

A integração entre a proteção de dados pessoais e o desenvolvimento responsável de sistemas de Inteligência Artificial (IA) é *desideratum* urgente para a construção de um ambiente digital permeado por agentes inteligentes, que seja seguro e confiável.

Nesse esforço, duas perspectivas antagônicas se apresentam. A primeira, procedimental, vê a IA responsável pelas lentes do direito à proteção de dados, tomando a IA como caso particular do processamento de dados pessoais, que merece o escrutínio da legislação em cada etapa de tratamento, cada uma delas como potencial risco à autodeterminação informacional. A segunda, finalística, coloca a privacidade e proteção de dados pessoais em segundo plano, concentrando-se nos efeitos da aplicação do sistema de IA e ponderando seus benefícios potenciais frente aos valores éticos que compõem seu uso responsável, dentre eles, a privacidade, ao lado da confiabilidade, segurança, beneficência, não-maleficência, não-discriminação, transparência etc.

No presente documento, defende-se um caminho intermediário nesta oposição, a partir do princípio de *precaução substantiva* na interpretação da legislação de proteção de dados, à luz da ética de desenvolvimento e aplicação responsável da inteligência artificial.

Parte-se da concepção da IA como *agente* sociotécnico, sistema que reúne (e não se reduz) a dimensão de processamento informático e a dimensão organizacional humana, com a observação do impacto efetivo do resultado da ação desse agente (aplicação do sistema) sobre a privacidade e proteção de dados pessoais, não só como valor subjetivo, mas como valor social. A observação desse impacto é que determina a adequação das garantias procedimentais quanto ao tratamento de dados pessoais, tornando a substantiva a proteção da autodeterminação informacional. Ou seja, havendo impacto significativo sobre a projeção da personalidade individual ou ao valor social da

[1] Agradeço a Josie de Menezes Barros, Bernardo de Souza Dantas Fico, João Moreira Marquesini Salles Navas, Beatriz de Sousa e Ana Laura Azevedo Costa, pesquisadores da Associação Lawgorithm, pelo apoio na pesquisa para a realização do estudo.

privacidade, adotam-se as precauções e procedimentos para efetivamente garantir direitos e valores sociais objetivos protegidos pela legislação de proteção de dados.

Trata-se, efetivamente, de interpretar se o contexto e o resultado da aplicação do sistema de IA tornam a projeção da personalidade individual no ambiente digital, em primeiro lugar, relevante, e em segundo lugar, digna de proteção, considerando-se os riscos e virtudes do sistema. A precaução substantiva, como princípio orientador da intervenção com base na legislação de proteção de dados, pode revelar serem contraproducentes a adoção de restrições procedimentais ao tratamento, assim como pode revelar serem insuficientes as restrições previstas, para que se garanta o direito fundamental à proteção de dados.

Para desenvolver e explicitar a precaução substantiva na interpretação da legislação de proteção de dados, à luz da IA Responsável, este documento abordará, na Seção 2, a evolução do direito à proteção de dados pessoais na Europa, que parte de uma perspectiva finalística na interpretação jurisprudencial constitucional, traduz-se em uma abordagem procedimental na construção legislativa, e, ao final, recebe uma infusão doutrinária e jurisprudencial informada pelo valor social da privacidade e proteção de dados, que transborda a análise circunscrita à uma proteção egocêntrica do titular. A legislação brasileira, inspirada na europeia, já é influenciada pela perspectiva finalística e a consideração desse valor social, que se manifesta na maior abertura e amplitude na especificação de bases legais para o tratamento de dados pessoais.

Na Seção 3, será examinada a construção da ética e regulação da Inteligência Artificial, essencialmente estruturada na abordagem de riscos, que prioriza a avaliação dos impactos concretos dos *outputs* dos sistemas, o que se coaduna com o conceito de precaução substantiva, na qual o valor social da privacidade e proteção de dados são ponderados à luz da ética de IA Responsável. A análise de impacto de sistemas de IA e a sua classificação de risco, diz respeito aos efeitos concretos do emprego do sistema sobre indivíduos (direitos fundamentais) e valores sociais (democracia, sustentabilidade ambiental, ordem econômica e estabilidade financeira) e não sobre características, processos, componentes ou propriedades do sistema.

Na Seção 4, o documento tratará de junções críticas na interpretação da legislação da proteção de dados, onde a perspectiva procedimental apresenta vícios de interpretação, que podem implicar limitações ao desenvolvimento tecnológico e a benefícios sociais e econômicos, sem trazer contrapartidas relevantes em termos de proteção de individualidades, são eles:

(i) **"exigir a explicação de tudo"**: apegar-se à explicação causal na justificativa de tratamento de cada ponto de dado utilizado pelo sistema de IA;

(ii) **"enfatizar o tratamento em vez do resultado":** enfatizar o tratamento de dados, sem considerar o resultado e impacto final do tratamento com o desenvolvimento e implementação do sistema de IA;

(iii) **"supervalorizar a máquina"**: entender a decisão automatizada como operação meramente informática e não como ação sociotécnica;

(iv) **"confundir supervisão com revisão"**: acenar para o *desideratum* da supervisão humana no ciclo de vida a IA para exigir a revisão humana do output do sistema de IA

A aplicação do princípio de precaução substantiva pode contornar tais vícios, conciliando a proteção de dados pessoais com a inovação trazida por sistemas de IA e promovendo, assim, seu desenvolvimento e emprego de modo seguro, confiável e responsável.

2. A EVOLUÇÃO DA PROTEÇÃO DE DADOS PESSOAIS: ENTRE A ABORDAGEM FINALÍSTICA E PROCEDIMENTAL

Este item resgata a origem finalística da legislação e da jurisprudência da proteção de dados pessoais na Alemanha e Europa e sua evolução.

Na origem, a tutela se dirigia à proteção da *projeção da personalidade individual na esfera pública*, como decorrência da privacidade e da liberdade de expressão (e.g. direito a controlar aspectos da personalidade disponíveis ao público).[2] A chamada *autodeterminação informacional*, tal como cristalizada pelos tribunais europeus, diz respeito ao risco de impacto do tratamento de dados à personalidade individual e à democracia, neste último caso, em função da inibição da participação do cidadão na esfera pública, causada pela incerteza quanto às informações extraídas dos dados.

A tradução desse valor finalístico na legislação, por enfatizar a *accountability*, acabou por se transformar em análise procedimental de governança, em que o *processamento* de dados, em cada uma de suas etapas, passou a ocupar o centro da análise, em uma visão egocentrada no titular de dados pessoais.

Em seguida, a preocupação com o desenvolvimento da IA e o processamento de dados para promoção de interesses coletivos provocou uma releitura da legislação de proteção de dados em sentido oposto ao egocentrismo do titular e à visão procedimental, a partir de ferramentas como o legítimo interesse, impacto global do tratamento e à ponderação da proteção de dados como valor social *vis a vis* conjunto mais amplo de valores e interesses coletivos.

O "transplante" dessa evolução legislativa e jurisprudencial para o ordenamento brasileiro já incorpora aspectos finalísticos no próprio conjunto de bases legais, como a finalidade de proteção à saúde ou a finalidade de proteção ao crédito, que foram apenas incorporados ao direito europeu pela via jurisprudencial, a partir de análise do propósito e impacto global do tratamento frente a outros interesses individuais e coletivos relevantes.

2.1. Origem Finalística da Proteção de Dados

O termo direito à proteção de dados, ou mesmo a expressão "titular de dados", contêm um abuso de linguagem, pois a correspondente legislação não protege a propriedade ou

[2] WESTIN, Alan F. Privacy and freedom. Washington and Lee Law Review, v. 25, n. 1, p. 166, 1968.; WARREN, Samuel; BRANDEIS, Louis. The right to privacy. In: Killing the Messenger: 100 Years of Media Criticism. Columbia University Press, 1989. p. 1-21.

a detenção exclusiva do dado pessoal.[3] O bem fundamental, digno de defesa, não é propriamente o dado (o que seria inútil quando os dados já estão em posse de terceiros), tampouco a informação extraída do dado (pois essa é uma construção alheia do significado)[4], mas a *projeção da personalidade individual na esfera pública*. [5] Os primeiros precedentes europeus enfatizam a proteção, de um lado, do desenvolvimento da personalidade individual (dimensão privada) e, de outro, os impactos que a incerteza quanto ao uso dos dados pode provocar na participação democrática do indivíduo na sociedade, o que não se reduz à proteção da liberdade de expressão, daí a cunhagem de um novo direito fundamental derivado pelo tribunal constitucional alemão, no célebre julgamento contrário à Lei do Recenseamento da População de 1983 ("*Volkszählungsgesetz*")[6]: o direito à *autodeterminação informacional*.

Note-se que, naquele julgado, o uso dos dados pessoais coletados para fins de elaboração de políticas públicas foi considerado compatível com a autonomia informativa, dado que políticas gerais, anonimizadas, não tocam qualquer personalidade individual. Apenas um dispositivo da legislação foi declarado inconstitucional, pois autorizava o uso dos dados para alocação de alunos, identificáveis, nas escolas municipais.[7]

Houve, desse modo, o reconhecimento pelo tribunal que, para o livre desenvolvimento da personalidade, é necessária a participação ativa do indivíduo na sua própria representação na esfera informacional,[8] ou seja, uma prerrogativa individual em relação a sua autoapresentação (*Selbstdarstellung*)[9] na sociedade da informação. Isso significa permitir a *participação individual na determinação do fluxo adequado da informação*, na medida em que esta afete a projeção da personalidade na esfera pública, o que é

---

[3] VON LEWINSKI, Kai. Die Matrix des Datenschutzes: besichtigung und ordnung eines begriffsfeldes. Mohr Siebeck, 2014., p. 4-5.

[4] Gabriele Britz, *Informationelle Selbstbestimmung zwischen rechtswissenschaftlicher Grundsatzkritik und Beharren des Bundesverfassungsgerichts*, in: W. Hoffmann-Riem (Hg.), Offene Rechtswissenschaft, Tübingen 2010, p. 561-596.

[5] Como coloca Hans Peter Bull, *Sinn und Unsinn des Datenschutzes*, Tübingen 2015, p. 27 ss, protege-se a pessoa, ou a personalidade individual contra os efeitos do uso ilegítimo da informação na esfera pública.

[6] BVerfG, Urteil des Ersten Senats vom 15. Dezember 1983 - 1 BvR 209/83.

[7] ABRUSIO, Juliana; MARANHÃO, Juliano; CAMPOS, Ricardo. Proteção de dados pessoais no STF e o papel do IBGE. Disponível em: https://www.irib.org.br/noticias/detalhes/artigo-undefined-conjur-a-protecao-de-dados-pessoais-no-stf-e-o-papel-do-ibge-undefined-por-juliano-maranhao-ricardo-campos-e-juliana-abrusio. Acesso em: 01.02.2025.

[8] WESTIN, Alan F. Privacy and freedom. Washington and Lee Law Review, v. 25, n. 1, p. 166, 1968.

[9] VESTING, Thomas. Freie Entfaltung durch Selbstdarstellung: Eine Rekonstruktion des allgemeinen Persönlichkeitsrechts aus Art. 2 Abs. 1 GG. 2008; Gabriele Britz, Freie Entfaltung durch Selbstdarstellung, Tübingen (Mohr Siebeck) 2007, 93 S., Kritische Justiz 4 (2008), S. 473-475.

determinado contextualmente (qual tipo de informação, em relação a qual agente, em qual contexto, para quais fins, pode ser de que forma transmitida).[10]

Portanto, o cerne do problema está no controle da finalidade da coleta[11] e a quem as informações são destinadas (poder do destinatário de aumentar ou não a interferência estatal na formação da decisão individual), sempre com o propósito de proteger o indivíduo contra ameaças à projeção de sua personalidade na infosfera. São estes os vetores para a avaliação da constitucionalidade do processamento de dados e não propriamente a natureza e a quantidade dos dados pessoais compartilhados.[12] É exatamente tal perspectiva de realização da personalidade no âmbito social o cerne da proteção de dados pessoais, como destacado por Hoffman-Riem, [13] ao colocar que "*o direito à autodeterminação informacional é, em consequência, não um direito de defesa privatístico do indivíduo que se põe a parte da sociedade, mas objetiva possibilitar a cada um uma participação em processos de comunicação*", de modo que "*a liberdade em comum – não pode ser orientada para um conceito limitador da proteção à expansão egocêntrica, mas deve ser entendida como o exercício da liberdade em reciprocidade*."

Esse destaque de Hoffman-Riem decorre do próprio texto do famoso julgado alemão, onde se lê: "*o direito à 'autodeterminação informacional' não é, contudo, garantido sem limitação. Não oferece ao indivíduo controle absoluto ou ilimitado sobre os 'seus' dados pessoais; em vez disso, o indivíduo desenvolve sua personalidade dentro da comunidade social e depende da comunicação com os outros*" (BVerfGE 27, p. 15, tradução livre).

Nas duas dimensões de proteção- (i) a democrática e (ii) de livre desenvolvimento da personalidade- a garantia de defesa, que justifica a intervenção do Estado, pressupõe a ameaça à personalidade individual.

Na dimensão democrática, o desconhecimento quanto à existência do tratamento e quanto a quem o realiza, retira do sujeito a oportunidade de avaliar as consequências do seu comportamento, bem como as reações dos seus interlocutores na comunicação, o que inibe a ação individual. Com isso, a aplicação indiscriminada de informações derivada de seu processamento descontrolado colocaria em risco o funcionamento de *"uma comunidade democrática livre baseada na capacidade dos seus cidadãos de agir e participar"*. Ao passo que a dimensão da personalidade pode ser afetada quando os dados *"podem ser (combinados) com outras recolhas de dados para formar uma imagem pessoal*

[10] Marion Alberns, *Informationelle Selbstbestimmung*. Baden-Baden 2005, p. 86 ss.
[11] SCHLINK, Bernhard. Die Amtshilfe: Ein Beitrag zu einer Lehre von der Gewaltenteilung in der Verwaltung. Berlim: Duncker & Humblot, 1982, p. 192.
[12] ALBERS, Marion. Informationelle selbstbestimmung. Baden-Baden: Nomos, 20052005, p. 212.
[13] HOFFMANN-RIEM, Wolfgang. Rechtliche Rahmenbedingungen, em Der neue Datenschutz, Helmut Bäumler (org.) Neuwied/Kriftel, Luchterhand, 1998, p. 13.

parcial ou amplamente completa sem que o titular dos dados possa controlar suficientemente a sua exatidão e utilização"[14].

Desse modo, não se pode perder de vista, na defesa ou garantia da autodeterminação informacional, ou do direito fundamental à proteção de dados pessoais, em primeiro lugar, se o propósito e efeito concreto do tratamento impacta determinada personalidade individual e, em segundo lugar, se esse impacto afeta a *infosfera* como um espaço em comum, o que implica a consideração e ponderação de valores além da privacidade, como a não-discriminação, a transparência e a democracia.

Como é natural, a tradução legislativa concentrou-se nos meios para se atingir tais finalidades almejadas pela proteção de dados pessoais, o que sobrelevou obrigações de *accountability* com respeito a procedimentos de governança de dados, a serem observados em cada etapa de tratamento, incluindo a verificação de sua base legal.

2.2. Transição para a Legislação com Obrigações Procedimentais: *Accountability*

A evolução do direito europeu de proteção de dados passou da ênfase jurisprudencial na preservação da autodeterminação informacional, como garantia substantiva, para o modelo de responsabilização procedimental (*accountability*).[15] A transição para a abordagem procedimental ocorre de maneira gradual e reflete a mudança na relação entre titulares de dados, os próprios dados e os controladores.

Com o crescente avanço da informática e da digitalização, o volume e a complexidade das operações de tratamento cresceram significativamente. Percebeu-se que o indivíduo, isoladamente, não possuía as ferramentas e o conhecimento necessários para gerenciar todos os riscos associados ao processamento de dados e arcar com o ônus de monitorar práticas de tratamento dos agentes de tratamento com os quais se relaciona.

Diante dessa limitação, a legislação europeia foi gradualmente incorporando regras, instituições e mecanismos de fiscalização, que exigem um compromisso contínuo por parte dos controladores de dados,[16] criando-se a ideia de demonstrar objetivamente o cumprimento normativo por meio da adoção de procedimentos de governança e

[14] ALEMANHA. Tribunal Constitucional Federal. Decisão de 15 de dezembro de 1983. BVerfGE 65, 1 (42).
[15] UNIÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016 (GDPR). Artigo 5(2).
[16] UNIÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016 (GDPR). Artigos 4(7) e 24.

transparência perante os titulares. Nesse processo, a Diretiva 95/46/CE[17], apesar de não vinculante, já indicava a necessidade de se estruturar mecanismos internos para dar maior solidez à proteção de dados,[18] mas foi o GDPR (Regulamento 2016/679)[19] que deu o passo definitivo rumo a uma abordagem robusta de governança e verificação.[20]

Em vigor desde 2018, o GDPR tornou a *accountability* um princípio fundamental da regulação, determinando que o controlador não só seja responsável pelo cumprimento dos princípios de proteção de dados, mas seja também capaz de comprovar essa conformidade. Assim, o cumprimento da finalidade de proteção da autodeterminação informacional (perspectiva dos efeitos e propósitos) passa a ter menor peso do que o cumprimento dos procedimentos (perspectiva condicional), ou seja, a conformidade tem a ver com a antecipação de riscos, correção, mitigação e documentação das iniciativas. Isso passa por aspectos procedimentais como treinamento de funcionários, revisões constantes de procedimentos e adoção de padrões de segurança técnicos e organizacionais para mitigação de riscos.

Assim, o direito europeu partiu de uma abordagem jurisprudencial consubstanciada na proteção de um direito fundamental do cidadão, para a incorporação de mecanismos de fiscalização que transferiram a responsabilidade para os controladores, em termos de prestação procedimental de contas, o que acabou por consagrar, na GDPR, a chamada abordagem baseada em riscos.

Essa mudança de perspectiva, em que o foco se desloca do sujeito do dado (titular) e a proteção da sua personalidade (autodeterminação) para a gestão procedimental de riscos pelo controlador (*accountability*), decorreu da dificuldade do exercício concreto de direitos pelo titular em uma sociedade que assistiu à digitalização de tudo. Porém, não se pode desprezar que acabou por deslocar o efetivo alvo da proteção. O foco na gestão de riscos pelo controlador enfatizou a preocupação com as etapas de tratamento, independentemente do resultado, ao passo que a abordagem baseada em direitos

---

[17] UNIÃO EUROPEIA. Diretiva 95/46/CE do Parlamento Europeu e do Conselho, de 24 de outubro de 1995. Relativa à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados. Jornal Oficial das Comunidades Europeias, L 281, p. 31-50, 23 nov. 1995. Disponível em: https://eur-lex.europa.eu/eli/dir/1995/46/oj.

[18] ALHADEFF, Joseph; VAN ALSENOY, Brendan; DUMORTIER, Jos. The accountability principle in data protection regulation: origin, development and future directions. In: **Managing privacy through accountability**. London: Palgrave Macmillan UK, 2012. p. 49-82.

[19] UNIÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016 (GDPR). Artigo 5(2).

[20] QUELLE, Claudia. Enhancing compliance under the general data protection regulation: the risky upshot of the accountability-and risk-based approach. **European Journal of Risk Regulation**, v. 9, n. 3, p. 502-526, 2018.

enfatizava o impacto efetivo na projeção da personalidade individual com o *resultado* do tratamento.

A abordagem baseada em direitos naturalmente aproxima o bem protegido do meio de proteção uma vez que seu exercício se desenvolve pelo e para o titular de dados pessoais, ao passo que a abordagem baseada em riscos dissocia o bem protegido do meio de proteção, tornando os meios de proteção um fim em si mesmo para o controlador, o que, por consequência, modula a atuação da autoridade, já que esta última tem no comportamento do controlador o alvo de sua fiscalização.

Essa redução da proteção efetiva do bem jurídico ao mero cumprimento dos meios de gestão, portanto a dissociação entre o direito e o meio de sua efetivação, não passou desapercebida pela doutrina, que, próxima à formatação final da GDPR, já apontava para o risco de que a proteção de dados se torne *"um mero exercício gerencial desprovido de qualquer substância e, em última análise, sem significado"*.[21] Gellert, por exemplo, já procurava opor à abordagem baseada em riscos, a abordagem baseada no *princípio substantivo de precaução,* antevendo que cumprir os meios gerenciais procedimentais pode ser insuficiente ou mesmo contraproducente para a garantia do livre desenvolvimento da personalidade individual.[22]

A responsabilização procedimental na abordagem baseada em riscos firmou-se como modelo de referência na abordagem legislativa contemporânea na área de tecnologia e em campos nos quais se exige controle e mitigação de riscos que não podem ser completamente eliminados pela natureza da atividade. Porém, o questionamento inicial sobre o formalismo na implementação da abordagem baseada em riscos ganha força na evolução doutrinária europeia em termos de *precaução substantiva*, ao se questionar que o objetivo da legislação não poderia ser somente alcançar "*um nível de riscos aceitável para a sociedade*", conforme parâmetros previstos em lei, mas sim *assegurar substantivamente direitos e promover benefícios sociais*.

2.3. Resgate Europeu do Valor Social da Privacidade e sua Ponderação com Interesses Públicos

A abordagem baseada no princípio de *precaução substantiva* também se baseia nos procedimentos de gerenciamento de risco no tratamento de dados, mas exige a constante postura crítica pelo intérprete e aplicador, no sentido de considerar o *efetivo*

[21] GELLERT, Raphaël. Data protection: a risk regulation? Between the risk management of everything and the precautionary alternative. Int'l Data Priv. L., v. 5, p. 3, 2015.
[22] *Ibidem.*

valor social da privacidade e da proteção de dados, o que dá abertura para diferentes mecanismos de implementação, com foco nos resultados da operação de tratamento.[23]

Essa dimensão social da privacidade como valor a ser protegido no ambiente digital foi delimitada com clareza no julgado do Tribunal Constitucional Alemão que reconheceu o direito fundamental chamado de "IT-Grundrecht", ou o *"direito fundamental à confidencialidade e integridade dos sistemas de tecnologia da informação"*,[24] com base no qual foi considerada inconstitucional a implantação de sistema de acesso policial privilegiado a plataformas para trocas de e-mail para investigações a suspeitos. Importante notar que, naquele julgado, o argumento oposto à garantia de segurança pública não foi propriamente o direito à proteção da personalidade individual de determinado investigado, mas sim a *confiança coletiva em uma plataforma tecnológica*, que seria abalada em função de vulnerabilidade introduzida pelo Estado para fins de vigília, com consequências diretas sobre a liberdade de expressão e a democracia.

Questão semelhante vem sendo debatida no STF sobre o bloqueio do Whatsapp (a ADPF 403 e a ADI 5527), em que os Ministros Edson Fachin e Rosa Weber proferiram votos no sentido de inconstitucionalidade de qualquer interpretação do inciso II do art. 7º e inciso III do art. 12 da Lei 12.965/2014 que leve à exigibilidade de acesso excepcional a conteúdo de mensagem criptografada ponta-a-ponta, uma vez que, defendem os ministros, tal medida poderia enfraquecer a confidencialidade de aplicações na internet. Note-se que não se protege aqui o direito subjetivo do investigado à privacidade, mas a confiança da coletividade na tecnologia, ou seja a privacidade em grupo ou seu valor social.[25]

Por sua vez, considerar o valor social da privacidade e proteção de dados como bens protegidos e assegurados constitucionalmente traz abertura para, de um lado, se considerar a privacidade como um valor coletivo (*group privacy*)[26] e, de outro, se considerar outros valores socialmente relevantes para o objetivo específico do tratamento de dados pessoais.[27] Esse último aspecto é destacado na abertura do

---

[23] BENNET, C. J.; RAAB, Charles D. The governance of Privacy. Policy Instruments in Global Perspective, MIT, 2006.

[24] *Online-Durchsuchung, BVerfGE 120, 274 (2008) e BVerfGE 141, 220 (2010).*

[25] MARANHÃO, J. O STF e a chave para os direitos no mundo pós-pandemia. Valor Econômico - O Globo, São Paulo, 12 maio 2021.

[26] *Mittelstadt, B. (2017). From Individual to Group Privacy in Big Data Analytics. Philo- sophy & Technology, 30(4), 475-494. https://doi.org/10.1007/s13347-017- 0253-7.*

[27] *Report of the Data Ethics Commission of the Federal Government. (2019). Federal Ministry of the Interior, Building and Community and Federal Ministry of Justice and Consumer Protection. https://datenethikkommission.de/wp-content/uploads/191128_ DEK_Gutachten_bf_b.pdf.*

"*Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*" do *European Data Protection Board* - EDPB[28]:

> "By protecting the fundamental right to data protection, GDPR supports these opportunities and promotes other EU fundamental rights, including the right to freedom of thought, expression and information, the right to education or the freedom to conduct a business. In this way, GDPR is a legal framework that encourages responsible innovation"

Ou seja, a proteção de dados é vista não como um fim em si, reduzida aos procedimentos legais, mas como veículo para promoção de outros direitos fundamentais e sociais, dentro de uma análise ética da inovação.

Nesse sentido, Augsberg e Ulmstein,[29] em artigo escrito durante a pandemia de COVID-19 sobre a soberania de dados no campo da saúde, defende que o uso de IA e Big Data Analytics em pesquisa, diagnóstico e tratamento médico demanda gestão de riscos atenta aos impactos efetivos do processamento sobre a saúde e a privacidade, de uma perspectiva não individual, mas coletiva.

Observam que, na era de Big Data e da IA, *"uma justa economia de dados não pode focar em um ponto de dado específico ou conjuntos de dados individuais, para assegurar a liberdade",* devendo-se pensar de uma perspectiva coletiva da privacidade, ou seja, em mecanismos de governança que propiciem maior privacidade em geral, ao lado de outros benefícios sociais.

De um lado, argumentam, diversos tipos de processamento de dados em massa por sistemas de IA podem levar a uma série de inferências relevantes, não antecipadas. De outro, novas tecnologias de gestão de dados permitem o exercício dinâmico de prerrogativas pelos titulares, o que significa ampliar a possibilidade de participação e reclamações de modo dinâmico.

Na visão tradicional procedimental, seria dever do controlador avaliar cada grupo de dados e sua referência a indivíduo determinável, para, no limite, na ausência de outras bases legais, buscar consentimento informado sobre seu processamento, seja qual for o propósito do tratamento. Já em uma abordagem de precaução substantiva, sensível ao impacto coletivo na privacidade em grupo e sobre a saúde, adotam-se medidas efetivas e adequadas a depender do propósito final ou do resultado do tratamento de dados. Por exemplo, ao tratar de consentimento, os autores sustentam que o uso de Big Data ou IA

---

[28] Disponível em: https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-ar t-64/opinion-282024-certain-data-protection-aspects_en.

[29] AUGSBERG, S.; ULMSTEIN, U. Requisitos de Consentimento Modificados: O Direito de Proteção de Dados Pode Aprender Com o Direito Da Saúde. In: CAMPOS, R.; ABBOUD, G.; NERY JR., N. (Org.). Proteção de dados e regulação. São Paulo: Thomson Reuters, 2020.

no corpo de dados para definir o tratamento de determinado indivíduo deve partir do consentimento informado ou restrito como medida de precaução substantiva, já o mesmo princípio de precaução aponta para medidas mais flexíveis de consentimento alargado ou consentimento dinâmico, quando os dados são utilizados para pesquisa e mesmo a utilização sem consentimento quando tratamos de utilização e pesquisa para promoção de saúde pública.[30] A própria tecnologia tem permitido novos mecanismos de controle por parte dos sujeitos de dados, como o consentimento dinâmico ou *data trusts*, permitindo o maior desenvolvimento da economia de dados e uso dos mesmos para o desenvolvimento tecnológico e promoção de valores coletivos, como a proteção à saúde.

O contexto da pandemia e a necessidade de tratamento de dados para a promoção da saúde pública foi um ponto de inflexão a respeito da necessidade de maior flexibilidade interpretativa da legislação de proteção de dados europeia, com maior foco no efetivo impacto do tratamento sobre a individualidade *versus* seu impacto sobre a saúde pública, ao lidar com o processamento de dados para combate à COVID-19 e monitoramento da quarentena.[31] Diferentes países europeus recorreram a soluções tecnológicas de rastreamento de contatos (*contact tracing apps*) e de coleta de dados de pacientes para fins epidemiológicos. Embora tais iniciativas pudessem suscitar questionamentos sobre adequação aos procedimentos legais de tratamento autorizado de dados, autoridades adotaram interpretação finalística que reconhecia a legitimidade do uso de dados pessoais para conter a disseminação do vírus.[32]

Dentre os esforços para o combate à pandemia estava o emprego de inteligência artificial, o que resgatou questionamentos já colocados pelos europeus sobre os possíveis entraves que a interpretação inflexível da legislação de proteção de dados poderia trazer ao desenvolvimento tecnológico naquele continente, seja quanto à disponibilidade de

---

[30] AUGSBERG, S.; ULMSTEIN, U. Requisitos de Consentimento Modificados: O Direito de Proteção de Dados Pode Aprender Com o Direito Da Saúde. In: CAMPOS, R.; ABBOUD, G.; NERY JR., N. (Org.). Proteção de dados e regulação. São Paulo: Thomson Reuters, 2020.

[31] INFORMATION COMMISSIONER'S OFFICE. Apple and Google joint initiative on COVID-19 contact tracing technology. Wilmslow: ICO, 2020. Disponível em: https://ico.org.uk/media/about-the-ico/documents/2617653/apple-google-api-opinion-final-april-2020.pdf; EUROPEAN COMMISSION. Digital contact tracing: learning from the experiences of European countries. Brussels: European Commission, 2023. Disponível em: https://commission.europa.eu/system/files/2023-02/DigitalContactTracingStudy.pdf; AUSTRALIAN INFORMATION COMMISSIONER. *Privacy update on the COVIDSafe App*. Disponível em: https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/covid-19/privacy-update-on-the-covidsafe-app.

[32] PORTUGAL. Comissão Nacional de Proteção de Dados. Orientações sobre os tratamentos de dados pessoais de saúde regulados no Decreto n.º 8/2020 https://www.cnpd.pt/media/1bbppegs/orientações_decreto_8_2020.pdf; UNITED KINGDOM. Information Commissioner's Office. COVID-19 and information rights: reflections and lessons learnt from the Information Commissioner November 2021 https://ico.org.uk/media/about-the-ico/documents/4019157/covid-19-report.pdf.

dados para pesquisa e desenvolvimento, seja perante a capacidade de inferência dos sistemas de inteligência artificial. [33] Nessa linha, pesquisa conduzida a pedido do Parlamento Europeu concluiu que a aplicação da legislação de proteção de dados não limitaria o avanço da inteligência artificial, desde que a sua interpretação se volte para o propósito ou resultados das inferências produzidas pela ferramenta sob análise.[34] Dessa forma, os direitos dos titulares deveriam estar focados principalmente na transparência sobre como seus dados são inicialmente coletados e tratados, mas não como óbice ao aproveitamento de aplicações finais e de seus resultados, principalmente aquelas que sejam despersonalizadas ou tragam benefícios sociais relevantes.

Assim, em vez de uma interpretação rígida quanto aos procedimentos de gerenciamento de riscos, no sentido exigir que o controlador busque uma base legal e, no limite, o consentimento, para cada grupo individual de dados, propõe-se abordagem que avalie a razoabilidade, transparência e a proporcionalidade do tratamento como um todo, o que significa olhar para *propósito* e para o *resultado* da aplicação de IA sobre a privacidade individual e sobre o sujeito de dados pessoais. [35]

Ou seja, olha-se para o *output* da IA com o objetivo de se avaliar a razoabilidade do uso do dado de entrada, e não para a pertinência individual do uso de cada dado no tratamento correspondente à etapa de *input* ou mesmo correspondente ao treinamento do sistema. Trata-se de perspectiva que olha a privacidade como valor social ao lado de outros valores que informam a ética de IA (ver abaixo), compondo não apenas o direito a proteção de dados, mas o direito a inferências razoáveis pelos sistemas de IA perante os indivíduos afetados, conceito que tem na transparência seu pilar fundamental.[36] Com isso, tem-se um novo direcionamento interpretativo da legislação de proteção de dados, com base no art. 6º, inc. VIII da LGPD, que traz o princípio de *prevenção de danos*, que se desloca, de um lado, do *procedimento formal* de mitigação de risco para a *precaução substantiva* na proteção do bem jurídico, [37] e, de outro, da *visão egocentrada no titular* para a *proteção do valor social da privacidade*.

---

[33] INFORMATION COMMISSIONER'S OFFICE. *Guidance on AI and Data Protection*. Wilmslow: ICO, 2023. Disponível em: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/.

[34] SARTOR, Giovanni; LAGIOIA, Francesca. *The impact of the General Data Protection Regulation (GDPR) on artificial intelligence*. Brussels: European Parliament, 2020. DOI: 10.2861/293.

[35] EUROPEAN DATA PROTECTION BOARD. *Opinion 28/2024 on certain data protection aspects of the proposed European Health Data Space*. Brussels: EDPB, 2024. Disponível em: https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en.

[36] *WACHTER, Sandra; MITTELSTADT, Brent. A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Colum. Bus. L. Rev., p. 494, 2019.*

[37] VON GRAFENSTEIN, Maximilian. The principle of purpose limitation in data protection laws. Nomos Verlagsgesellschaft mbH & Co. KG, 2018.Disponível em: https://www.nomos-

Ao considerar procedimentos de gerenciamento de riscos como ponto de partida e não como fim em si, a precaução substantiva significa não só que procedimentos ulteriores aos previstos em lei possam se tornar deveres quando a personalidade individual estiver em risco, mas também que se esvaziam os deveres procedimentais quando não há impacto relevante sobre a personalidade individual como resultado do tratamento de dados.

Para se compreender o significado da privacidade e da proteção de dados como valor social, no contexto do tratamento para desenvolvimento e emprego de sistemas de IA, é importante entender como a privacidade é incorporada dentro da ética e da governança do uso confiável e responsável de sistemas de IA, objeto do próximo item.

elibrary.de/de/10.5771/9783845290843/the-principle-of-purpose-limitation-in-data-protection-aws.

3. ÉTICA E REGULAÇÃO DA IA BASEADA EM RISCOS

O debate sobre o desenvolvimento e uso responsável da Inteligência Artificial, diferentemente da criação jurisprudencial da proteção de dados como um direito à autodeterminação informacional, já nasceu informado pela abordagem baseada em riscos, ou seja, construída, no debate europeu, não sobre direitos dos usuários ou pessoas impactadas, mas da perspectiva de procedimentos de governança a serem adotados por desenvolvedores, fornecedores e operadores de tais sistemas.

Porém, o debate em torno da ética de IA tem-se estruturado a partir da consideração dos impactos dos sistemas a direitos fundamentais ou a bens coletivos, de modo que as medidas de mitigação de riscos são desenvolvidas com base em categorização dos tipos e propósitos de aplicações de IA. O AI Act Europeu aborda a regulação da IA como tarefa de criação de requisitos de *segurança de produto* (*product safety*), em termos de seus impactos sobre direitos. Assim, o AI Act é, ao mesmo tempo, sobre *segurança de produto* e sobre *direitos*, pois a segurança é definida em termos de desenvolvimento e uso responsável de IA, o que significa uso atento aos riscos e que incorpore os correspondentes mecanismos para sua mitigação e proteção de potenciais afetados.

Isso vale – e talvez mesmo em grau maior – para o debate legislativo brasileiro em torno do PL 2338/2023, que dá particular ênfase, de um lado, à proteção efetiva de direitos e, de outro à setorialização e autorregulação, como mecanismos para aumentar a eficácia de medidas de controle e de governança, considerando os propósitos de sistemas de IA em domínios específicos de aplicação.

O objetivo desta Seção 3 é evidenciar (i) como a abordagem baseada em riscos na regulação da IA está imbuída da preocupação com a proteção efetiva de direitos fundamentais (precaução substantiva), de modo que a integração entre a legislação de proteção de dados e a regulação de IA devem partir da consideração dos propósitos e efeitos da aplicação do sistema; e (ii) como a *precaução substantiva* em relação a proteção de dados pessoais pode informar a solução de conflitos com os demais valores protegidos pela IA responsável. Para tanto, primeiro trataremos da ética de IA e as iniciativas europeia e brasileiras de regulação e, em seguida, traremos alguns exemplos sobre os confrontos, a serem sopesados, entre o valor social da privacidade e outros valores promovidos no âmbito da IA Responsável.

3.1 Regulação da IA como Regulação de seus Impactos

Iniciativas de regulação da inteligência artificial tem empregado a abordagem baseada em riscos, impondo conjuntos de medidas obrigatórias de governança apropriadas ao

grau de risco e gravidade dos impactos sobre direitos fundamentais. A legislação mais completa em vigor, nesse sentido, é o AI Act da União Europeia. Tal legislação foi concebida a partir de uma estrutura híbrida entre a abordagem procedimental e a proteção de direitos fundamentais. Seu objetivo declarado é promover sistemas de IA responsáveis, portanto sistemas compatíveis com o exercício de direitos fundamentais, impondo obrigações de procedimentos técnicos e organizacionais para garantir a segurança de produto (*product safety*).[38]

Na verdade, como destacam Almada e Petit, o AI Act europeu une duas tradições regulatórias: a tradição de segurança de produto e a de proteção de direitos fundamentais. A linguagem de proteção a direitos fundamentais e mesmo impactos a bens coletivos está presente desde o Relatório do High Level Expert Group (AI HLEG)[39], passando pela avaliação de impacto regulatório[40], o preâmbulo justificador do diploma e permeando todo o texto da lei, em particular, na formulação da última versão. Já se entreveem, porém, alguns desafios, como a necessidade de adaptação constante de regras fixas de governança aos avanços da tecnologia (*pacing problem*), a tendência a se aplicar a lei pelas lentes da segurança de produto (*regulatory lens*), ou seja, seguir a visão procedimental e não aquela baseada na garantia efetiva de direitos fundamentais e a herança regulatória europeia que circunda o AI Act, também estruturada em segurança de produto (*path dependence*).[41] O que se propõe neste alerta, semelhante à crítica de Gellert quanto ao procedimentalismo da GDPR, pouco antes de sua entrada em vigor, é justamente uma interpretação voltada para os instrumentos de governança informada e modulada, portanto, flexível, em relação ao propósito de efetiva proteção dos direitos fundamentais e valores sociais implicados, ou seja, não uma governança gerencial, mas uma proteção substantiva.

Como a IA é tratada pela legislação como produto, o AI Act estabelece os requisitos a serem atendidos para que os mesmos possam ser colocados em circulação no mercado

---

[38] ALMADA, Marco; PETIT, Nicolas. The EU AI Act: Between the rock of product safety and the hard place of fundamental rights. Common market law review, v. 62, n. 1, 2025.

[39] Os relatórios produzidos pelo AI HLEG podem ser acessados no seguinte link, sendo especialmente interessante o intitulado "Policy and Investment Recommendations for Trustworthy AI": https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai. Neste documento, publicado em meados de 2019, o grupo apresentou 33 recomendações para orientar a IA confiável em direção à sustentabilidade, crescimento, competitividade e inclusão, trazendo considerações relativas a direitos fundamentais.

[40] EUROPEAN COMMISSION. *Impact assessment of the regulation on artificial intelligence*. Bruxelas, 21 abr. 2021. Disponível em: https://digital-strategy.ec.europa.eu/en/library /impact-assessment-regulation-artificial-intelligence.

[41] Almada, M;Petit, N. The AI Act: between the rock of product safety and the hard place of fundamental rights. Common Market Law Review 62: 85–120, 2025.

europeu.[42] Tais requisitos, dentro da abordagem considerada "baseada em riscos", que informa essa regulação, variam conforme a categorização de risco do produto e a regulação assume a possibilidade de se definir exaustivamente todas as aplicações nessas categorias conforme os potenciais efeitos adversos sobre direitos fundamentais. Assim, ao considerar como "*falhas de produto*" os potenciais impactos sobre direitos, o AI Act incorpora, de modo oblíquo, a abordagem baseada em direitos em uma estrutura regulatória procedimental. [43]

Ao se voltar para os riscos em termos de impactos sobre direitos,[44] a abordagem procedimental do AI Act dá abertura para interpretação centrada na promoção dos direitos fundamentais relativos ao emprego de sistemas de IA e dos impactos concretos da tecnologia na sociedade, permitindo-se, assim, maior flexibilidade por meio da consideração dos propósitos da aplicação. Ou seja, os procedimentos de governança que garantem a segurança do produto traduzem-se nas melhores práticas para *efetivamente promover os benefícios almejados* e, ao mesmo tempo, mitigar riscos potenciais para a sociedade e impactos negativos sobre direitos individuais, o que é congênere à ideia de *precaução substantiva*.

No Brasil, o debate em torno do Projeto de Lei 2338/2023 e a evolução de sua redação também apontam para um desenho regulatório de *precaução substantiva*. Apesar do texto legal ser estruturado em medidas de governança a serem adotadas conforme o risco da aplicação, típico da abordagem procedimental, dois aspectos do debate brasileiro, que se sobrelevam mesmo em relação ao europeu, merecem destaque: o elenco de direitos fundamentais e a abordagem setorial.

No que concerne a direitos fundamentais, a versão original do PL 2338/23 colocava a proteção a direitos fundamentais em primazia, na medida em que abordava medidas de governança como deveres em relação a direitos de pessoas afetadas por sistemas de IA. Ou seja, partia da proteção a direitos fundamentais, por essência, fazendo com que as medidas de governança surgissem como meios para sua concretização. Ocorre que tal abordagem atraía para o P. Judiciário a competência para fiscalização, na medida que criava pretensões de cidadãos a medidas de governança perante fornecedores, o que poderia resultar em insegurança jurídica. Gradualmente, o elenco de "direitos à IA Responsável", nas versões sucessivas, foi sendo esvaziado, centrando-se a regulação e

[42] EBERS, Martin. Standardizing AI-The Case of the European Commission's Proposal for an Artificial Intelligence Act. The Cambridge handbook of artificial intelligence: global perspectives on law and ethics, 2021.
[43] ALMADA, Marco; PETIT, Nicolas. The EU AI Act: Between the rock of product safety and the hard place of fundamental rights. Common market law review, v. 62, n. 1, 2025.
[44] DUFOUR, Raimond et al. AI or more? A risk-based approach to a technology-based society. Oxford Business Law Blog, v. 2021, n. September 16, 2021.

fiscalização de medidas de governança como obrigações dos desenvolvedores, fornecedores e aplicadores perante autoridades administrativas. Apesar de ter restado apenas a previsão de direitos ligados ao princípio de transparência e supervisão humana, o PL 2338/23 continua com a afirmação da instrumentalidade das medidas de governança e mitigação de riscos ali previstas, isto é, como meios para a promoção efetiva e concreta dos valores que compõem a responsabilidade no desenvolvimento e uso da IA, dentre eles, a proteção de dados pessoais.

Além disso, durante na evolução do debate legislativo, o Projeto de Lei 2338/2023 reformulou sua abordagem inicialmente transversal por mecanismos mais específicos de fiscalização setorial e de autorregulação. Esses incluem autorregulação e códigos de conduta validados pelas autoridades competentes. Para aprimorar ainda mais a responsabilidade da indústria, a legislação promove a autorregulação e programas de certificação, alinhando medidas regulatórias formais à adesão às melhores práticas e padrões éticos no desenvolvimento e uso de IA. Tais alterações são indicativas da preocupação com a concretização substantiva de direitos fundamentais no desenvolvimento e emprego da IA, na medida em que a abordagem setorial abre espaço para maior flexibilidade e especificidade para avaliar e enfrentar os riscos específicos da IA em cada setor econômico, permitindo a adoção de soluções mais efetivas conforme o estado da arte do desenvolvimento da IA em cada setor. Por seu turno, mecanismos de autorregulação permitem uma abordagem mais eficiente para resolver os desafios técnicos decorrentes da aplicação das normas, favorecendo respostas mais rápidas, flexíveis e aderentes às realidades técnicas enfrentadas pelos próprios agentes econômicos.

Trata-se de linha bastante próxima à defendida por Hoffman-Riem como meio de regulação da IA responsável,[45] com a construção de cooperação público privada, mecanismos de autorregulação, certificação e especificação setorial de melhores práticas, capazes de proteger os direitos sociais e individuais envolvidos, na qual o Estado assume um papel garantidor perante a regulação e fiscalização privada e voluntária (Gewahrleistungsstaat)[46], como meio de proteger e promover a confiança do cidadão na tecnologia, o que vê como um desdobramento do direito fundamental à tecnologia *IT-Grundrecht*, mencionado no item 2.1.

[45] HOFFMANN-RIEM, W. Artificial Intelligence as a Challenge to Law and Regulation. em WISCHMEYER, T. e RADMACHER, T. Regulating Artificial Intelligence, Springer, 2020, p. 75 e ss.
[46] RUGE, R. Die Gewahrlestungsverantwortung des Staates und der Regulatory State. Duncker & Humboldt, Berlin, 2004.

3.2. Dilemas Éticos e a *Precaução Substantiva* quanto à Proteção de Dados Pessoais na Governança Responsável do Desenvolvimento e Uso da IA

Como visto, a valoração ética de sistemas de IA, usualmente referida pelos conceitos de "IA confiável" ou "IA responsável", [47] é construída a partir da consideração dos benefícios dos sistemas de IA- nos mais diferentes campos, por meio da potencialização do conhecimento e aumento de eficiência e produtividade- frente aos riscos inerentes à tecnologia, como os riscos de efeitos danosos causados por erros, efeitos discriminatórios, violações à privacidade e proteção de dados e opacidade em relação ao uso ou aos critérios que determinam os *outputs* dos sistemas.

É em torno desses riscos que se elencam os valores a serem promovidos, desde a concepção, desenvolvimento, emprego e o monitoramento, tais como: *transparência* (deve estar claro para o usuário que ele interage com um sistema artificial), *explicabilidade* (divulgação de informações ao interessado que permitam ao usuário entender os critérios para tomada de decisão), *fairness* ou *não discriminação* (evitar que os sistemas incorporem vieses que possam ter como efeito direto ou indireto prejudicar grupos minorizados ou desfavorecidos), *confiabilidade* (desenvolver técnicas para elevar a acurácia dos sistemas e buscar mecanismos de supervisão humana para minimizar erros),[48] *não maleficência* (sistemas de IA não podem prejudicar humanos), *responsabilidade* (no sentido de prestação de contas e reparação de possíveis danos) e *privacidade/proteção de dados* (sistemas não devem ter por objeto ou efeito violar a privacidade ou tratar dados sem controle de finalidade).[49]

Mesmo que possa haver algum consenso quanto a valores abstratos na conceituação de *"desenvolvimento e uso responsável da IA"*, observa-se, nos diversos documentos acadêmicos ou de organizações governamentais ou não governamentais que trataram do tema, divergências quanto a seu alcance e implementação prática, em razão das diferentes soluções possíveis diante de um conflito entre os valores envolvidos.[50] Para lidar com esses desafios, diferentes modelos de governança para gestão de riscos da IA,

---

[47] EC HIGH LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE. Orientações éticas para uma IA de confiança. 2018. Disponível em: https://doi.org/10.2759/2686. Acesso em: 15 jul. 2024.
[48] MARANHÃO, Juliano. Atributos de confiabilidade e segurança na governança de sistemas de Inteligência Artificial. In: BIONI, Bruno R.; CUEVA, Ricardo V. B..; MENDES, Laura S.; ALVES, Fabrício M. (orgs.). Inteligência Artificial e Regulação. São Paulo: Editora Gen-Jurídico. No prelo.
[49] JOBIN, Anna; IENCA, Marcello; VAYENA, Effy. The global landscape of AI ethics guidelines. Nature Machine Intelligence, v. 1, n. 9, p. 389-399, 2019.
[50] JOBIN, Marcello Ienca and VAYENA, Effy, 'The Global Landscape of AI Ethics Guidelines' (2019) 1 Nature Machine Intelligence 389.

ou de certificação de qualidade, vem sendo propostos por entidades de referência,[51] com abordagem mais transversal ou setorial (com foco em determinados domínios de aplicação), de modo a delimitar métricas confiáveis e especificar medidas de implementação eficazes.[52]

Esse problema pode ser ilustrado em diferentes conflitos entre os valores que compõem o conceito de IA responsável, em particular em relação a temas chave em legislações de proteção de dados, como o consentimento do titular e quanto à exigência de explicabilidade de decisões automatizadas com base em perfilamento (*profiling*). Diante desse conflito, o princípio da precaução substantiva pode trazer soluções equilibradas, promovendo a garantia substantiva do direito fundamental à proteção de dados pessoais, sem limitar o desenvolvimento tecnológico.

Exemplo 1 – *Credit Scoring*

Exemplo de uso de IA envolvendo *trade-off* entre privacidade *versus* benefícios sociais trazidos pela confiabilidade (acurácia) dos sistemas está nos modelos, cada vez mais complexos, para pontuação de risco de inadimplência (*credit scoring*), com efeitos

[51] Ver, por exemplo: IEEE P2863™ Recommended Practice for Organizational Governance of Artificial Intelligence; IEEE P2894™ Guide for an Architectural Framework for Explainable Artificial Intelligence; IEEE Std 7000™ – 2021 – Model Process for Addressing Ethical Concerns During System Design; IEEE 7001™ – 2021 – Standards for Transparency of Autonomous Systems; IEEE P7003™ – Standard for Algorithmic Bias Considerations. ISO/IEC TS 4213:2022 (Assessment of machine learning classification performance), ISO/IEC 8183:2023 (Data life cycle framework), ISO/IEC 20546:2019 (Big data - Overview and vocabulary), ISO/IEC TR 20547-1:2020 (Big data reference architecture - Part 1: Framework and application process), ISO/IEC TR 20547-2:2018 (Big data reference architecture - Part 2: Use cases and derived requirements), ISO/IEC TR 20547-3:2020 (Big data reference architecture - Part 3: Reference architecture), ISO/IEC TR 20547-5:2018 (Big data reference architecture - Part 5: Standards roadmap), ISO/IEC 22989:2022 (Artificial intelligence concepts and terminology), ISO/IEC 23053:2022 (Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)), ISO/IEC 23894:2023 (Guidance on risk management), ISO/IEC TR 24027:2021 (Bias in AI systems and AI aided decision making), ISO/IEC TR 24028:2020 (Overview of trustworthiness in artificial intelligence), ISO/IEC TR 24029-1:2021 (Assessment of the robustness of neural networks - Part 1: Overview), ISO/IEC 24029-2:2023 (Part 2: Methodology for the use of formal methods), ISO/IEC TR 24030:2021 (Artificial intelligence (AI) - Use cases), ISO/IEC TR 24368:2022 (Overview of ethical and societal concerns), ISO/IEC TR 24372:2021 (Artificial intelligence (AI) - Overview of computational approaches for AI systems), ISO/IEC 24668:2022 (Process management framework for Big Data analytics), ISO/IEC 25059:2023 (Software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - Quality model for AI systems), ISO/IEC 38507:2022 (Governance of IT - Governance implications of the use of artificial intelligence by organisations).

[52] MARANHÃO, Juliano; NAVAS, João. Certificação como instrumento de regulação da Inteligência Artificial no AI Act. In: VAINZOF, Rony; GUTIERREZ, Andriei; GODINHO, Gustavo; KRASTINS, Alexandra(orgs.) Comentários ao EU AI Act: uma abordagem prática e teórica do Artificial Intelligence Act da União Europeia. 1ª ed. São Paulo: Thomson Reuters Brasil, 2024. Pp. 259-279.

significativos sobre o titular de dados, dada sua relevância para acesso a crédito.[53] Historicamente, a eficiência e confiabilidade dos sistemas de *credit scoring* tem aumentado quanto maior a quantidade e variedade de dados utilizados e maior a complexidade dos sistemas de IA empregados.[54] Essa maior escala e variedade de dados significa maior intensidade de uso de dados pessoais, seja para o treinamento dos modelos, seja para inferência quanto à classificação de risco individual. Ocorre que, apesar do impacto sobre personalidades individuais, quanto maior a confiabilidade do sistema menor o risco de crédito, o que propicia aumento de sua oferta e redução do *spread* bancário, com benefícios ao desenvolvimento econômico e à democratização do acesso a crédito, além de propiciar maior segurança ao próprio sistema financeiro, razão pela qual o emprego do estado da arte para controle de riscos traduz-se, inclusive, em obrigação regulatória (no Brasil, a Resolução 4557/2017 do Banco Central do Brasil, que dispõe da estrutura de gerenciamento de riscos para instituições financeiras).

A avaliação da adequação da prática de *credit scoring* da perspectiva do princípio precaução substantiva deve compreender, inicialmente, o impacto sobre a privacidade e à proteção de dados pessoais como bem coletivo e como bem individual. Tal prática, que emprega intensivamente sistemas de IA, envolve duas fases principais em relação ao processamento de dados: a fase de o desenvolvimento, em particular, do uso de dados pessoais para o treinamento de sistemas de IA e a fase de inferência, ou seja, de uso dos sistemas de inteligência artificial para classificação de determinado indivíduo nos índices de risco delineados.

Na fase de treinamento, o impacto sobre personalidades individuais é reduzido, pois o resultado do tratamento é um modelo com categorias de devedores, despersonalizados, de modo que se sobrepõe o benefício social e econômico de aumento da oferta e redução do custo do crédito. Aqui, o princípio de precaução substantiva aponta para a admissibilidade do tratamento, o que é fator relevante para o desenvolvimento e aperfeiçoamento da tecnologia, uma vez que podem ser realizados diversos testes e modelagens tentativas que sequer encontrarão aplicação, portanto, sequer terão perspectiva de efeito sobre qualquer personalidade individual.

[53] MARANHÃO, Juliano Souza; CAMPOS, Ricardo Resende. Proteção de dados de crédito na lei geral de proteção de dados. Direito Público, v. 16, n. 90, 2019.
[54] THE ECONOMIST. A brief history—and future—of credit scores. The Economist, [s.l.], 6 jul. 2019. Disponível em: https://www.economist.com/international/2019/07/06/a-brief-history-and -future-of-credit-scores.

Foi justamente em função dessa ponderação que a jurisprudência e posterior regulamentação na Europa reconheceram haver legítimo interesse como base legal para tratamento de dados pessoais para o fim de *credit scoring*.[55] Inclusive a disponibilização de dados por instituições financeiras para diferentes *bureaux* de crédito foi vista como medida de incentivo à competição que admitia o compartilhamento independente de consentimento.[56] É digno de nota também, que, na legislação brasileira, o propósito de proteção ao crédito é previsto como uma base legal própria (art. 7º inc. X da LGPD), o que reflete a inclinação do legislador brasileiro em direção à consideração dos resultados do tratamento de dados.[57]

Já na fase de inferência, o impacto sobre a personalidade individual é significativo, tendo em vista que o objetivo é realizar avaliação capaz resultar na recusa de acesso a bens. Aqui, a precaução substantiva aponta para a importância de medidas efetivas e garantia de direitos dos sujeitos de dados afetados quanto à qualidade, fidedignidade e base legal para emprego dos dados individuais usados para a avaliação.

A fase de inferência pode apresentar duas situações: (i) o índice de risco de inadimplência é base para uma decisão automatizada sobre a concessão de empréstimo ou de acesso a bem (ii) o resultado da classificação em índices de risco de inadimplência é apenas premissa para tomada de decisão humana.

No primeiro caso, aplicam-se imediatamente as garantias legais, como a necessidade de consentimento quanto à decisão automatizada no caso da legislação europeia (art. 22 GDPR), ou, no caso da legislação brasileira, os direitos à explicação, revisão e auditoria pela autoridade para detecção de possíveis traços discriminatórios (art. 20 e parágrafos da LGPD).

No segundo caso, apesar de não haver propriamente decisão integralmente automatizada, nos termos da lei, o princípio de *precaução substantiva* pode estender os

---

[55] MARANHÃO, Juliano Souza; CAMPOS, Ricardo Resende. Op. Cit.

[56] Em um caso decidido pelo tribunal de segunda instância de Berlim, em 2013, sobre a necessidade ou não de consentimento como base legal para o tratamento e compartilhamento de dados referentes à capacidade de crédito e avaliação de crédito por pontuação, o tribunal, baseando-se no § 29 BDSG (deutsche Bundesdatenschutzgesetz – Lei Federal Alemã de Proteção de Dados), excluiu a necessidade de consentimento. Segundo a argumentação do tribunal, a coleta, armazenamento ou tratamento de dados pessoais para fins de transmissão é permitida se houver interesse legítimo. O caso concreto referia-se a uma agência de crédito, que, pelo julgado, estaria autorizada a recolher informações sobre a concessão da quitação residual da dívida junto do registo de devedores e a armazená-las durante três anos para efeitos de prestação de informações a potenciais mutuantes. KG, Urteil vom 7.2.2013 – 10 U 118/12 (LG Berlin). Sobre o assunto, ver: KG: KG: Auskunftei darf Restschuldbefreiung drei Jahre speichern (ZD 2013, 189). Há também outros casos nesse sentido. Para tanto, ver BGH, Urt. v. 22. 2. 2011 – VI ZR 120/10 (OLG Jena).

[57] MARANHÃO, Juliano Souza; CAMPOS, Ricardo Resende. Op. Cit.

procedimentos de governança previstos, caso não haja, organizacional ou efetivamente, discricionariedade do humano para divergir da recomendação baseada no *score* de crédito, ou mesmo quando, havendo competência de humano, ela não seja de fato exercida. Nesse sentido, recente decisão da Corte de Justiça Europeia,[58] comentada em maior detalhe no item 4.3 abaixo, equipara a decisão automatizada aqueles sistemas de *credit scoring* nos quais a análise contextual fática demonstre não haver efetiva discricionariedade humana. Este é um exemplo sobre como a precaução substantiva pode apontar a insuficiência dos procedimentos formais legais para então proteger de modo concreto o direito fundamental à proteção de dados pessoais.

No caso de *credit scoring*, seja por força jurisprudencial (Europa) seja por previsão expressa na legislação (LGPD) há o reconhecimento de legítimo interesse para esse tipo de tratamento. Mas mesmo em relação a dados sensíveis, onde não se admite o reconhecimento de legítimo interesse, a análise da perspectiva da precaução substantiva e do uso responsável de IA pode resultar no dever ético do tratamento de dados, independentemente de consentimento dos titulares ou outra base legal, ao menos no que diz respeito à fase de treinamento, para construção de modelos que possam trazer benefícios e promover direitos individuais e sociais.

Exemplo 2 – *Dados de Saúde na Pandemia de COVID-19*

Durante a pandemia, diversos sistemas de IA foram desenvolvidos, testados e alguns aplicados, em diferentes países, para aumentar a eficiência do diagnóstico de infecção, permitir diagnóstico remoto de insuficiência respiratória, recomendar internação ou tipo de tratamento, gerir de leitos, monitorar a quarentena, dentre outros fins.[59] O desenvolvimento e aplicação desses sistemas envolvia a coleta e processamento de dados de saúde, tanto de pessoas infectadas, quanto de pessoas saudáveis, disponíveis nos sistemas de saúde e em outras fontes.

No Brasil diversos projetos foram implementados para este fim, como o SPIRA, desenvolvido pelos Center4AI da USP/IBM, para realizar diagnóstico remoto de insuficiência respiratória de pacientes de COVID-19, como forma de triagem para internação do paciente.[60] O projeto, tanto em relação ao desenvolvimento do sistema

[58] C-634/21, Schufa Holding, ECLI:EU:C:2023:957, (Dec. 7, 2024).

[59] CALANDRA, Davide et al. Artificial Intelligence to fight COVID-19 outbreak impact: an overview. EUROPEAN JOURNAL OF SOCIAL IMPACT AND CIRCULAR ECONOMY., v. 1, n. 3, p. 84-104, 2020.

[60] TUDOCELULAR. *Coronavírus: Projeto Spira recebe mais voluntários para aumentar precisão de diagnóstico de insuficiência respiratória*. Disponível em: https://www.tudocelular.com/seguranca/noticias/n157802/sistema-usp-diagnostico-insuficiencia-respiratoria-covid19-inteligencia-artificial-spira.html.

de IA quanto a sua implementação, tocava em diversas questões éticas sobre responsabilidade dos médicos, autonomia do paciente, transparência e privacidade[61]. Durante o treinamento do sistema, foram coletados, mediante consentimento, dados de voz de pacientes internados além da coleta por meio do site de gravações de falas por pessoas saudáveis e por pessoas infectadas, mas não internadas. Como o treinamento do sistema exigia grande volume de dados de pessoas saudáveis recorreu-se a áudios disponíveis no Museu da Pessoa (https://museudapessoa.org/), surgindo a questão sobre a necessidade de consentimento específico para utilização dos dados para desenvolvimento do sistema. Considerando a urgência da questão de saúde pública seria proibitivamente custoso o esforço de localização e coleta de consentimento específico sobre o uso dos dados para este fim, valendo uma ponderação dos valores subjacentes atento aos interesses em jogo e impactos sobre a privacidade. Ocorre que sistemas de IA aplicados a dados de voz discursivos ou não são capazes de inferir uma série de informações sensíveis, como biometria, orientação sexual, distúrbio psíquico, etc.,[62] e a própria aplicação pretendida envolvia inferência sobre estado de saúde, portanto dados sensíveis, o que impediria o uso de legítimo interesse como base legal.

A análise do valor social da privacidade perante a promoção da saúde pública exige que se separe a fase de desenvolvimento da fase de inferência. Na fase de treinamento, os dados são utilizados para gerar um modelo estatístico que represente as observações de áudios, extraindo padrões verbais e não verbais (respiração, intervalos entre palavras) durante a fala de pessoas saudáveis, não saudáveis e pacientes sem ou com insuficiência respiratória. Tais padrões compõem um espaço latente com representações matemáticas dos dados- não com os próprios dados, mas com clusters ou vetores ou pesos associados a correlações estatísticas. O modelo gerado sobre padrões de fala de pessoas saudáveis ou não saudáveis, portanto, é formado por dados despersonalizados, uma vez que a individualidade não é relevante para o padrão de saúde que se pretende generalizar.

[61] Para uma análise das questões éticas envolvidas em uso de dados de voz para inferências sobre saúde com sistemas de IA, ver ALMADA, M. A. L.; MARANHÃO, J.S.A. Voice-based diagnosis of covid-19: ethical and legal challenges. INTERNATIONAL DATA PRIVACY LAW, v. 11, p. 63-75, 2021.
[62] Jacob Leon Kröger, Otto Hans-Martin Lutz and Philip Raschke, 'Privacy Implications of Voice and Speech Analysis – Information Disclosure by Inference' in Michael Friedewald and others (eds), *Privacy and Identity Management. Data for Better Living: AI and Privacy: 14th IFIP WG 9.2, 9.6/11.7, 11.6/SIG 9.2.2 International Summer School, Windisch, Switzerland, August 19–23, 2019, Revised Selected Papers* (Springer International Publishing 2020) 248.

Nesse quadro, considerando esse propósito de uso, o valor de promoção da saúde pública e sobrepõe-se ao valor social da privacidade, mesmo que, em fases iniciais do tratamento, para se gerar o modelo estatístico de representação, sejam empregados e processados dados pessoais. Já na fase de inferência em que será extraída a probabilidade de determinado indivíduo estar acometido de COVID ou de insuficiência respiratória grave, considerando o impacto da informação sobre a liberdade individual e privacidade, o valor social da proteção de dados sobrepõe-se à saúde pública, devendo haver consentimento daquele indivíduo que se dispõe a fazer a triagem ou o diagnóstico por meio da gravação de áudio.

Como aponta Solove, a legislação de proteção de dados em diferentes países europeus assim como em Estados norte-americanos baseia-se na categorização de dados sensíveis, além do que, tais legislações consideram sensíveis aqueles dados a partir dos quais seja possível inferir dados sensíveis. Mas se dados sensíveis são todos os dados dos quais podemos inferir informações sensíveis, na era da IA praticamente todos os dados são sensíveis (e.g. qualquer registro de voz é dado sensível, já que sistemas de IA podem obter identificação biométrica, inferir orientação sexual, raça e condição de saúde). Assim, propõe que a sensibilidade do dado seja associada ao uso ou informação que se pretende extrair do dado (o dado de voz não é, per se, sensível, mas usar voz para detectar insuficiência respiratória ou orientação sexual é sensível). Já a abordagem de categorização de tipos de dados, independentemente do uso que deles venha a ser feito, é contraproducente, diante do atual poder de inferência de sistemas de IA, pois *"nearly all personal data can be sensitive, and the sensitive data categories can swallow up everything"*.[63] A proteção de dados, como precaução substantiva, apenas faz sentido a partir da identificação do dano potencial à personalidade individual que pode resultar ao final do processo de tratamento como um todo.

Outra encruzilhada ética na interface entre proteção de dados pessoais e inteligência artificial está no conflito entre, de um lado, a confiabilidade, que se liga à produção eficiente dos resultados e benefícios para o qual o sistema é desenhado e, de outro, a transparência em termos informações sobre os critérios relevantes para determinação dos resultados do sistema (explicabilidade).[64]

A explicabilidade está presente em diferentes legislações, como a europeia (art. 22 GDPR e a brasileira, art. 20 LGPD) como direito do sujeito de dados em relação a decisões

[63] SOLOVE, Daniel J. Data is what data does: regulating based on harm and risk instead of sensitive data. Northwestern University Law Review, 118:1081 (2024).

[64] "Here, then, is a core and, for the moment, unavoidable trade-off in designing algorithmic accountability regimes: Interpretability often comes only at the cost of efficacy" (p. 7). ENGSTROM, David Freeman; HO, Daniel E., Artificially Intelligent Government: A Review and Agenda. In: VOGL, Roland (org). Big Data Law, forthcoming 2020, March 9, 2020.

automatizadas, trazendo como contrapartida o dever de disponibilizar uma justificativa humanamente inteligível sobre os outputs, ou, no caso brasileiro, sujeitando os controladores a uma auditoria pela autoridade para verificar discriminação. O bem protegido pelo direito e correspondente obrigação está, em última análise na possibilidade de contestar a decisão, a partir da inadequação dos critérios, de imprecisões nos seus dados pessoais empregados que levaram à sua categorização em determinado perfil ou inferência, ou a presença de traços discriminatórios que possam dificultar acesso a bens ou direitos.

Existe, porém um trade-off entre confiabilidade, em termos de capacidade preditiva, na qual é baseada a decisão automatizada, e a explicabilidade do modelo. Modelos mais complexos, como redes neurais profundas, têm poder de predição bastante elevado, mas são consideradas *"caixas pretas"* em razão da opacidade quanto aos critérios relevantes para seus resultados (decisões, predições, recomendações). [65] Por outro lado, modelos mais transparentes, como árvores de decisão ou métodos lineares, permitem compreender claramente os fatores envolvidos nas decisões, mas frequentemente apresentam desempenho inferior em termos de confiabilidade. Um exemplo no campo da saúde pode ilustrar as questões éticas que podem surgir diante desse *trade-off*.

Exemplo 3 – Explicabilidade de Diagnósticos de Câncer em Manchas de Pele

O câncer de pele tem sido responsável pelo maior índice de mortalidade relativo a essa doença e o diagnóstico precoce é decisivo no tratamento, razão pela qual diversos sistemas de IA vem sendo desenvolvidos para possibilitar diagnóstico remoto, ou maior precisão no diagnóstico, por meio da análise de imagens de manchas na pele.[66] Diversas metodologias de aprendizado de máquina, incluindo *deep learning*, tem sido empregadas, com resultados promissores, onde alguns sistemas alcançam precisão na identificação de melanomas próxima a 100%.[67] Além da complexidade do modelo análise da imagem da mancha de pele pelos sistemas computacionais é feita sobre sua decomposição em milhares de pixels, o que é invisível ao olhar humano, o que torna o output ininteligível para humanos, mesmo para médicos especialistas que não tem

[65] LEHR, D. e OHM, P. Playing with the data: what Legal Scholars Should Learn About Machine learning. UC Davis Law Review, v. 51, Dec. 2017.

[66] MELARKODE N, SRINIVASAN K, QAISAR SM, PLAWIAK P. AI-Powered Diagnosis of Skin Cancer: A Contemporary Review, Open Challenges and Future Research Directions. Cancers (Basel). 2023 Feb 13;15(4):1183.

[67] HALL, Emma. *AI software achieves 100% melanoma detection rate*. FMAI Hub, 17 out. 2023. Disponível em: https://www.fmai-hub.com/ai-software-achieves-100-melanoma-detection-rate/.

acesso a uma explicação causal da recomendação de diagnóstico.[68] Algumas técnicas desenvolvidas em XAI (Inteligência Artificial Explicável) permitem uma aproximação, como a análise de sensibilidade, que indica qual área de *pixels* na imagem tem maior correlação com o resultado, sendo, porém, insuficiente como justificação causal ou mesmo com justificação em conotações menos exigentes de explicação (contrastividade ou contestabilidade).[69] No atual estado da arte, tais sistemas não substituem o especialista, que toma a recomendação como subsídio para a análise clínica. Nesse quadro, considerando o valor ético de transparência e autonomia do paciente, frente ao princípio de responsabilidade do médico, deveria este último informar sobre o resultado preditivo do sistema de IA, quando discordar de sua recomendação? Caso o diagnóstico de melanoma por IA seja incorporado como melhor prática em medicina baseada em evidências, haveria a obrigação de dispor de mecanismos de explicação do resultado como condição para uso da tecnologia?

Aparte o conflito ético entre transparência quanto ao uso de IA no diagnóstico e a responsabilidade final do médico pela terapia a ser adotada, uma exigência de explicabilidade em termos humanamente inteligíveis, seja de justificação causal (qual elemento da imagem da mancha na pele está ligado a determinado sintoma de melanoma), de contrastividade (quais aspectos da imagem cuja alteração geraria um resultado diverso) [70] ou mesmo contestabilidade (capacidade do afetado contestar a decisão) dificilmente seria atendida. Nesse caso, os benefícios do uso de sistemas complexos, mas confiáveis, de *deep learning*, em termos de diagnóstico preciso e tratamento precoce, não deveriam ser obstados, apenas em função da dificuldade ou mesmo impossibilidade de se fornecer meios de explicação. Antes, a avaliação deve contrastar o benefício social e individual da tecnologia, com o impacto da ausência de explicação causal.

Da perspectiva do médico, que opera o sistema, a ausência de uma explicação causal reduz o valor da informação resultante do sistema de IA. Já em relação ao paciente, deve-se examinar o quão em risco estão os bens protegidos pelo direito à explicação. Assim, a possibilidade de contestação da decisão não se aplica, tendo em vista que, apesar da

[68] HOLZINGER, A. , LANGS, G. , DENK, H., ZATLOUKAL, K. MULLER, H. Causability and explainability of artificial intelligence in medicine. *WIREs Data Mining and Knowledge Discovery*, 2019, 9.

[69] MARANHÃO, J. S. A.; COZMAN, F. G. ; ALMADA, M. . Concepções de explicação e do direito à explicação de decisões automatizadas. In: Rony Vainzof, Andrei Gutierrez. (Org.). Inteligência Artificial : Sociedade Economia e Estado. 1ed.São Paulo: Thomson Reuters Brasil, 2021, v. 1, p. 137-154.

[70] MITTELSTADT, Brent; RUSSELL, Chris; WACHTER, Sandra. Explaining explanations in AI. In: Proceedings of the conference on fairness, accountability, and transparency. 2019. p. 279-288.

autonomia do paciente, em bioética é responsabilidade do médico a definição do tratamento. Também não se aplica a retificação de dados pessoais no momento da inferência, uma vez que a classificação é de determinada imagem da mancha sobre a qual é aplicado o modelo matemático de representação com a composição de funções objetivas e respectivos pesos, que melhor detecta o padrão de melanomas, não havendo aqui impacto sobre a personalidade individual, seja no treinamento do modelo, seja na inferência.

Por fim, o risco de discriminação pode estar presente quanto à qualidade e precisão da predição para diferentes grupos (por exemplo, sistema treinado com dados de pacientes de hospitais privados, com maior renda e com predominância de determinada raça), a depender da representatividade da base dados usada para o treinamento, mas esse seria um aspecto da confiabilidade e diria respeito ao valor social, não individual, de privacidade, pois não se daria por alocação privilegiada ou minorizada de determinado perfil individual. Portanto, inexistindo impacto sobre personalidade individual, novamente a confiabilidade, considerando os benefícios sociais do sistema em questão, parecem prevalecer sobre o valor social de privacidade e proteção de dados (concretizada no exercício do direito à explicação de decisões automatizadas).

A dificuldade em relação à explicabilidade, embora não seja suficiente para justificar a proibição de uso desses sistemas, pode ser abordada por outros mecanismos, caso haja impacto significativo sobre a pessoa afetada, como, por exemplo, o caso de erro médico, sendo possível fornecer informações relevantes em diversas camadas por meio do aparato institucional, como perícias, auditorias, que podem trazer elementos para as partes e o julgador em eventual conflito.[71]

[71]WISCHMEYER. T. Artificial Intelligence and Transparency: opening the black box, em WISCHMEYER, T. e RADMACHER, T. Regulating Artificial Intelligence, Springer, 2020, p. 75 e ss.

4. PROTEÇÃO DE DADOS PESSOAIS À LUZ DA IA RESPONSÁVEL

Entre 1996 e 1997, os dois *matches* disputados entre Garry Kasparov e o sistema de inteligência artificial Deep Blue ocuparam o noticiário internacional, pelo equilíbrio das partidas e posterior vitória da segunda versão do sistema sobre o então campeão mundial de xadrez.[72] Além do avanço tecnológico, aquele celebrado marco permite refletir sobre a interação homem-máquina e, consequentemente, qual a posição a ser adotada pelo direito no papel de regulação da inteligência artificial.

Para lidar com sucesso em suas partidas com o Deep Blue, seria inútil para Kasparov adotar a *perspectiva física*, isto é, considerar os impulsos elétricos e conexões entre os transistores ou a operação de processamento do hardware que resultava nas imagens projetadas na tela do computador, com os movimentos de seu oponente. Também seria inútil para Kasparov adotar a *perspectiva do programa* ou da lógica de processamento dos dados para determinação dos movimentos. Isso porque, ao passo que Kasparov vislumbrava 3 jogadas por segundo, o Deep Blue calculava cerca de 3 mil jogadas por segundo, fazendo os respectivos cálculos de probabilidade de sucesso das jogadas subsequentes. Assim, para lidar com sucesso com os movimentos apresentados pelo Deep Blue, a única estratégia bem sucedida foi reconhecer um comportamento da máquina e adotar uma *perspectiva intencional*, [73] buscando entender *qual o propósito e quais resultados* o sistema objetivava alcançar com tais movimentos, ainda que soubesse que o sistema Deep Blue não possuía consciência, intenção ou entendimento sobre o contexto da partida disputada, respondendo apenas aos múltiplos cálculos do programa a partir dos *inputs* recebidos.

Da mesma forma, ao construir modelos de gestão de risco e regulação no sentido de obter uma IA responsável, a perspectiva relevante é aquela do comportamento exibido pelo sistema por meio do seu propósito, ou seja, para aquilo que foi desenhado, e por meio dos resultados apresentados.

Embora haja ações humanas envolvidas desde o *design*, passando pelo desenvolvimento e implementação do sistema, a segmentação nas diferentes etapas e procedimentos de tratamento de dados não permite apreender o sistema como um todo, que envolve, como veremos abaixo, não apenas a dimensão técnico-informática, como também a dimensão humana envolvida (IA como sistema sociotécnico). É desse sistema, tomado como

[72] WIKIPÉDIA: a enciclopédia livre. *Match Garry Kasparov vs. Deep Blue*. Disponível em: https://pt.wikipedia.org/wiki/Match_Garry_Kasparov_vs_Deep_Blue. Acesso em: 3 abr. 2025.
[73] Dennet, D. The intentional stance. MIT Press, 1989.

verdadeiro *"agente"* que se exige um comportamento responsável, em termos de promoção de valores sociais e proteção de direitos fundamentais, em particular a privacidade e proteção de dados.

Voltando para a pergunta colocada no início desse estudo, reduzir a inteligência artificial a um conjunto de operações de processamento de dados, sem atentar para seu resultado e propósito, é incapaz de captar esse aspecto de comportamento ou mesmo de se conceptualizar adequadamente o que seria um comportamento responsável para sistemas de IA. A visão ficaria restrita à fiscalização do comportamento responsável de cada controlador em cada etapa de tratamento, isoladamente considerada, levando-se em consideração apenas o valor individual da privacidade e proteção de dados. Como vimos acima, essa perspectiva procedimental, limita a fiscalização pela autoridade a, no máximo, uma garantia de níveis aceitáveis de risco, de uma perspectiva egocentrada no titular, sendo incapaz de resultar em garantia substantiva do direito fundamental à proteção de dados pessoais.

Já a noção de precaução substantiva, presente na origem jurisprudencial da proteção dados pessoais, como garantia de direitos fundamentais, embora possa ter ficado de certo modo opaca na formulação legislativa, vem sendo resgatada no contexto do avanço da Inteligência Artificial, justamente pela necessidade de concretização de valores não só individuais como sociais. Ou seja, a proteção de dados pessoais como valor social passa a ser um componente da construção de sistemas (sociotécnicos) de IA, que exibam comportamento responsável, em seus objetivos e resultados.

Nesse sentido, Graeden et. al.,[74] ao analisar a legislação europeia e norte-americana de proteção de dados propõem, perante a IA, um paradigma regulatório e de intervenção pelas autoridades baseada em resultados (*outcomes-based* regulation paradigm). Conforme observam, abordagens que enfoquem tipos de dados de operações de tratamento tem por efeito criar uma tensão entre inovadores e reguladores que pode limitar o acesso de consumidores a tecnologias novas, muitas vezes vitais, e falha em protegê-los efetivamente contra danos a direitos. A abordagem *outcomes-based* por eles proposta é mais congênere à forma pela qual os engenheiros desenvolvem sistemas de IA, i.e. enfatizando soluções para trazer melhorias em resultados, de modo que a regulação, em vez de limitar técnicas e procedimentos, interferindo na liberdade criativa, deve circunscrever as restrições aos resultados. Argumentam assim, que abordagens baseadas nos resultados e propósitos do sistema direcionam os engenheiros para criar

[74] GRAEDEN, Ellie; STEVENS, Tess; KNODEL, Mallory; BENNETT, Ashley; ROSADO, David; HENDRICKS-STURRUP, Rachele; REISKIND, Andrew; LEITNER, John; LEKAS, Paul; DEMOOY, Michelle. *An outcomes-based paradigm for data and AI regulation*. Lawgorithm: ibero-american Journal on Artificial Intelligence and Law, 2025. A ser publicado.

soluções que evitem tais resultados, sem restringir procedimentos ou métodos, sendo, portanto, mais efetiva tanto para mitigar potenciais danos, como também para promover a inovação e seus benefícios.

A abordagem proposta é natural quando efetivamente se adota a perspectiva da *IA como agente* (sociotécnico). O direito, usualmente, perante agentes, regula os resultados de suas ações e não os movimentos corporais correspondentes: proíbe-se o assassinato ou a lesão corporal, não "o apertar do gatilho" ou "o empunhar de um objeto cortante"; da mesma forma, para proteger pedestres, regras focam nas situações que possam provocar atropelamento e não no tipo de veículo envolvido. O tratamento da IA pelo direito deve ser o mesmo, considerando os resultados de seu comportamento, que envolve o processamento de dados, e não considerar procedimentos de tratamento de modo isolado, como um fim em si mesmo.

Tal postura regulatória, que lê a IA pelos óculos da proteção de dados, e não o contrário, pode prejudicar a inovação, e, portanto, o melhor aproveitamento dos benefícios sociais que a IA pode gerar,[75] sem trazer garantia efetiva à privacidade e proteção de dados.

Isso não significa colocar à margem a proteção de dados pessoais em prol do desenvolvimento tecnológico. Pelo contrário, a proteção de dados pessoais está no cerne da regulação europeia sobre serviços, mercados digitais e inteligência artificial.[76] A privacidade e proteção de dados devem compor os valores estruturantes da IA responsável, [77] mas não pode ser um condicionante para a IA, ou uma precondição para seu desenvolvimento. Essa leitura é congênere com o resgate da precaução substantiva na interpretação da legislação de proteção de dados e no debate doutrinário que vê a abordagem de segurança de produto nas legislações sobre IA como mecanismo *apenas instrumental*, i.e., como simples meio, quando a preocupação está em assegurar, efetivamente, direitos fundamentais e valores sociais que podem, ao mesmo tempo ser ameaçados e promovidos por sistemas de IA.

Nesta Seção abordaremos os desafios que surgem na aplicação da LGPD a sistemas de IA, em sua visão procedimental e egocentrada no titular, oferecendo uma interpretação baseada no princípio de prevenção (art. 6º, inc.) como precaução substantiva, de modo a

---

[75] MARANHÃO, Juliano. **IA e o risco do medo**. JOTA, 2023. Disponível em: https://www.jota.info/artigos/ia-e-o-risco-do-medo. Acesso em: 06 mar. 2025.

[76] ZANFIR-FORTUNA, Gabriela, Follow the (personal) Data: Positioning Data Protection Law as the Cornerstone of EU's 'Fit for the Digital Age' Legislative Package (March 15, 2024). EDPS at 20 Anniversary Volume, Forthcoming June 2024, Available at SSRN: https://ssrn.com/abstract=4794182 or http://dx.doi.org/10.2139/ssrn.4794182

[77] MARANHÃO, J.; NAVAS, J. Certificação como instrumento de regulação da Inteligência Artificial no AI Act. In: VAINZOF, R.; GUTIERREZ, A.; GODINHO, G.; KRASTINS, A. (Coords.). **Comentários ao EU AI Act.** 2024.

conciliar a proteção de dados pessoais como valor social, como um dos valores formadores da chamada IA Responsável. A análise será voltada para quatro vícios de interpretação baseados na leitura procedimental, a ser enfrentado em cada subitem:

(i) **"exigir a explicação de tudo"**: apegar-se à explicação causal na justificativa de tratamento de cada ponto de dado utilizado pelo sistema de IA;
(ii) **"enfatizar o tratamento em vez do resultado":** enfatizar o tratamento de dados, sem considerar o resultado e impacto final do tratamento com o desenvolvimento e implementação do sistema de IA;
(iii) **"supervalorizar a máquina"**: entender a decisão automatizada como operação meramente informática e não como ação sociotécnica;
(iv) **"confundir supervisão com revisão"**: acenar para o *desideratum* da supervisão humana no ciclo de vida a IA para exigir a revisão humana do output do sistema de IA

4.1. Primeiro "Vício": *Exigir a Explicação de Tudo* (ou o Apego à Justificação de Cada Ponto de Dado em sua Relação Causal com o Propósito de Tratamento)

Em 2020, o aprendizado de máquina, empregado por pesquisadores do MIT, foi responsável pela descoberta da *halicina*, um novo antibiótico capaz de combater tipos de bactérias resistentes a todos os antibióticos então conhecidos. Essa conquista não só ilustra os benefícios sociais da IA, como ajuda a entender a forma pela qual são obtidos seus resultados. Os esforços humanos para desenvolvimento de antibióticos, até então, envolviam enormes recursos para que especialistas pudessem estipular dentre as moléculas de medicamentos conhecidos, quais teriam fatores causalmente ligados à sua eficácia para inibir o desenvolvimento bacteriano, desde peso atômico até tipos de ligações moleculares. Humanos, porém, são limitados em sua capacidade de processamento e análise de grandes quantidades moléculas, além de se limitarem aos fatores conhecidos para estabelecerem suas hipóteses ou predições. O sistema de IA empregado, por sua vez, foi treinado para reconhecer padrões estatísticos dentre moléculas eficazes para combate a bactérias, aplicando-os na análise de uma biblioteca de 61.000 moléculas. O sistema não só analisou características antibacterianas conhecidas pelos especialistas, como identificou novos atributos não só desconhecidos, como não conceptualizáveis em termos causais, que se mostraram estatisticamente relevantes para que fosse identificada uma determinada molécula, nomeada em homenagem ao HAL do filme "*2001 uma Odisséia no Espaço*", de Stanley Kubrick, que se

mostrou, após testes, um antibiótico não tóxico, efetivo e diverso de todos os até então conhecidos. Apesar da sua eficácia, o programa não entende, nem explica causalmente por que a molécula funciona e os desenvolvedores ou especialistas no domínio de aplicação também não encontram uma explicação intuitiva a partir do próprio conhecimento. Ou seja, não se sabe como ou por que a halicina cumpre seu propósito; sabe-se que ela o cumpre.

O caso da halicina mostra como a exigência de base legal vinculada a uma justificação causal do propósito de cada ponto ou tipo de dado empregado no desenvolvimento do sistema de IA pode impor limites de confiabilidade ou mesmo inviabilizar seu desenvolvimento pela própria característica técnica desses modelos, nos quais, muitas vezes, não é possível fornecer explicação causal ou relação meio e fim para as correlações estatísticas entre diferentes pontos de dados.

Na interpretação procedimental da legislação de proteção de dados, espera-se que o controlador defina finalidade específica para tratamento de cada dado ou cada tipo de dado pessoal processado. Ou seja, espera-se que se atribua explicação causal ou uma relação entre o dado como meio para se atingir o propósito de tratamento almejado pelo controlador. Contudo, sistemas de IA baseados em modelos matemáticos complexos com significativo volume de dados não permitem que se estabeleça uma relação linear entre o tipo de dado, seu tratamento e a finalidade pretendida.

Ocorre que modelos baseados em aprendizagem de máquina extraem padrões a partir das regularidades e correlações dos dados coletados para estabelecer classificações e realizar predições, sem que, necessariamente, seja explicitamente programado especificando-se cada fator ou atributo relevante para tal classificação ou predição.[78] Em estágios iniciais de emprego de sistemas computacionais para predição e mesmo em modelos menos complexos de aprendizado de máquina há uma espécie de "*engenharia de fatores*" (*feature engineering*), em que especialistas em determinado domínio, com base em seu conhecimento ou com base em modelos teóricos, selecionam as variáveis explicativas relevantes como hipóteses causais de determinado resultado, objeto de predição. Em modelos de aprendizado de máquina mais complexos, porém, com a enorme disponibilidade de dados e elevada capacidade de processamento

[78] Na aprendizagem supervisionada, os modelos aprendem a partir de exemplos previamente classificados ou rotulados para reproduzir os padrões identificados. Já a aprendizagem não supervisionada ocorre quando não existem categorias pré-definidas nos dados. Nesse caso, o algoritmo explora os próprios dados para encontrar padrões ou agrupamentos naturais, identificando estruturas ou relações sem orientação prévia. Por outro lado, a aprendizagem por reforço é baseada em tentativa e erro, onde um agente (por exemplo, um robô ou um programa) interage com o ambiente e aprende com os resultados de suas ações. AKASH, Mamidi Sai. Artificial intelligence & neural networks. International Journal of Soft Computing and Artificial Intelligence, v. 3, n. 2, p. 37-44, 2015.

computacional, a lógica se inverte. A chave para a seleção das variáveis explicativas do modelo passa a ser feita a partir da mensuração de correlações estatísticas entre as variáveis de *input* e a variável de *output*. Tais correlações são calculadas a partir de milhões de observações pelo sistema computacional.[79]

A relevância ou utilidade de um determinado dado (pessoal ou não) para um sistema de IA, portanto, é avaliada estatisticamente, pelo seu potencial contributivo para incrementar a acurácia do modelo, ou seja, para refinar a função objetiva generalizada a partir das observações e elevar sua confiabilidade e capacidade preditiva, sem que seja possível atribuir relação causal imediata, entre o dado e o resultado da predição.[80] O tratamento do ponto de dados individual, considerado isoladamente, muitas vezes não possui uma justificativa clara, sendo relevante apenas quando inserido no contexto amplo do conjunto de dados usado para treinamento ou inferência dentro de determinado modelo.

Assim, por exemplo, um sistema suficientemente complexo de *scoring* de crédito não parte de uma engenharia humana sobre quais fatores seriam por hipótese razoáveis para causar a inadimplência (e.g. dados de renda anual, histórico de adimplemento, dívidas em curso, patrimônio). Parte-se, antes, do tipo de modelo matemático que é alimentado pelos dados disponíveis, de modo que o próprio sistema encontre correlações significantes entre os pontos de dados, resultando em predições confiáveis, correlações essas que podem ser mesmo surpreendentes para a interpretação humana.

Assim, por exemplo, diferentes hábitos de uso de telefone celular mostraram-se relevantes, com uso de metodologia de árvores randômicas de decisão empregadas em modelos de *credit scoring*, para predizer se usuários de telefonia pré-paga seriam bons candidatos para clientes de celular pós-pago. Mais do que isso, tais hábitos uso do telefone celular, como realização de chamadas internacionais, intervalos de carregamento de créditos ou diferentes cortes de períodos entre chamadas, dentre centenas de pontos de dados, mostram-se não só relevantes para predizer adimplência de contas de telefonia, como adimplência em geral. Tais fatores não são intuitivos para humanos em termos de relação causal, mas o modelo matemático empregado encontra correlações significantes, de modo que a ausência desses dados no modelo, reduzem sua acurácia. Impedir o seu uso, por incapacidade de explicação causal desses itens como fatores de proteção ao crédito, poderia inviabilizar a tomada de empréstimos por

---

[79] ANDERSON, Chris. The end of theory: The data deluge makes the scientific method obsolete. **Wired magazine**, v. 16, n. 7, p. 16-07, 2008.

[80] LONDON, Alex John. Artificial intelligence and black-box medical decisions: accuracy versus explainability. **Hastings Center Report**, v. 49, n. 1, p. 15-21, 2019.

boa parte da população brasileira que, embora desbancarizada, possui telefone celular e, além de limitar a redução de riscos para o sistema financeiro.

O exemplo coloca em questão se a aplicação da legislação de proteção de dados deveria exigir uma justificativa compreensível para humanos para cada ponto de dado individualmente considerado, ou se deveria prevalecer a precisão e a justiça geral do modelo, seu impacto na inclusão financeira e na capacidade de ser robusto o suficiente para atender as normas de gestão integrada de riscos do mercado financeiro.

Isso porque os dados selecionados não são necessariamente aqueles que o humano supõe causalmente relevantes, mas aqueles com maior correlação estatística com a determinação do *output*. Esse descompasso entre correlação estatística e explicação causal pode ser também ilustrado por sistema desenvolvido para predizer a divergência entre juízes da corte de apelação norte-americana. Se as hipóteses formuladas por humanos que seriam causadoras de divergência apontavam para a diferença de orientação ideológica dos magistrados ou posições em precedentes semelhantes, os fatores que mais guardavam correlação para o sucesso de predição de divergência não diziam respeito ao perfil progressista ou conservador dos magistrados, mas, antes, diziam respeito à (i) posição em que os juízes se sentavam durante o julgamento, (ii) tamanho do voto e (iii) número de citações do voto. [81]

Assim, os modelos não são formulados a partir de, nem buscam hipóteses causais. Busca-se, a partir de uma grande massa de dados, obter uma regra de decisão ou função objetiva probabilística, conforme o modelo matemático escolhido (regressão linear, árvore de decisão, florestas randômicas, redes neurais, etc.) que atribua pesos relativos aos dados de entrada, de modo a obter a predição mais acurada da variável de saída.[82] Nesse processo de múltiplas observações, dentro do desenvolvimento do modelo, as próprias variáveis de entrada, i.e. os tipos de dados a serem empregados no sistema, podem ser escolhidos ou descartados, conforme seu grau de correlação ou sua significância estatística, considerando todo o conjunto de dados simultaneamente testados (não ocorre um teste individual de cada dado em relação ao output).

Ou seja, é o resultado ou o sucesso das múltiplas observações de predições processadas pelo sistema informático, a partir de diferentes conjuntos de variáveis, com diferentes pesos, que determina quais variáveis de entrada e com qual peso seriam relevantes. Assim, para algoritmos mais complexos de aprendizado de máquina, em que lidamos com milhões de observações, de múltiplos pacotes de variáveis, com diferentes

[81] CHEN, Daniel L.; PARTHASARATHY, Adithya; VERMA, Shivam. The Genealogy of Ideology: Predicting Agreement and Persuasive Memes in the US Courts of Appeals. 2017.
[82] LEHR, David; OHM, Paul. Playing with the data: what legal scholars should learn about machine learning. UCDL Rev., v. 51, p. 653, 2017.

sopesamentos, é praticamente impossível conectar um *input* específico a determinado *output*.[83] Com isso, há um desencontro inevitável entre a otimização matemática das predições típicas dos modelos de *machine learning* com a escala humana de raciocínio e estilo de interpretação semântica.[84]

Desse modo, a exigência de justificativa legal para o tratamento de cada ponto de dados para treinamento ou inferência de um sistema de IA pode ser incompatível com a natureza das operações realizadas pelo modelo. Tal exigência de justificativas causais para o tratamento não seria atendida nem mesmo com base nas técnicas atuais de explicabilidade para os modelos mais complexos.[85] Em geral, as abordagens disponíveis para explicabilidade em inteligência artificial limitam-se a indicar o impacto das variáveis sobre a acurácia do modelo, sua influência em decisões individuais ou oferecer simplificações interpretáveis do modelo original.[86]

Dessa perspectiva, como aponta Solove, deixa de fazer sentido, no campo da IA, inquirir sobre a adequação ou categoria de cada tipo de dado usado, previamente ao resultado gerado por seu uso. O dado e sua relevância (estatística e *não* causal) é definido pelo resultado de seu uso no sistema. Integrando-se a proteção de dados pessoais ao uso responsável do sistema de IA, valendo-se de precaução substantiva (como expressão do princípio de prevenção, art. 6º, inc. VIII da LGPD), deve-se avaliar o impacto do sistema de IA em todos os valores envolvidos, e não só sobre a dimensão individualista e egocentrada de privacidade.

No exemplo de *credit scoring* acima, em que dados sobre hábitos de telefonia de população desbancarizada podem, pelo sistema de IA, propiciar sua inclusão e acesso a crédito, estimular o consumo e a economia em geral, em vez de se perguntar pela sensibilidade *per se* dos dados empregados, a questão a ser colocada é se tal uso tem resultado sobre elementos sensíveis da personalidade individual e se a privacidade, como valor social, deveria prevalecer perante os demais valores envolvidos.

Dessa forma, uma abordagem mais adequada ao contexto dos sistemas de IA deve deslocar o foco da justificativa individualizada por ponto de tratamento para uma

---

[83] WISCHMEYER, Thomas. Artificial intelligence and transparency: opening the black box. In: Regulating artificial intelligence. Cham: Springer International Publishing, 2019. p. 75-101.

[84] Burrell, J. How the machine thinks: understanding opacity in machine learning algorithms. Big Data Soc. 3. 2016. https://doi.org/10.1177/2053951715622512.

[85] KESARI, Aniket et al. A Legal framework for explainable artificial intelligence. Center for Law & Economics Working Paper Series, v. 9, 2024.

[86] MARANHÃO, J. S. A.; COZMAN, F. G. ; ALMADA, M. . Concepções de explicação e do direito à explicação de decisões automatizadas. In: Rony Vainzof, Andrei Gutierrez. (Org.). Inteligência Artificial : Sociedade Economia e Estado. 1ed.São Paulo: Thomson Reuters Brasil, 2021, v. 1, p. 137-154.

perspectiva mais integrada, na qual a análise é dirigida à relevância dos dados para a finalidade geral do modelo e ao impacto global de sua aplicação, sob pena de inviabilizar a implementação de sistemas e modelos de IA responsáveis, social e economicamente benéficos.

Isso implica uma leitura extensiva de bases legais que apontam resultados na LGPD, como, por exemplo, o inc. X do art. 7º da LGPD, tomando-se como dado para proteção ao crédito não apenas aqueles que pela inteligibilidade causal humana pareçam razoavelmente levar a essa finalidade, mas virtualmente todos os dados que sejam capazes de, em determinados modelos de *scoring de crédito*, elevar a confiabilidade do sistema. Ou seja, não é o tipo de dado individualmente considerado que importa para definir sua própria relevância, antes é a confiabilidade do modelo de IA que atribui relevância para todos os dados que contribuam com correlação significante para incrementar sua acurácia.

O mesmo vale para a interpretação do inc. VIII do art. 7º e inc. II, "f" do art. 11, portanto aplicável a dados sensíveis, que trata tutela à saúde, em procedimentos realizados por profissionais de saúde ou serviços de saúde. Perante a IA, em vez de restringir a aplicação dessa base legal para dados empregados pelo médico para uma intervenção específica, a leitura extensiva, com a incorporação da IA gradualmente como melhor prática em medicina baseada em evidência, pode considerar a IA como parte integrante de um serviço de saúde e, portanto, de procedimentos com essa finalidade, o que autoriza o emprego dos dados relevantes para desenvolvimento do modelo, seja na fase de treinamento, seja na fase de inferência.

4.2. Segundo "Vício": Enfatizar o Tipo de Dado ou de Tratamento (Sem Olhar para o Impacto Concreto da Aplicação do Sistema de IA)

Outro vício decorrente da leitura estritamente procedimental da LGPD consiste em se privilegiar a análise isolada de cada tipo de dado pessoal e de cada etapa do tratamento em vez de se enfatizar o resultado do tratamento e seu efetivo impacto sobre personalidades individuais, direitos fundamentais e sobre o valor social da privacidade.

A dificuldade aparece principalmente em relação ao treinamento de sistemas de IA de propósito geral em que dados pessoais são utilizados para se extrair padrões gerais de determinada população ou conjunto de documentos, que por sua natureza, são despersonalizados. Ou seja, quando dados pessoais deixam de ser personalizados ou são anonimizados no curso do processo de treinamento, na construção do modelo de IA, ou ainda, nas inferências geradas pelo sistema de IA.

Recentemente, a ANPD determinou, em medida preventiva,[87] a suspensão do treinamento do sistema Meta AI, chatbot generativo baseado no grande modelo de linguagem (*Large Language Model*) Llama, a partir de textos escritos por usuários brasileiros das redes sociais controladas pela Meta. Segundo a nota da ANPD, haveria risco de *"dano irreparável"* à proteção de dados daqueles usuários por entender em juízo preliminar, não haver base legal para as etapas de coleta, armazenamento e processamento dos textos dos usuários nas redes. A intervenção foi revertida pela Agência, após a demonstração de medidas de transparência sobre a finalidade de uso e propriedades do sistema desenvolvido.

Por sua vez, o EDPB, destacando o valor social da proteção de dados e a promoção de direitos fundamentais conexos, como visto acima, analisou o tema em sua *"Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models"*. O parecer, segundo as regras de exercício de competência daquele órgão, fica limitado às perguntas colocadas na consulta, que se direcionaram a esclarecer quando e como um modelo de IA pode ser considerado anônimo e como controladores podem demonstrar legítimo interesse nas fases de desenvolvimento e implementação do modelo.

A resposta a tais perguntas seguiu a metodologia tradicional de análise de anonimidade e legítimo interesse. Ou seja, segundo o EDPB, o modelo é considerado anônimo se o controlador dificilmente for capaz de extrair informações individuais a partir do modelo, ou se for insignificante a possibilidade de usuários extraírem informações pessoais do sistema em suas requisições (*prompts*). Por sua vez, o controlador pode mostrar legítimo interesse pelo tradicional teste triplo de identificação do interesse, da necessidade do processamento para seu alcance e da ausência de sobreposição de direitos fundamentais dos sujeitos de dados sobre tal interesse.

Antes de analisar tal posicionamento na visão procedimental, vale esclarecer elemento técnico relevante sobre o desenvolvimento e emprego desses modelos.

Embora possa haver aplicações diversas do desenvolvimento de modelos fundacionais nas fases de codificação, apropriada para tarefas de predição, mineração e extração de dados, de um lado, e de decodificação, apropriada para tarefas de geração de novo conteúdo, de outro, tais modelos podem e são usualmente empregados como base para criação de sistemas de IA generativa com diferentes graus de generalidade em seu propósito. A construção desses modelos parte do armazenamento ou coleta, de algum

[87] BRASIL. Autoridade Nacional de Proteção de Dados. **Despacho Decisório nº 20/2024**. *Diário Oficial da União: seção 1*, Brasília, DF, 2024. Disponível em: https://www.in.gov.br /en/web/dou/-/despacho-decisorio-n-20/2024/pr/anpd-569297245. Acesso em: 13 mar. 2025.

modo, de enorme base de dados. Tais dados são usados no treinamento para que se sejam extraídos vetores com pesos, ou *embeddings*,[88] dentro de determinado modelo matemático (de complexidade diversa). Tais vetores representam padrões observados nos dados e permitem encontrar funções apropriadas para instanciações confiáveis (com diferentes graus de acurácia), capazes de gerar novos conteúdos. Uma vez extraídos esses padrões matemáticos é esse espaço latente o conteúdo utilizado para a realização de inferência geradora de novos conteúdos, não sendo mais necessário se acessar os dados originários.[89]

É exatamente por esse motivo, aliás, que ocorrem as chamadas "alucinações" em grandes modelos de linguagem, com geração de referência a fatos inexistentes, uma vez que tais modelos são treinados para detectar padrões em escrita, simulando diálogos humanos, mas sem verificação de fontes, ou seja, sem consulta à base de dados originária. A pesquisa e desenvolvimento nesse campo tem caminhado no sentido de acoplar a tais sistemas habilidades de deliberação, avaliação e verificação de informações geradas nos textos, áudios ou vídeos, esforço que, na verdade, evidencia o descolamento dos dados originais, tanto no modelo resultante, quanto no processo de inferência.

Como tais modelos são treinados sobre imensa base de dados, por sua natureza, detectam padrões gerais de escrita, de diálogo, de voz, de geração de imagens e de edição de vídeos, que não preservam aspectos de individualidade, ou seja, são, por sua natureza, anonimizados. É claro que, a depender da exposição e relevância pública de determinados indivíduos com dados na internet, o sistema pode gerar outputs adequados e informativos sobre tal indivíduo com elementos pessoais, mas o que se "aprende" nesse processo, não são aspectos da identidade e personalidade individual, mas informações públicas sobre pessoas notórias.

Seria bastante custoso e desafiador para desenvolvedores e operadores criar novos sistemas de IA com o propósito de inferir informações pessoais identificáveis, a partir do espaço latente de funções que compõem o modelo matemático. Na verdade, um dos campos de maior investimento no setor está na capacidade de tais sistemas realizarem pesquisas de modo a fornecer informações fidedignas. Mesmo assim, tais investimentos não se direcionam a reverter espaços latentes nos dados originários, mas acoplar outros sistemas capazes de realizar tais tarefas. Assim, é bastante improvável e faz pouco sentido para o negócio, desenvolver sistemas confiáveis de IA de propósito geral em que

[88] IBM. (n.d.). *Vector Embedding*. Disponível em: https://www.ibm.com/br-pt/think/topics /vector-embedding. Acesso em 29 de janeiro de 2025.
[89] GUADAMUZ, Andres. A scanner darkly: Copyright liability and exceptions in artificial intelligence inputs and outputs. GRUR International, v. 73, n. 2, p. 111-127, 2024.

seja possível resgatar informações pessoais, a partir do modelo matemático empregado. Até porque, quanto mais confiável o sistema, melhor desempenha seu propósito de detectar padrões gerais e, portanto, mais descolado ele é da base de dados originária.

Com exceção de celebridades ou personas públicas, é bastante improvável que usuários consigam provocar as chamadas "regurgitações" de dados pessoais ou mesmo de obras autorais usados no treinamento, a não ser que utilizem nas requisições os próprios dados ou trechos de conteúdo proprietário, ou seja, as regurgitações são estratagemas artificiosos que pressupõem o conhecimento dos dados que se pretende extrair, não se tratando de uma revelação ou exposição.[90]

Por esse motivo, o EDPB, apesar de descrever os critérios formais de caracterização da anonimização, praticamente permite presumir que tais modelos não seriam congêneres a uma extração pelo controlador ou pelos usuários de dados pessoais.

Vale dizer, o propósito ou resultado pretendido por tais sistemas é detectar padrões gerais de escrita ou de qualquer conteúdo ou estilos, em geral, de obras intelectuais e a partir dos padrões registrados no modelo matemático treinado, gerar novos conteúdos. Por isso, o seu impacto ou a pretensão de sua ação é congênere, por concepção, à anonimização.

É claro que, em etapa anterior, há coleta e possível armazenamento de dados. Mas nas etapas subsequentes de desenvolvimento desses modelos, os dados são anonimizados, pois justamente não interessa ao seu propósito o que for específico ou individualizado. Da perspectiva do propósito ou do resultado, portanto, não cabe falar em danos à personalidade individual ou à proteção de dados pessoais. Esse resultado, normalmente, não será produzido. E uma vez reconhecida a natureza anonimizada da aplicação e dos resultados gerados pelo sistema, tem menor importância indagar sobre justificação do treinamento desses sistemas com base em legítimo interesse. Se o resultado é anonimizado, não há impacto sobre a projeção de personalidades individuais. Novamente, a precaução substantiva aponta para a inexistência de um objeto de proteção como resultado do processamento, que justifique a imposição de procedimentos restritivos ou de controle do tratamento.

Sendo independente de análise de legítimo interesse, o treinamento de sistemas de propósito geral pode processar inclusive dados sensíveis para detecção de padrões nos modelos. Por exemplo, um modelo multimodal treinado para descrever imagens de radiografias com vocabulário médico de modo a auxiliar o trabalho de profissionais de saúde, apesar de coletar dados de imagens radiografadas individuais e conteúdos de

[90] MARANHÃO, JULIANO. The New York Times versus OpenAI. Jota, 13 jan. 2024.

relatórios médicos e prontuários tem como propósito e resultado extrair padrões correlacionando aspectos da imagem com observações médicas, sendo irrelevantes aspectos da individualidade que identifiquem os indivíduos ou mesmo especificidades individualizadas de sua condição de saúde. Desse modo, respeitadas informações e medidas de transparência a respeito da fase de coleta e armazenamento para treinamento, não se vislumbram, dentro de uma análise de precaução substantiva, impactos sobre a personalidade individual desses sistemas.

Isso não significa dizer que a atuação da ANPD estaria excluída ou seria despicienda. A proteção deve consistir na plena transparência quanto aos dados coletados e seu processamento, assegurando-se que o modelo de IA e sua aplicação não identificará indivíduos, ainda que seus dados tenham feito parte do amplo volume de dados usados para treinamento.

A perspectiva procedimental, porém, assume que, a simples coleta e armazenamento, por serem atividades regradas no gerenciamento de riscos, já implicam, per se, violação à personalidade. Sobrevalorizar o procedimento implica assumir que a tarefa da proteção de dados seria gerenciar formalmente níveis aceitáveis de riscos sem se comprometer, efetivamente, com a proteção substantiva de direitos fundamentais, como preconiza o princípio de precaução substantiva.

Ao lado dessa consideração, dentro de uma análise que incorpora a proteção de dados, como valor social, à ética de IA responsável, cabe indagar sobre os impactos da restrição de acesso a dados sobre o desenvolvimento do sistema e dos benefícios da tecnologia.

Os grandes modelos fundacionais de texto, imagem, áudio e vídeos que ocuparam espaço de mercado são, em grande maioria, desenvolvidos por empresas estrangeiras globais, sendo que algumas delas, principalmente as líderes, buscam fazer a adaptação do modelo geral às especificidades de cada país, ou mesmo regionais ou locais, o que significa a necessidade de refinamento do modelo com treinamento a partir de conteúdo local. O impacto desse refinamento dos modelos para adequação ao contexto brasileiro tem o potencial de incorporar aspectos gerais da cultura nacional ou regional e evitar visões estereotipadas que o modelo tenha incorporado a partir de conteúdo estrangeiro, o que é um traço do chamado *colonialismo digital*. Ou seja, o treinamento com conteúdo local é benéfico do ponto de vista de mitigação do risco de estereótipos estrangeiros que podem levar a efeitos discriminatórios, sendo desejável adotar medidas para sua mitigação.[91]

[91] MOLLEMA, Warmhold Jan Thomas. Decolonial AI as Disenclosure. **Open Journal of Social Sciences**, v. 12, n. 2, p. 574-603, 2024.

Ilustra-se a imprecisão quanto a cultura local com a incapacidade de detectar expressões locais, como a falha do *chatbot* multimodal MetaAI em gerar imagens adequadas com a requisição *"homem aranha plantando bananeira"*, em que o output entrega o homem aranha plantando um pé de banana, em vez de posição de ponta cabeça apoiado pelas mãos.[92] Já o estereótipo decorrente da visão estrangeira sobre determinada localidade pode ser ilustrado por um sistema em que requisições de imagens de CEO mexicano entregam, predominantemente, imagens de homens brancos com bigodes, vestidos de terno. Portanto, o aumento de confiabilidade e adequação do modelo para as peculiaridades locais, por meio do treinamento com conteúdo nacional, traz consequências sociais positivas, seja para o usuário seja para a preservação da cultura e fomento à inclusão e não-discriminação.

Portanto, os riscos e benefícios promovidos pela inteligência artificial devem ser avaliados em uma perspectiva do resultado do sistema em termos dos valores sociais promovidos, dentre eles o valor social da privacidade e proteção de dados, seguindo o princípio precaução substantiva e não prevenção procedimental de riscos, considerados inerentes ao tipo de dado ou ao tipo de tratamento.

4.2.1 Analogia com Text and Data Mining na Legislação Autoral

Restrições da legislação ao uso de dados para treinamento de sistemas de IA traz preocupações quanto à perspectiva de desenvolvimento tecnológico. Boa parte do desenvolvimento de sistemas de IA dá-se no campo acadêmico com pesquisas voltadas ao desenvolvimento de novas ou aperfeiçoamento de metodologias, sem que o sistema seja aplicado ou traga quaisquer efeitos sobre a dimensão de direitos individuais relativos aos dados usados para treinamento.

Mesmo no âmbito da pesquisa e desenvolvimento da indústria, normalmente chegamos a travar contato apenas com sistemas bem-sucedidos, que são disponibilizados e comercializados no mercado, muito embora, para que se chegue a esse resultado, muitos sistemas e experimentações prévias são descartadas, sem qualquer uso ou resultado de aplicação do modelo experimental.

Ressalvadas as precauções e controles de comissões éticas na academia ou na indústria tais pesquisas científicas, experimentações e ensaios de desenvolvimento de sistemas de IA são cruciais para o avanço da tecnologia e não oferecem riscos significativos aos

[92] G1. *Meta AI de WhatsApp e Instagram gera respostas erradas, mas não é possível desativar recurso; veja como limitar.* 25 out. 2024. Disponível em: https://g1.globo.com/tecnologia/noticia/2024/10/25/meta-ai-de-whatsapp-e-instagram-gera-respostas-erradas-mas-nao-e-possivel-desativar-recurso-veja-como-limitar.ghtml.

titulares de direitos sobre os dados usados para treinamento. Assim, restrições ao uso desses dados como a exigência de autorização dos autores em relação a obras protegidas ou dos titulares sobre o uso de seus dados pessoais, podem inviabilizar esses exercícios intelectuais, inibindo o progresso tecnológico.

No campo da propriedade intelectual, já vem sendo debatidas e adotadas legislações que excepcionem a necessidade de autorização para uso de dados para treinamento, tanto para fins de pesquisa, quanto para fins de experimentações por organizações empresariais.

O Reino Unido foi um dos primeiros países[93] a excepcionar direitos de propriedade intelectual para o treinamento e desenvolvimento de modelos de inteligência artificial com conteúdo protegido, permitindo a reprodução de cópias para fins de pesquisa científica sem propósito comercial[94], desde que haja conhecimento suficiente pelo sujeito detentor de direitos. [95]

Recentemente, acentuaram-se os debates sobre se essa previsão permitiria o uso comercial dessas cópias por terceiros com fins comerciais.[96] É comum que base de dados utilizadas para pesquisas subsidiadas tenham de ser publicizadas. Nesse caso, companhias privadas reutilizam essas bases de dados para desenvolvimento de seus próprios modelos. Nesse contexto, após um procedimento de consulta pública com a finalidade de compreender aspectos entre propriedade intelectual e inteligência artificial, o *Intellectual Property Office* do Reino Unido (UKIPO) havia optado por expandir a exceção acima, de forma a permitir a cópia de obras protegidas por *text and data mining* inclusive para usos comerciais,[97] entretanto, a iniciativa foi interrompida em razão de críticas, como a emitida pelo Comitê Digital e de Comunicações da Casa dos Lordes do Reino Unido.[98]

---

[93] FIIL-FLYNN, Sean M. et al. Legal reform to enhance global text and data mining research. Science, v. 378, n. 6623, p. 951-953, 2022.

[94] HARGREAVES, Ian. Digital opportunity. A review of intellectual property and growth, v. 5, 2011.

[95] Copyright, Designs and Patents Act 1988, Artigo 29A.

[96] GUADAMUZ, Andres. A scanner darkly: Copyright liability and exceptions in artificial intelligence inputs and outputs. GRUR International, v. 73, n. 2, p. 111-127, 2024.

[97] REINO UNIDO. Artificial intelligence and intellectual property: copyright and patents - government response to consultation. [S. l.], 28 jun. 2022. Disponível em: https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents/outcome/artificial-intelligence-and-intellectual-property-copyright-and-patents-government-response-to-consultation. Acesso em: 24 mar. 2025.

[98] REINO UNIDO. House of Lords. Communications and Digital Committee. Artificial intelligence and the creative industries: summary of conclusions and recommendations. Londres: Authority of the House of Lords, 2023. Disponível em: https://publications.parliament.uk/pa/ld5803/ldselect/ldcomm/125/12502.htm. Acesso em: 24 mar. 2025.

Em função disso, o UKIPO procedeu a novo processo de consulta pública[99], ainda em discussão, em que se considera a possibilidade de adotar um modelo semelhante ao da União Europeia[100], em que a exceção abrangeria fins comerciais, mas que os detentores de direitos poderiam se opor à cópia de suas obras por mecanismos de *opt-out*.[101]

Na União Europeia, a legislação sobre direitos autorais está ancorada na Diretiva 2001/29/CE, que estabelece uma série de exceções à proteção dos direitos autorais, especialmente quando se trata do desenvolvimento de pesquisas. A Diretiva permite o uso de obras protegidas para fins de pesquisa científica sem a necessidade de autorização do titular dos direitos autorais, desde que o uso seja "não comercial" e a obra não seja utilizada para fins lucrativos (artigo 5º, nº 3, alínea b)[102]. A partir da reforma da Lei de Copyright da UE, e considerando a emergência das tecnologias de inteligência artificial, surgiram discussões acerca da flexibilização desses direitos para permitir que os algoritmos de IA possam treinar modelos sem a necessidade de autorização dos detentores dos direitos autorais[103]. Estas discussões resultaram na *Directive on Copyright in the Digital Single Market* (Diretiva 790/2019) que, em seu artigo 4º[104], trata especificamente da exceção para a análise de texto e dados (*text and data mining* – TDM) e permite que as obras protegidas sejam utilizadas para fins de treinamento de IA sem a necessidade de autorização dos titulares dos direitos autorais para fins de análise de dados, incluindo tanto para pesquisa científica quanto para finalidades comerciais, não obstante os autores mantenham seus direitos sobre as obras e materiais protegidos, desde que tenha expressamente se manifestado, como, por exemplo, por meio de mecanismos de leitura por máquina quando a obra for publicamente disponível na *internet*

---

[99] Conferir mais detalhes em https://www.gov.uk/government/consultations/ copyright-and-artificial-intelligence.

[100] Diretiva (EU) 2019/790, Artigo 4º.

[101] REINO UNIDO. Intellectual Property Office. Artificial intelligence and copyright: consultation. Newport: Intellectual Property Office, 2024. Disponível em: https://assets.publishing.service.gov.uk/media/6762c95e3229e84d9bbde7a3/241212_AI_and_Copyright_Consultation_print.pdf. Acesso em: 24 mar. 2025.

[102] "Os Estados-membros poderão prever exceções ou limitações aos direitos de reprodução e comunicação pública, no caso de uso de obras para fins de pesquisa científica não comercial."

[103] Em 2021, a Comissão Europeia publicou uma Consulta Pública sobre a Revisão da Legislação de Direitos Autorais da UE com foco na necessidade de adaptar as regras de direitos autorais à era digital. A consulta envolveu discussões sobre o uso de dados para treinamento de IA e o impacto das tecnologias emergentes na proteção de direitos autorais.

[104] "Os Estados-Membros poderão prever uma exceção ou limitação ao direito de reprodução e ao direito de comunicação pública para permitir a utilização de obras, identificadas ou não, por qualquer pessoa, para análise de texto e dados com fins de pesquisa científica ou comercial, desde que a utilização seja realizada por qualquer meio que não infrinja substancialmente os interesses legítimos do autor."

4.3. Terceiro Vício: A Supervalorização da Máquina no Processo Decisório

Outro problema trazido pela visão procedimental, que reduz a IA a um complexo de processamento computacional de dados, está na visão de que sistemas de IA seriam apenas sistemas informáticos, ignorando-se o envolvimento dos humanos e de seu papel na tomada de decisões.[105] Tal visão tende a ver quaisquer *outputs* de sistemas de IA como espécies de decisões automatizadas, ainda que se limitem a estabelecer predições ou fazer recomendações.

A noção de decisão automatizada precisa ser compreendida não apenas a partir das funcionalidades técnicas de um sistema específico de IA, mas sim à luz de sua natureza sociotécnica,[106] que engloba, por um lado, o componente técnico, ou seja, o *software* concebido e desenvolvido como produto informático, implantado de forma autónoma (*stand-alone*) ou incorporado em equipamentos, e, por outro, o componente humano, ou seja, as pessoas envolvidas no ciclo de vida do *software* de inteligência artificial *autônomo* ou incorporado, desde a sua concepção, passando pelo seu desenvolvimento, teste, validação e monitoramento. Essa perspectiva sociotécnica pressupõe que o papel do sistema informático seja analisado com atenção às práticas organizacionais, especialmente no que diz respeito à presença ou ausência de discricionariedade humana na tomada da decisão final.

Tal aspecto tem particular relevo na discussão sobre a interpretação do conceito de "decisão automatizada", da qual se derivam obrigações e direitos na legislação de proteção de dados, como no art. 22[107] do GDPR e no art. 20 da LGPD. Vale notar que há diferenças significativas entre ambas as legislações a esse respeito. [108]

Na GDPR prevê-se expressamente o direito do sujeito de dados não estar sujeito a decisões integralmente automatizadas, com algumas exceções, entre elas o consentimento do titular. Já na LGPD prevê-se apenas o direito a revisão de decisões automatizadas, o que implicitamente implica o reconhecimento de um direito dos controladores a empregar decisões automatizadas.

---

[105] KUDINA, Olya; VAN DE POEL, Ibo. A sociotechnical system perspective on AI. Minds and Machines, v. 34, n. 3, p. 21, 2024.

[106] AKBARIGHATAR, Pouria; PAPPAS, Ilias; VASSILAKOPOULOU, Polyxeni. A sociotechnical perspective for responsible AI maturity models: Findings from a mixed-method literature review. International Journal of Information Management Data Insights, v. 3, n. 2, p. 100193, 2023.

[107] Article 22 (1) The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

[108] ALMADA, Marco; MARANHÃO, Juliano. Contribuições e limites da lei geral de proteção de dados para a regulação da inteligência artificial no brasil. Revista direito público, v. 20, p. 385-413, 2023.

Por outro lado, o art. 22 da GDPR aplica-se a decisões automatizadas que incluam o perfilamento pessoal. Já a letra do art. 20 da LGPD traz direitos de explicação e revisão de decisões automatizadas *"incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade"*. Além disso, o art. 22 da GDPR refere-se apenas a decisões que *afetem significativamente os direitos dos titulares ou interesses equiparáveis*, assegurando, apenas em relação a estas, o direito à revisão humana. Por sua vez, o art. 20 da LGPD fala em quaisquer decisões que afetem interesses, sem prever o direito à revisão.

Uma leitura literal e procedimental do art. 20 da LGPD poderia considerar como decisão automatizada as próprias predições e classificações de um indivíduo em determinado perfil, ainda que tal classificação não implique qualquer efeito sobre direitos ou interesses do sujeito de dados.

Tal interpretação, porém ignora a dimensão sociotécnica e a consideração acerca do impacto de tais sistemas na dimensão individual da personalidade, como reza o princípio de precaução substantiva. O direito à revisão, ou mesmo à revisão humana, ou o direito à explicação (dos critérios que levam ao output resultante do sistema), que formam o núcleo dos artigos 22 da GDPR e 20 da LGPD, fazem sentido diante de consequências sobre a esfera individual decorrentes do processamento computacional. Ocorre que, se o processamento não causa impacto direto sobre a esfera individual, mas serve apenas como subsídio para decisão discricionária humana, então não cabe exigir tais direitos. Assim, por exemplo, sistema que classifica currículos pessoais em critérios relevantes para o contratante, fazendo recomendações ou classificações dos melhores currículos, não é, em si, uma decisão automatizada, mas uma informação relevante para a tomada de decisão. Se a decisão sobre contratação for discricionária e adotada por humano, não há a sujeição do candidato à máquina. Uma vez acoplada tal predição com uma determinação automática de contratação, temos então uma decisão integralmente automatizada.

Para interpretar a disposição presente nos artigos 22 da GDPR e 20 da LGPD, dentro da abordagem de precaução substantiva e a partir dos resultados do sistema de IA, deve-se examinar exatamente o que significa uma decisão individual inteiramente automatizada. A letra do art. 20 refere-se a *"decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses"*, podendo-se inferir que os interesses afetados são aqueles concernentes ao próprio sujeito de dados em questão.

Decisão individual é aquela que determina uma posição ou atitude com relação a alguém com efeitos vinculantes significativos sobre a esfera de tal pessoa. [109] Assim, como consequência desse ato, o indivíduo dever ter sua esfera de direitos ou interesses de alguma forma afetados. A GDPR exige ainda que tais decisões tenham efeitos significativos, o que, de acordo com as orientações sobre decisões automatizadas do Article 29 Working Party,[110] significa a capacidade de influenciar substancialmente as circunstâncias, o comportamento ou as escolhas dos indivíduos envolvidos, gerar um impacto prolongado ou permanente sobre o titular dos dados ou levar à exclusão ou discriminação das pessoas afetadas.

Ou seja, da perspectiva da proteção substantiva de direitos, a aplicação do art. 20 da LGPD não pode contemplar procedimentos inteiramente automatizados que não constituam "decisões individuais". Assim, se a análise de empréstimo bancário se baseia em sistema de *credit scoring*, a pontuação gerada pelo sistema não é uma decisão individual, mas, a depender do grau de discricionariedade e das práticas decisórias, pode ser apenas uma informação relevante para a tomada de decisão humana, que consistirá na concessão ou não do empréstimo.

Para que se tenha uma decisão integralmente automatizada, por exemplo, de concessão ou não de crédito, é preciso examinar a IA como sistema sociotécnico para entender o papel e o grau de intervenção humana no resultado final. Ou seja, decisões *"tomadas unicamente com base em tratamento automatizado de dados pessoais"* compreende apenas aquelas decisões em que não há participação humana significativa no processo, ou seja, o ser humano envolvido não deve simplesmente aceitar o resultado automatizado, mas deve ter autoridade e competência para influenciar a decisão, considerando todos os dados relevante. Tal análise é contextual, com a verificação de elementos fáticos sobre o grau de participação humana e sua discricionariedade, não cabendo dela prescindir para se classificar de antemão tipos de sistemas como integralmente automatizados.

Nessa linha, a Corte de Justiça Europeia[111] em caso envolvendo a negativa de concessão de crédito com base em *credit scoring*, considerou tal decisão como integralmente automatizada apenas após verificar que no contexto específico de decisões adotadas não houve participação efetiva e discricionariedade humana. A partir desse precedente, autoridades nacionais de proteção de dados passaram a considerar uma série de fatores

[109] PALMIOTTO, Francesca. When is a decision automated? A taxonomy for a fundamental rights analysis. German law journal, v. 25, n. 2, p. 210-236, 2024.
[110] ARTICLE 29 DATA PROTECTION WORKING PARTY. Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679. 2018.
[111] C-634/21, Schufa Holding, ECLI:EU:C:2023:957, (Dec. 7, 2024).

para definição de uma decisão como integralmente automatizada, dentre eles: (i) a estrutura organizacional da empresa na qual é empregado o sistema de IA, (ii) linhas de reporte sobre o conteúdo das decisões, (iii) cadeias de aprovação e grau de discricionariedade dentro das atribuições de competência da organização, (iv) presença de decisões divergentes à recomendação ou predição do sistema de IA, (v) treinamento efetivo da equipe, (vi) políticas internas de revisão.[112]

Portanto, a constatação e reconhecimento de uma decisão como integralmente automatizada, conforme prevista pela LGPD, pressupõe a análise detalhada da aplicação do sistema de IA em sua consideração sociotécnica, em termos de responsabilidade organizacional e grau efetivo de intervenção humana. Mesmo decisões aparentemente automatizadas, mas que representem uma etapa do processo decisório, precisam ser analisadas dentro de toda a estrutura de uma organização para a tomada de decisão final que tenha o potencial de impactar uma personalidade individual. Assim, por exemplo, o cadastramento ou descadastramento de usuários de uma plataforma pode ser feito de modo automatizado, simplificando-se e propiciando maior celeridade ao ingresso do usuário (prestador ou tomador de serviço; vendedor ou consumidor), o que seria inviável com um cadastramento pessoal ou mesmo com revisão humana de cada cadastro. Porém, em caso de negativa de cadastramento ou descadastramento, que pode afetar negativamente o interesse individual do usuário, pode haver na organização em questão, uma revisão ou "recurso" já previsto a humano ou departamento. Cabe, nesses casos, a análise contextual, pelo exercício de fato da revisão, para se determinar se de fato se trata de uma decisão automatizada, ou de decisão discricionária humana com um processo inicial de filtragem por máquina.

4.4. Quarto Vício: Confundir Supervisão Humana com Revisão

Diretamente relacionado ao vício de se supervalorizar o papel da máquina, sem considerar a dimensão sociotécnica de sistemas de IA, está a confusão entre a exigência de revisão humana de uma decisão inteiramente automatizada e o *desideratum* de supervisão humana no ciclo de vida de sistemas de IA. A distinção entre esses conceitos é fundamental, mas por vezes mal compreendida, o que compromete tanto a proteção efetiva de interesses coletivos relevantes no ambiente digital, quanto o avanço tecnológico.

[112] PALMIOTTO, Francesca. Op. Cit.

Embora os mecanismos de revisão sejam importantes para tratar questões individuais, o foco em uma supervisão humana robusta na concepção e no monitoramento de sistemas de IA está alinhado com o objetivo de garantir práticas seguras e éticas no uso da tecnologia.

Importante lembrar que sistemas de IA não necessariamente são desenvolvidos para simular determinada habilidade, ação ou pensamento humano, nem voltados para substituir capacidades de humanos individual ou coletivamente considerados. Sistemas de IA podem objetivar ideais de comportamento, ou, em oposição, buscar se aproximar do comportamento humano efetivo. Assim, por exemplo, *chatbots* que interagem em conversações em áudio simulam ações cujo objetivo é aproximar-se da forma como os humanos dialogam e se expressam, o que pode incluir pausas para respiração e reflexão, hesitações, como meio para se propiciar conforto na interação humano-máquina. Já um sistema voltado para operar veículos autônomos não objetiva se aproximar do comportamento padrão de humanos na direção, mas seguir um ideal de comportamento, o que inclui o respeito estrito às regras de trânsito e capacidade de visão e tempo de reação bastante superiores às humanas.

Assim, para diversas aplicações, em que se objetiva um ideal de ação para produção de resultados eficientes, a revisão humana sequer se coloca. Por exemplo, o acionamento automático de *air bag* em veículos pode ter seu desempenho otimizado com uso de IA, a partir da previsão do impacto, o que não comporta, por motivos óbvios, a revisão humana. Tal consideração abrange não só tarefas idealizadas inalcançáveis pelas habilidades humanas, como também tarefas que, embora possam ser desempenhadas por humanos, implicariam esforço insuperável de uniformização ou coordenação de revisores.

Tome-se como exemplo os sistemas utilizados por plataformas de intermediação de serviços de transporte de passageiros para distribuir as chamadas de corridas entre os motoristas cadastrados. A decisão de distribuição das corridas é realizada com base em algoritmo desenvolvido para trazer maior eficiência logística e aumentar a qualidade do serviço. Embora, em tese, humanos possam realizar a tarefa de gestão e direcionamento das corridas, uma plataforma com centenas de milhares de motoristas cadastrados e milhões de usuários implica expressiva granularidade para alocação de corridas que deve empregar critérios uniformes para assegurar qualidade e eficiência do serviço. Para alcançar tal granularidade e especificação do direcionamento, seria necessário, em primeiro lugar, equipe numerosa de designadores das corridas, cada um capaz de processar número limitado de chamadas, e, em segundo lugar, gestores e coordenadores

para que não ocorra sobreposição ou falta de motorista designado para alguma chamada. Tais equipes gestoras deveriam ser responsáveis pela uniformização dos critérios de alocação, que são instanciados por humanos necessariamente com algum grau de subjetividade, pelo monitoramento das designações para evitar conflitos ou lacunas entre corridas. Além de tal estrutura organizacional implicar custos proibitivos, seria de difícil coordenação e uniformização, potencialmente inviabilizando o negócio de intermediação online de corridas.

O serviço de intermediação por plataformas tem eficácia e alcançaram sucesso justamente por trazer uma solução tecnológica para o problema de coordenação entre pares para alocação dos prestadores e usuários, de modo que não faz sentido reinserir a dificuldade de coordenação por meio de uma exigência de revisão humana de cada designação de um serviço. Vale considerar, por outro lado, dentro de análise de precaução substantiva, que o impacto de tais decisões de alocação, em termos de privacidade e proteção de dados pessoais, é pouco significativo para os usuários e mesmo para motoristas.

Se a revisão humana pode ser inviável ou mesmo contraproducente, é sempre desejável a supervisão do sistema de IA durante o seu ciclo de vida e em particular no monitoramento dos resultados de sua aplicação. Wimmer e Doneda (2021) consideram um conceito mais abrangente de "intervenção humana", que compreende os direitos à revisão e à explicação de decisões automatizadas, mas que *"pode se materializar por meio da participação de agentes humanos nos processos de tomada de decisão algorítmica de distintas formas e em diferentes momentos do ciclo de vida do sistema"*.[113] Ou seja, a "intervenção humana" pode ser entendida como conceito abrangente que inclui *revisão*, *explicação* e *supervisão humana*. Se, de um lado, a LGPD estabelece somente os direitos à revisão e à explicação,[114] o conceito de supervisão humana ganhou destaque no âmbito da regulação de IA.

A revisão de decisões automatizadas estabelecida na LGPD é, por natureza, uma ferramenta *ex post* disponível ao titular de dados sujeito a decisão automatizada, com a finalidade de garantir a contestabilidade da decisão. Já a supervisão humana ("*human oversight*") refere-se ao envolvimento de humano(s) no processo algorítmico. A supervisão humana é frequentemente associada à ideia do "*human-on-the-loop*", em que o elemento humano é responsável por monitorar o processo de decisão algorítmica e tem

[113] WIMMER, Miriam; DONEDA, Danilo. "Falhas de IA" e a Intervenção Humana em Decisões Automatizadas: Parâmetros para a Legitimação pela Humanização. Direito Público, *[S. l.]*, v. 18, n. 100, 2022. DOI: 10.11117/rdp.v18i100.6119. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/6119. Acesso em: 6 mar. 2025.
[114] LGPD, Art. 20.

a possibilidade de intervir a qualquer momento.[115] A diferença entre os dois conceitos é evidenciada nos Princípios para Decisões Automatizadas elaborados pelo *European Law Institute*, que institui um princípio para cada conceito:[116]

Princípio Orientador nº 9: **Supervisão/ação humana**	Princípio Orientador nº 10: **Revisão humana de decisões significativas**
O operador deve garantir supervisão humana razoável e proporcional sobre a **operação da decisão automatizada**, levando em consideração os riscos envolvidos e os direitos e interesses legítimos potencialmente afetados pela decisão.	A revisão humana de **decisões significativas selecionadas** com base na relevância dos efeitos legais, na irreversibilidade de suas consequências ou na gravidade do impacto sobre direitos e interesses legítimos deve ser disponibilizada pelo operador.

Fonte: European Law Institute. Tradução Livre. Grifo nosso.

Enquanto a revisão humana é disponibilizada pelo operador ao usuário ou sujeito de dados afetado pela decisão individual automatizada, com a finalidade de oferecer meios para contestação da decisão, a supervisão humana cumpre outro papel. Note-se que a exigência de supervisão humana para sistemas de alto risco no AI Act europeu não obriga os operadores a disponibilizar meios de ação para os afetados, mas apenas que sistemas de alto risco sejam desenvolvidos de forma a possibilitar supervisão humana[117] pelos próprios operadores, o que deve ser feito por pessoas capacitadas seja nos aspectos técnico informáticos, seja no domínio de aplicação.[118]

Fink (2025) entende que tal previsão cumpre duas finalidades: (i) a garantia de adequação do *output* e (ii) garantia de adequação do processo decisório.[119] Quanto ao *output*, o objetivo é que o humano seja capaz de detectar erros, imprecisões ou falhas do sistema para adotar medidas de mitigação, tendo em vistas potenciais impactos do sistema sobre direitos e valores sociais. A supervisão pode envolver a intervenção para correção do

[115] Ibid.
[116] EUROPEAN LAW INSTITUTE. Op. cit. Tradução livre.
[117] AI Act, Art. 14.
[118] AI Act, Art. 26.
[119] FINK, Melanie, op. cit., pp. 7-9.

output, mas em contextos nos quais habilidades humanas sejam necessárias (e.g. contexto social ou empatia). Mesmo nesses casos, a organização que opera o sistema de IA deve estruturar procedimentos para evitar falsos negativos e positivos quanto à intervenção humana, que pode comprometer a eficiência do sistema de IA.[120] Já quanto ao processo, o objetivo seria aumentar a confiabilidade dos sistemas, tanto da perspectiva dos indivíduos afetados (maior confiança em razão do envolvimento humano) quanto dos tomadores de decisão (maior agência e discricionariedade pela possibilidade de intervir).[121]

Conforme esclarece o European Law Institute, a supervisão humana não pode ser confundida com a intervenção caso a caso, o que desnaturaria o processo de automação, devendo consistir antes no monitoramento para que o sistema seja ajustado para atingir com maior eficiência e precisão o seu propósito.[122]

Assim, na concepção de uso responsável da IA, o foco deve estar na exigência de supervisão humana em decisões-chave no desenvolvimento e monitoramento do emprego de sistemas de IA. A supervisão humana efetiva não requer envolvimento integral no funcionamento do sistema, muito menos a revisão pontual de decisões automatizadas por um humano, o que comprometeria os benefícios da automação, como redução de custos, eficiência e ganhos de escala.[123]

---

[120] Ibid, p. 13.

[121] "From the perspective of individuals affected by AI systems, the addition of a human can safeguard procedural rights, including the right to a reasoned decision, the right to be heard, or the right to an effective remedy.41 Affected individuals may also simply trust a process more or only feel 'seen' and treated with dignity when other humans are involved. However, many process-oriented goals focus on the perspective of the individual who interacts with the AI system, such as a decision maker or someone using a chatbot. Human oversight requirements create space for the exercise of human judgment, protecting human agency and autonomy. This has been highlighted as particularly important in the public sector where discretion creates room for the decision-maker to take into account context and respond to novel, marginal, and individual circumstances, and thus needs to be preserved." Ibid, p. 8.

[122] EUROPEAN LAW INSTITUTE. Op. cit.

[123] "Human oversight should not compromise the benefits in cost reduction, effectiveness, and economies of scale gained by introducing automation." European Law Institute, op. cit., p. 22.

5. CONSIDERAÇÕES FINAIS

Sistemas de inteligência artificial trazem importante desafio para a interpretação da legislação de proteção de dados pessoais, em função do poder de inferência dessa tecnologia, o que torna bastante complexo o controle de finalidade do uso dos dados.

Tal possibilidade faz com que procedimentos de governança prescritos legalmente se tornem obsoletos para efetivamente garantir a proteção da personalidade individual ou da privacidade como valor social, exigindo medidas efetivas para promover o direito fundamental à autodeterminação informacional, a não-discriminação no ambiente digital e uma esfera pública democrática.

Categorizar dados, por exemplo, como pessoais ou não pessoais, sensíveis ou não sensíveis, pelo seu tipo ou pelo procedimento de tratamento, pode deixar de fazer sentido uma vez que não é o tipo, mas o uso e o efeito do uso pela IA do dado para individualizar ou produzir resultados que impactam direitos individuais.

Daí a inadequação em se encarar sistemas de IA como conjuntos de processamento de dados, e a necessidade de concebê-los como *agentes sociotécnico* que produzem resultados no ambiente virtual e físico, observando seus impactos efetivos sobre a personalidade individual e a privacidade como valor social, de modo a garantir proteção concreta de direitos individuais e coletivos.

Ao mesmo tempo, o excessivo rigor procedimental deixa de fazer sentido quando o desenvolvimento da IA ou a operação da IA como agente sociotécnico não resultar em impacto substantivo sobre a personalidade individual e a privacidade como valor social.

O princípio de precaução substantiva, ao enfatizar a análise do impacto dos sistemas de IA (como agentes) sobre a privacidade, tal como desenvolvido neste documento, propõe-se a orientar a interpretação da legislação de proteção de dados pessoais para, de um lado, oferecer essa proteção efetiva quando houver ameaça concreta a direito fundamental e reduzir constrangimentos procedimentais desnecessários quando a ameaça não estiver presente ou não for significativa, permitindo-se o desenvolvimento tecnológico e a inovação.

A precaução substantiva também orienta a ponderação sobre os valores sociais e direitos individuais em torno do desenvolvimento e uso responsável da Inteligência Artificial de modo que permite integração mais harmônica entre a legislação de proteção de dados e a ética de IA.

Em particular, mostramos como a precaução substantiva pode superar alguns vícios interpretativos na leitura procedimental da legislação de proteção de dados, observando-se com rigor os impactos sobre a personalidade e sobre a privacidade como bem coletivo,

nas fases de treinamento e desenvolvimento de sistemas de IA e na fase de inferência e operação desses sistemas.

O vício em se exigir justificativa causal para cada ponto de dado usado no treinamento ou aplicação da IA pode ser superado com a pergunta sobre o momento no qual a aplicação do dado pode afetar a personalidade individual, com a aplicação das salvaguardas correspondentes. Como visto, o risco apenas se apresenta na fase de inferência quando a aplicação do sistema pode implicar limitação ao acesso a bens e direitos, devendo-se concentrar aqui as medidas protetivas e não na fase de desenvolvimento do sistema.

O vício em se considerar cada etapa de tratamento como um fim em si a exigir medidas protetivas ignora o efeito concreto do desenvolvimento e aplicação do sistema de IA. Observar esses efeitos, à luz da precaução substantiva, permite adotar medidas procedimentais de governança protetivas apenas para usos arriscados, que impactem a personalidade individual.

O vício em se sobrevalorizar o papel da informática na tomada de decisão pela IA como agente resolve-se, à luz da precaução substantiva e da IA responsável, com a análise concreta do contexto, discricionariedade dos humanos envolvidos e verificação de fato do exercício dessa discricionariedade, de modo a se propiciar uma leitura adequada dos papeis relevantes na interação humano-máquina que ocorre na IA como sistema sociotécnico.

Por fim, o vício em se exigir a revisão humana de toda e qualquer decisão automatizada ignora os diferentes papeis assumidos pela tecnologia na sua relação com humanos, quando a precaução substantiva e o uso responsável da IA recomendam a supervisão humana, que deve estar presente, na IA como sistema sociotécnico, em todo o ciclo de vida da IA.

BIBLIOGRAFIA

ABRUSIO, Juliana; MARANHÃO, Juliano; CAMPOS, Ricardo. Proteção de dados pessoais no STF e o papel do IBGE. Disponível em: https://www.irib.org.br/noticias/detalhes/artigo-undefined-conjur-a-protecao-de-dados-pessoais-no-stf-e-o-papel-do-ibge-undefined-por-juliano-maranhao-ricardo-campos-e-juliana-abrusio. Acesso em: 01.02.2025.

AKASH, Mamidi Sai. Artificial intelligence & neural networks. International Journal of Soft Computing and Artificial Intelligence, v. 3, n. 2, p. 37-44, 2015.

AKBARIGHATAR, Pouria; PAPPAS, Ilias; VASSILAKOPOULOU, Polyxeni. A sociotechnical perspective for responsible AI maturity models: Findings from a mixed-method literature review. International Journal of Information Management Data Insights, v. 3, n. 2, p. 100193, 2023.

ALALOUL, Wesam Salah; QURESHI, Abdul Hannan. Data processing using artificial neural networks. In: Dynamic data assimilation-beating the uncertainties. IntechOpen, 2020.

ALBERS, Marion. Informationelle selbstbestimmung. Baden-Baden: Nomos, 2005, p. 212.

ALBERNS, Marion. Informationelle Selbstbestimmung. Baden-Baden, 2005, p. 86 ss.

ALEMANHA. Tribunal Constitucional Federal. Decisão de 15 de dezembro de 1983. BVerfGE 65, 1 (42).

ALHADEFF, Joseph; VAN ALSENOY, Brendan; DUMORTIER, Jos. The accountability principle in data protection regulation: origin, development and future directions. In: Managing privacy through accountability. London: Palgrave Macmillan UK, 2012. p. 49-82.

ALMADA, Marco; MARANHÃO, Juliano. Voice-based diagnosis of covid-19: ethical and legal challenges. International Data Privacy Law, v. 11, p. 63-75, 2021.

ALMADA, Marco; MARANHÃO, Juliano. Contribuições e limites da Lei Geral de Proteção de Dados para a regulação da inteligência artificial no Brasil. Revista Direito Público, v. 20, p. 385-413, 2023.

ALMADA, Marco; PETIT, Nicolas. The EU AI Act: Between the rock of product safety and the hard place of fundamental rights. Common Market Law Review, v. 62, n. 1, p. 85–120, 2025

ANDERSON, Chris. The end of theory: The data deluge makes the scientific method obsolete. Wired magazine, v. 16, n. 7, p. 16-07, 2008.

ARTICLE 29 DATA PROTECTION WORKING PARTY. Guidelines on Automated Individual Decision-Making and Profiling for the Purposes of Regulation 2016/679. 2018.

AUGSBERG, S.; ULMSTEIN, U. Requisitos de consentimento modificados: o direito de proteção de dados pode aprender com o direito da saúde. In: CAMPOS, R.; ABBOUD, G.; NERY JR., N. (Org.). Proteção de dados e regulação. São Paulo: Thomson Reuters, 2020.

AUSTRALIAN INFORMATION COMMISSIONER. Privacy update on the COVIDSafe App. Disponível em: https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/covid-19/privacy-update-on-the-covidsafe-app.

BENNET, C. J.; RAAB, Charles D. The governance of privacy: policy instruments in global perspective. Cambridge: MIT Press, 2006.

BRASIL. Autoridade Nacional de Proteção de Dados. Despacho Decisório nº 20/2024. Diário Oficial da União: seção 1, Brasília, DF, 2024. Disponível em: https://www.in.gov.br/en/web/dou/-/despacho-decisorio-n-20/2024/pr/anpd-569297245. Acesso em: 13 mar. 2025.

BRITZ, Gabriele. Freie Entfaltung durch Selbstdarstellung, Tübingen (Mohr Siebeck), 2007, 93 S. In: Kritische Justiz, n. 4 (2008), p. 473-475.

BRITZ, Gabriele. Informationelle Selbstbestimmung zwischen rechtswissenschaftlicher Grundsatzkritik und Beharren des Bundesverfassungsgerichts. In: Hoffmann-Riem, W. (Org.). Offene Rechtswissenschaft. Tübingen, 2010, p. 561-596.

BULL, Hans Peter. Sinn und Unsinn des Datenschutzes. Tübingen, 2015, p. 27 ss – protege-se a pessoa, ou a personalidade individual contra os efeitos do uso ilegítimo da informação na esfera pública.

BURRELL, J. How the machine thinks: understanding opacity in machine learning algorithms. Big Data & Society, v. 3, 2016. Disponível em: https://doi.org/10.1177/2053951715622512. Acesso em: 3 abr. 2025.

CALANDRA, Davide et al. Artificial Intelligence to fight COVID-19 outbreak impact: an overview. European Journal of Social Impact and Circular Economy, v. 1, n. 3, p. 84-104, 2020.

CHEN, Daniel L.; PARTHASARATHY, Adithya; VERMA, Shivam. The Genealogy of Ideology: Predicting Agreement and Persuasive Memes in the US Courts of Appeals. 2017.

DENNET, Daniel. The intentional stance. MIT Press, 1989.

DUFOUR, Raimond et al. AI or more? A risk-based approach to a technology-based society. Oxford Business Law Blog, v. 2021, n. September 16, 2021.

EBERS, Martin. Standardizing AI – The Case of the European Commission's Proposal for an Artificial Intelligence Act. In: The Cambridge handbook of artificial intelligence: global perspectives on law and ethics, 2021.

EC HIGH LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE. Orientações éticas para uma IA de confiança. 2018. Disponível em: https://doi.org/10.2759/2686. Acesso em: 15 jul. 2024.

ERNST, Christian. Artificial intelligence and autonomy: self-determination in the age of automated systems. In: WISCHMEYER, T.; RADMACHER, T. (eds.). Regulating artificial intelligence. Springer, 2020. p. 53-73.

ENGSTROM, David Freeman; HO, Daniel E. Artificially Intelligent Government: A Review and Agenda. In: VOGL, Roland (org.). Big Data Law, forthcoming 2020. March 9, 2020.

EUROPEAN COMMISSION. Digital contact tracing: learning from the experiences of European countries. Brussels: European Commission, 2023. Disponível em: https://commission.europa.eu/system/files/2023-02/DigitalContactTracingStudy.pdf.

EUROPEAN COMMISSION. Impact assessment of the regulation on artificial intelligence. Bruxelas, 21 abr. 2021. Disponível em: https://digital-strategy.ec.europa.eu/en/library/impact-assessment-regulation-artificial-intelligence. Acesso em: 3 abr. 2025.

EUROPEAN DATA PROTECTION BOARD. Opinion 28/2024 on certain data protection aspects of the proposed European Health Data Space. Brussels: EDPB, 2024. Disponível em: https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282024-certain-data-protection-aspects_en.

EUROPEAN LAW INSTITUTE. Guiding Principles for Automated Decision-Making in the EU. ELI Innovation Paper. Disponível em: https://www.europeanlawinstitute.eu/projects-publications/publications/eli-innovation-paper-on-guiding-principles-for-automated-decision-making-in-the-eu/. Acesso em: 6 mar. 2025.

FICO, Bernardo; MARANHÃO, Juliano; VAINZOF, Rony. Riscos e oportunidades da regulamentação de decisões automatizadas e IA pela ANPD. Conjur, novembro de 2024. Disponível em: https://www.conjur.com.br/2024-nov-27/riscos-e-oportunidades-da-regulamentacao-de-decisoes-automatizadas-e-inteligencia-artificial-pela-anpd/. Acesso em: 6 mar. 2025.

FIIL-FLYNN, Sean M. et al. Legal reform to enhance global text and data mining research. Science, v. 378, n. 6623, p. 951-953, 2022.

FINK, Melanie. Human Oversight under Article 14 of the EU AI Act. Forthcoming in: Gianclaudio Malgieri, Gloria González Fuster, Alessandro Mantelero, and Gabriela Zanfir-Fortuna (eds), AI Act Commentary: A Thematic Analysis (Hart-Bloomsbury, forthcoming 2026). Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5147196. Acesso em: 6 mar. 2025.

G1. Meta AI de WhatsApp e Instagram gera respostas erradas, mas não é possível desativar recurso; veja como limitar. 25 out. 2024. Disponível em: https://g1.globo.com/tecnologia/noticia/2024/10/25/meta-ai-de-whatsapp-e-instagram-gera-respostas-erradas-mas-nao-e-possivel-desativar-recurso-veja-como-limitar.ghtml. Acesso em: 3 abr. 2025.

GELLERT, Raphaël. Data protection: a risk regulation? Between the risk management of everything and the precautionary alternative. Int'l Data Priv. L., v. 5, p. 3, 2015.

GRAEDEN, Ellie et al. An outcomes-based paradigm for data and AI regulation. Lawgorithm: Ibero-American Journal on Artificial Intelligence and Law, 2025. A ser publicado.

GUADAMUZ, Andres. A scanner darkly: Copyright liability and exceptions in artificial intelligence inputs and outputs. GRUR International, v. 73, n. 2, p. 111-127, 2024.

HALL, Emma. AI software achieves 100% melanoma detection rate. FMAI Hub, 17 out. 2023. Disponível em: https://www.fmai-hub.com/ai-software-achieves-100-melanoma-detection-rate/.

HARGREAVES, Ian. Digital opportunity. A review of intellectual property and growth, v. 5, 2011.

HOFFMANN-RIEM, W. Artificial Intelligence as a Challenge to Law and Regulation. In: WISCHMEYER, T.; RADMACHER, T. (eds.). Regulating Artificial Intelligence. Springer, 2020. p. 75 e ss.

HOFFMANN-RIEM, Wolfgang. Rechtliche Rahmenbedingungen, em Der neue Datenschutz. In: Bäumler, H. (Org.). Der neue Datenschutz. Neuwied/Kriftel: Luchterhand, 1998, p. 13.

HOLZINGER, A.; LANGS, G.; DENK, H.; ZATLOUKAL, K.; MULLER, H. Causability and explainability of artificial intelligence in medicine. WIREs Data Mining and Knowledge Discovery, 2019, v. 9.

IBM. Vector Embedding. [s.d.]. Disponível em: https://www.ibm.com/br-pt/think/topics/vector-embedding. Acesso em: 29 jan. 2025.

INFORMATION COMMISSIONER'S OFFICE. Apple and Google joint initiative on COVID-19 contact tracing technology. Wilmslow: ICO, 2020. Disponível em: https://ico.org.uk/media/about-the-ico/documents/2617653/apple-google-api-opinion-final-april-2020.pdf.

INFORMATION COMMISSIONER'S OFFICE. Guidance on AI and data protection. Wilmslow: ICO, 2023. Disponível em: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/artificial-intelligence/guidance-on-ai-and-data-protection/.

JOBIN, Anna; IENCA, Marcello; VAYENA, Effy. The global landscape of AI ethics guidelines. Nature Machine Intelligence, v. 1, n. 9, p. 389-399, 2019.

KESARI, Aniket et al. A legal framework for explainable artificial intelligence. Center for Law & Economics Working Paper Series, v. 9, 2024.

KRÖGER, Jacob Leon; LUTZ, Otto Hans-Martin; RASCHKE, Philip. Privacy Implications of Voice and Speech Analysis – Information Disclosure by Inference. In: FRIEDEWALD, Michael et al. (eds.). Privacy and Identity Management. Data for Better Living: AI and Privacy: 14th IFIP WG 9.2, 9.6/11.7, 11.6/SIG 9.2.2 International Summer School, Windisch, Switzerland, August 19–23, 2019, Revised Selected Papers. Springer International Publishing, 2020. p. 248.

KUDINA, Olya; VAN DE POEL, Ibo. A sociotechnical system perspective on AI. Minds and Machines, v. 34, n. 3, p. 21, 2024.

LEHR, D.; OHM, P. Playing with the data: what Legal Scholars Should Learn About Machine learning. UC Davis Law Review, v. 51, Dec. 2017.

LONDON, Alex John. Artificial intelligence and black-box medical decisions: accuracy versus explainability. Hastings Center Report, v. 49, n. 1, p. 15-21, 2019.

LORENZ, P.; PERSET, K.; BERRYHILL, J. Initial policy considerations for generative artificial intelligence. OECD Artificial Intelligence Papers, n. 1. Paris: OECD Publishing, 2023. Disponível em: https://doi.org/10.1787/fae2d1e6-en. Acesso em: 15 jul. 2024.

MARANHÃO, Juliano. Atributos de confiabilidade e segurança na governança de sistemas de Inteligência Artificial. In: BIONI, Bruno R.; CUEVA, Ricardo V. B.; MENDES, Laura S.; ALVES, Fabrício M. (Org.). Inteligência Artificial e Regulação. São Paulo: Editora Gen-Jurídico. No prelo.

MARANHÃO, Juliano. IA e o risco do medo. JOTA, 2023. Disponível em: https://www.jota.info/artigos/ia-e-o-risco-do-medo. Acesso em: 06 mar. 2025.

MARANHÃO, Juliano. O STF e a chave para os direitos no mundo pós-pandemia. Valor Econômico – O Globo, São Paulo, 12 maio 2021.

MARANHÃO, Juliano. The New York Times versus OpenAI. Jota, 13 jan. 2024.

MARANHÃO, Juliano; COZMAN, F. G.; ALMADA, M. Concepções de explicação e do direito à explicação de decisões automatizadas. In: VAINZOF, Rony; GUTIERREZ, Andrei (orgs.). Inteligência Artificial: Sociedade, Economia e Estado. São Paulo: Thomson Reuters Brasil, 2021. v. 1, p. 137-154.

MARANHÃO, Juliano Souza; CAMPOS, Ricardo Resende. Proteção de dados de crédito na Lei Geral de Proteção de Dados. Direito Público, v. 16, n. 90, 2019.

MARANHÃO, Juliano; NAVAS, João. Certificação como instrumento de regulação da Inteligência Artificial no AI Act. In: VAINZOF, Rony; GUTIERREZ, Andriei; GODINHO, Gustavo; KRASTINS, Alexandra (orgs.). Comentários ao EU AI Act: uma abordagem prática e teórica do Artificial Intelligence Act da União Europeia. 1. ed. São Paulo: Thomson Reuters Brasil, 2024. p. 259-279.

MELARKODE, N.; SRINIVASAN, K.; QAISAR, S. M.; PLAWIAK, P. AI-Powered Diagnosis of Skin Cancer: A Contemporary Review, Open Challenges and Future Research Directions. Cancers (Basel), v. 15, n. 4, p. 1183, 13 fev. 2023.

MITTELSTADT, B. From individual to group privacy in big data analytics. Philosophy & Technology, v. 30, n. 4, p. 475-494, 2017. https://doi.org/10.1007/s13347-017-0253-7.

MITTELSTADT, Brent; RUSSELL, Chris; WACHTER, Sandra. Explaining explanations in AI. In: Proceedings of the conference on fairness, accountability, and transparency, 2019. p. 279-288.

MOLLEMA, Warmhold Jan Thomas. Decolonial AI as Disenclosure. Open Journal of Social Sciences, v. 12, n. 2, p. 574-603, 2024.

PALMIOTTO, Francesca. When is a decision automated? A taxonomy for a fundamental rights analysis. German Law Journal, v. 25, n. 2, p. 210-236, 2024.

PORTUGAL. Comissão Nacional de Proteção de Dados. Orientações sobre os tratamentos de dados pessoais de saúde regulados no Decreto n.º 8/2020. Disponível em: https://www.cnpd.pt/media/1bbppegs/orientações_decreto_8_2020.pdf.

QUELLE, Claudia. Enhancing compliance under the general data protection regulation: the risky upshot of the accountability-and risk-based approach. European Journal of Risk Regulation, v. 9, n. 3, p. 502-526, 2018

REPORT OF THE DATA ETHICS COMMISSION OF THE FEDERAL GOVERNMENT. (2019). Federal Ministry of the Interior, Building and Community and Federal Ministry of Justice and Consumer Protection. Disponível em: https://datenethikkommission.de/wp-content/uploads/191128_DEK_Gutachten_bf_b.pdf.

REINO UNIDO. Artificial intelligence and intellectual property: copyright and patents - government response to consultation. [S. l.], 28 jun. 2022. Disponível em: https://www.gov.uk/government/consultations/artificial-intelligence-and-ip-copyright-and-patents/outcome/artificial-intelligence-and-intellectual-property-copyright-and-patents-government-response-to-consultation. Acesso em: 24 mar. 2025.

REINO UNIDO. House of Lords. Communications and Digital Committee. Artificial intelligence and the creative industries: summary of conclusions and recommendations. Londres: Authority of the House of Lords, 2023. Disponível em: https://publications.parliament.uk/pa/ld5803/ldselect/ldcomm/125/12502.htm. Acesso em: 24 mar. 2025.

REINO UNIDO. Intellectual Property Office. Artificial intelligence and copyright: consultation. Newport: Intellectual Property Office, 2024. Disponível em: https://assets.publishing.service.gov.uk/media/6762c95e3229e84d9bbde7a3/241212_AI_and_Copyright_Consultation_print.pdf. Acesso em: 24 mar. 2025.

RUGE, R. Die Gewährleistungsverantwortung des Staates und der Regulatory State. Berlin: Duncker & Humboldt, 2004.

SARTOR, Giovanni; LAGIOIA, Francesca. The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. Brussels: European Parliament, 2020. DOI: 10.2861/293.

SCHLINK, Bernhard. Die Amtshilfe: Ein Beitrag zu einer Lehre von der Gewaltenteilung in der Verwaltung. Berlim: Duncker & Humblot, 1982, p. 192.

SOLOVE, Daniel J. Data is what data does: regulating based on harm and risk instead of sensitive data. Northwestern University Law Review, v. 118, p. 1081, 2024.

THE ECONOMIST. A brief history—and future—of credit scores. The Economist, [s.l.], 6 jul. 2019. Disponível em: https://www.economist.com/international/2019/07/06/a-brief-history-and-future-of-credit-scores.

TUDOCELULAR. Coronavírus: Projeto Spira recebe mais voluntários para aumentar precisão de diagnóstico de insuficiência respiratória. Disponível em: https://www.tudocelular.com/seguranca/noticias/n157802/sistema-usp-diagnostico-insuficiencia-respiratoria-covid19-inteligencia-artificial-spira.html.

UNIÃO EUROPEIA. Diretiva 95/46/CE do Parlamento Europeu e do Conselho, de 24 de outubro de 1995. Relativa à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados. Jornal Oficial das

Comunidades Europeias, L 281, p. 31-50, 23 nov. 1995. Disponível em: https://eur-lex.europa.eu/eli/dir/1995/46/oj.

UNIÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016 (GDPR). Artigo 5(2).

UNIÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016 (GDPR). Artigos 4(7) e 24.

UNIÃO EUROPEIA. Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016 (GDPR). Artigo 5(2).

UNITED KINGDOM. Information Commissioner's Office. COVID-19 and information rights: reflections and lessons learnt from the Information Commissioner. November 2021. Disponível em: https://ico.org.uk/media/about-the-ico/documents/4019157/covid-19-report.pdf.

VESTING, Thomas. Freie Entfaltung durch Selbstdarstellung: Eine Rekonstruktion des allgemeinen Persönlichkeitsrechts aus Art. 2 Abs. 1 GG, 2008.

VON GRAFENSTEIN, Maximilian. The principle of purpose limitation in data protection laws. Nomos Verlagsgesellschaft mbH & Co. KG, 2018. Disponível em: https://www.nomos-elibrary.de/de/10.5771/9783845290843/the-principle-of-purpose-limitation-in-data-protection-aws.

VON LEWINSKI, Kai. Die Matrix des Datenschutzes: besichtigung und ordnung eines begriffsfeldes. Mohr Siebeck, 2014, p. 4-5.

WACHTER, Sandra; MITTELSTADT, Brent. A right to reasonable inferences: re-thinking data protection law in the age of big data and AI. Columbia Business Law Review, p. 494, 2019.

WARREN, Samuel; BRANDEIS, Louis. The right to privacy. In: Killing the Messenger: 100 Years of Media Criticism. Columbia University Press, 1989. p. 1-21.

WESTIN, Alan F. Privacy and freedom. Washington and Lee Law Review, v. 25, n. 1, p. 166, 1968.

WIKIPÉDIA: a enciclopédia livre. Match Garry Kasparov vs. Deep Blue. Disponível em: https://pt.wikipedia.org/wiki/Match_Garry_Kasparov_vs_Deep_Blue. Acesso em: 3 abr. 2025.

WIMMER, Miriam; DONEDA, Danilo. "Falhas de IA" e a intervenção humana em decisões automatizadas: parâmetros para a legitimação pela humanização. Direito Público, [S. l.], v. 18, n. 100, 2022. DOI: 10.11117/rdp.v18i100.6119. Disponível em:

https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/6119. Acesso em: 6 mar. 2025.

WISCHMEYER, Thomas. Artificial intelligence and transparency: opening the black box. In: WISCHMEYER, T.; RUXANDRA, C. (Ed.). Regulating Artificial Intelligence. Cham: Springer International Publishing, 2019. p. 75-101.

WU, Yu-chen; FENG, Jun-wen. Development and application of artificial neural network. Wireless Personal Communications, v. 102, p. 1645-1656, 2018.

ZANFIR-FORTUNA, Gabriela, Follow the (personal) Data: Positioning Data Protection Law as the Cornerstone of EU's 'Fit for the Digital Age' Legislative Package (March 15, 2024). EDPS at 20 Anniversary Volume, Forthcoming June 2024, Available at SSRN: https://ssrn.com/abstract=4794182 or http://dx.doi.org/10.2139/ssrn.4794182