

# Questões atuais sobre inteligência artificial: Riscos e emergência regulatória

Current issues about artificial intelligence: Risks and regulatory emergency



Recebimento em 03/12/2024

Aceito em 03/12/2024

**Flávio Dino de Castro e Costa<sup>1</sup>**  
<https://orcid.org/0009-0000-1115-8450>  
agenda.gmfd@stf.jus.br

**Estela Aranha<sup>2</sup>**  
estela.aranha@presidencia.gov.br

## RESUMO

Este artigo destina-se à análise de questões atuais e controversas, sob o ponto de vista jurídico, relacionadas ao avanço do uso da Inteligência Artificial (IA). Em especial, objetiva-se avaliar as possibilidades e os limites da regulação da IA no Brasil. O texto divide-se em três partes. A primeira delas consiste em uma aproximação conceitual ao tema. Em seguida, avança-se com a apresentação de alguns dos riscos do uso da IA em um contexto de desregulação. Por fim, avaliam-se as condições e os limites da regulação da IA no Brasil, com vistas a assegurar a promoção e proteção de direitos fundamentais. Conclui-se com a exposição dos motivos pelos quais é emergencial a regulação da IA pelo Estado.

**PALAVRAS-CHAVE:** Inteligência Artificial; Riscos; Regulação.

## ABSTRACT

This article is intended to analyze current and controversial issues, from a legal point of view, related to the advancement of the use of Artificial Intelligence (AI). In particular, the objective is to evaluate the possibilities and limits of AI regulation in Brazil. The text is divided into three parts. The first of them consists of a conceptual approach to the topic. Next, we proceed with the presentation of some of the risks of using AI in a context of deregulation. Finally, the conditions and limits of AI regulation in Brazil are assessed, with a view to ensuring the promotion and protection of fundamental rights. It concludes with the explanation of the reasons why the regulation of AI by the State is an emergency.

**KEYWORDS:** Artificial intelligence; Risks; Regulation.

## 1 INTRODUÇÃO

Na última década, a chamada Inteligência Artificial (IA) passou por um rápido processo de

---

<sup>1</sup> Autor convidado, Ministro do Supremo Tribunal Federal

<sup>2</sup> Autora convidada, Professora do Instituto Brasiliense de Direito Público, IDP, Brasil



desenvolvimento. Este não se deve apenas a avanços científicos fundamentais nas técnicas de IA, mas, principalmente, à concentração da indústria tecnológica, que se dedicou em construir os melhores algoritmos para as suas finalidades. Estes algoritmos foram conquistados a partir do aumento no poder computacional e de rede, e da capacidade de coletar e armazenar grandes quantidades de dados, possibilitando o uso de volumes cada vez mais massivos (WHITTAKER, 2021, OTERO, 2019).

Não sem razão, as recentes discussões acerca dos usos da IA têm se concentrado na necessidade de definição jurídica dos seus limites. Nesse contexto, coexistem, de um lado, um compreensível receio de que a intervenção estatal desidrate o potencial de expansão da inovação tecnológica, e, de outro, a insegurança causada pela ausência (ou insuficiência) de regulação digital, o que, além de inibir investimentos na área, ameaça o compromisso do Estado com a promoção e proteção de direitos fundamentais.

Alguns exemplos ilustram a dimensão do problema enfrentado.

Recentemente, denunciou-se à Organização das Nações Unidas (ONU) o suposto uso de IA por Israel, com a finalidade de definir alvos de ataques na Faixa de Gaza (ONU NEWS, 2024). O episódio é uma amostra dos impactos possíveis da “militarização da IA” para a automação da guerra, cujo debate ganha importância em razão da criação e desenvolvimento de sistemas próprios destinados à obtenção de vantagem estratégica bélica em diversos países, implicando, diretamente, sistemas de armas letais autônomos (em inglês, LAWs – Lethal Autonomous Weapons) e indiretamente, sistemas de apoio à decisão baseados em IA (Lavender/Where is Daddy? - Israel). Para tanto, são mobilizados elevados investimentos em ciência e tecnologia voltados estritamente para tal fim (THE ECONOMIST, 2019).

Esses usos podem representar riscos para a estabilidade internacional, nomeadamente através de uma escalada à velocidade das máquinas.

Em debates multilaterais, o Brasil tem apontado preocupação sobre a possibilidade de integração da inteligência artificial nos sistemas de comando, controle e comunicações nucleares (NC3), assim como a notícia de que os sistemas informáticos assistidos por IA são capazes de desenvolver, em horas, novos compostos químicos venenosos e de conceber novos agentes patogênicos (BRASIL, 2023).

Os impactos do uso da IA no mercado de trabalho são igualmente significativos e os riscos diversos. Enquanto o Objetivo de Desenvolvimento Sustentável N°8 das Nações Unidas propõe a promoção do emprego pleno e produtivo para todos, a adoção da IA para a automação do trabalho, com o incremento da produção e do lucro, traz consigo os perigos da discriminação, por exemplo na seleção automatizada de currículos.

Ademais, a tecnologia redesenha a divisão tradicional do trabalho, quando possibilita a realização de processos de produção de bens e serviços em locais externos a qualquer ambiente produtivo tradicional. Muitas das dinâmicas de socialização, organização, coordenação da produção e do trabalho passam a ocorrer em espaços virtuais. Utiliza-se da extração de enormes níveis de dados, que acaba gerando o tratamento indevido de dados pessoais e sensíveis de trabalhadores, num sistema de supervigilância. Também leva à decomposição do processo de produção de bens ou serviços em microtarefas, que passam



a ser coordenadas por meio de gestão algorítmica (LAUANDE; 2022).

Recente relatório do Fundo Monetário Internacional (FMI) sugere que quase 40% dos empregos em todo o mundo serão afetados por tecnologias de IA generativa, trazendo grandes desafios ao substituir certas profissões, reduzir a demanda por mão-de-obra e impactar os salários (CAZZANIGA, 2024.), alcançando, por exemplo, a indústria criativa.

À vista de tais questões, a perquirição inafastável é sobre o papel do Direito nesta conjuntura.

## 2 INTELIGÊNCIA ARTIFICIAL – DEFINIÇÃO E CONCEITOS CORRELATOS

A Organização para a Cooperação e Desenvolvimento Econômico (OCDE) aprovou uma definição de "inteligência artificial", que serviu de base para a nova legislação da União Europeia, e segue sendo a principal referência nas discussões internacionais:

“Um sistema de IA é um sistema baseado em máquina que, para objetivos explícitos ou implícitos definidos pelo ser humano, infere, a partir dos dados de entrada que recebe, como gerar resultados: previsões, conteúdo, recomendações ou decisões que podem influenciar ambientes físicos reais ou virtuais. Diferentes sistemas de IA tem diferentes níveis de autonomia e adaptabilidade após a implantação.” (livre tradução)

Os diversos sistemas de IA envolvem múltiplos patamares de automação e complexidade, e suas aplicações também variam em termos de escala e impacto potencial sobre as pessoas. Nesse sentido, destacam-se os sistemas de *Machine Learning* (ML) e de Inteligência Artificial Generativa, os quais guardam entre si semelhanças e algumas diferenças.

*Machine Learning* ou ML consiste na utilização de algoritmos para analisar dados, aprender com eles, fazer previsões e elaborar decisões informadas. Os modelos de ML aprendem examinando dados e usando técnicas estatísticas para identificar padrões, extrair insights e encontrar relações entre entradas e saídas. Com o tempo, tal aprendizado permite mudanças de comportamento para a melhoria do desempenho de tarefas. Assim, a partir dele é possível produzir resultados automatizados próximos daqueles a que chegaria uma pessoa em situação semelhante (SURDEN, 2014).

Por sua vez, a IA generativa compreende o aprendizado de máquina para gerar novos conteúdos. Ou seja, trata-se de um modelo de produção de ativos completamente novos (SURDEN, 2014).

Em síntese, o ML está focado na análise de dados para encontrar padrões e fazer previsões. A IA generativa, por outro lado, está focada na criação de novos dados que se assemelham a dados de treinamento. Como se pode deduzir, cada um à sua maneira, são sistemas que oferecem riscos associados a seus usos, conforme se destaca adiante.

## 3 DESREGULAÇÃO E RISCOS DO USO DA IA



Com o desenvolvimento da IA, capacidades antes dificilmente imagináveis emergem em um ritmo rápido e sem precedentes, oferecendo um potencial extraordinário. Dessa maneira, ao lado dos avanços, o desenvolvimento da IA, se desacompanhado de regulação, pode trazer sérios riscos aos direitos fundamentais. Alguns desses riscos são abordados a seguir.

### **3.1 A lógica da correlação**

Dentre os grandes impactos trazidos pela IA, preocupa a substituição da lógica da causalidade pela correlação. A partir da mineração de dados, que consiste em análises automatizadas de big data, os sistemas de ML induzem suas próprias regras a partir da análise de dados, correlações, classificações, criação de perfis, permitindo, a partir da aplicação de modelos estatísticos, o chamado sistema de tomada de decisão automatizada. Esses sistemas acabam alterando os processos decisórios que antes eram de competência exclusiva de seres humanos, produzindo uma nova lógica que não a da causalidade (POLLICINO; GREGORIO, 2021, p. 25-128).

Assim, na mineração de dados, os algoritmos permitem o discernimento de padrões, os quais refletem correlações, porém não são a prova inequívoca de relações causais. Um determinado padrão pode ser a indicação de um processo causal subjacente relevante, mas, também, pode ser o resultado de processos ou ruídos desinteressantes (BAROCAS; SELBST, 2016).

Confiar em correlações sem investigar efeitos causais traz o risco de que as correlações sejam ‘forçadas’ nos dados, e, mesmo que uma forte correlação estatística seja encontrada, isso só diz algo sobre um determinado (sub)grupo de pessoas, e não necessariamente sobre determinado indivíduo pertencente a esse grupo (ELBERS, NAVAS 2020).

Essa mudança tem sentido ainda maior quando as decisões têm efeitos legais ou outros impactos significativos na vida das pessoas, uma vez que elas podem minar as garantias procedimentais e substantivas relacionadas ao Estado Democrático de Direito (MICKLITZ et al., 2021).

A contenção desse risco passa imperativamente pela observância do princípio da supervisão humana, a ser concretizado com a revisão humana das decisões automatizadas que afetem direitos.

### **3.2 A opacidade**

Outra preocupação que se destaca em relação às técnicas avançadas é a opacidade de muitos sistemas de IA. A noção de IA “caixa-preta” refere-se a um cenário em que são vistos apenas os dados de entrada e os dados de saída, sem que se possa entender exatamente o que acontece no meio (PASQUALE, 2015).

Para os afetados por uma decisão algorítmica, ou por um documento que foi gerado de forma sintética por IA generativa, é importante compreender por que o sistema chegou a tal decisão ou redação. Também efeitos discriminatórios dos algoritmos são frequentemente baseados em mecanismos menos visíveis do que a simples utilização de um critério proibido. Os efeitos nocivos da discriminação não



intencional de um sistema de IA geralmente não são rastreáveis por meio de uma avaliação humana (BAROCAS; SELBST, 2016).

A opacidade e a natureza de “caixa preta” dos sistemas, somados à sua complexidade decorrente da crescente interconectividade dos sistemas computacionais, tornam muito mais complexa a definição de responsabilidades e a responsabilização pelos danos causados pelo uso de IA, colocando em risco o princípio da reparação integral.

### 3.3 A perfilização

Pode-se afirmar que o ML, a partir da mineração de dados e da aplicação de modelos estatísticos que usam uma lógica correlacional, não faz avaliações sobre indivíduos. Todavia, classifica-os como membros de um determinado grupo, distinguindo-os em relação a outros grupos.

Os indivíduos passam a ser avaliados pelas características do grupo em que alocados e não por uma particularidade própria. Essas características são definidas a partir de bases de dados históricas, em que o sistema busca correlações e padrões. Assim, embora um perfil se refira apenas a correlações entre dados que traduzem eventos passados, ele será usado para “determinar” o futuro – tanto predizendo eventos vindouros como se fossem inteiramente determinados, quanto decidindo realmente sobre o futuro das pessoas, como uma profecia autorrealizável (HILDEBRANDT, 2010).

Tecnologias de criação de perfil podem forjar um sistema de Justiça que responsabiliza os cidadãos não pelo que fazem, mas predominantemente por exibirem características que correspondem aos perfis daqueles que assim já o fizeram no passado. No âmbito criminal, isso corresponde à criminalização de “comportamentos indesejados”. Assim, uma vez que os dados corresponderam a um perfil criminal, o comportamento futuro do indivíduo será sobredeterminado, em uma espécie de punição “pré-crime” (HILDEBRANDT, 2010), inclusive mediante a influência no convencimento do julgador.

### 3.4 Os vieses discriminatórios

A discriminação pode ser um artefato do próprio processo tecnológico da Inteligência Artificial (BAROCAS, SELBST, 2016). Os vieses replicam e amplificam preconceitos existentes na sociedade, em especial as desigualdades estruturais: racial, de gênero e de classe, entre outras (POWLES; NISSENBAUM, 2018).

Friedman e Nissenbaum (1996), em pesquisa feita a partir de casos reais, desenvolveram uma classificação dos vieses existentes em sistemas computacionais, agrupando-os em três categorias: preexistentes, técnicos e emergentes.

O viés preexistente tem suas raízes em instituições, práticas e atitudes, e atua de forma independente, geralmente antes da criação do sistema, sendo por ele incorporado. Tal viés pode entrar em um sistema por meio de esforços explícitos e conscientes de indivíduos ou instituições, ou implícita e



inconscientemente, mesmo a despeito das melhores intenções.

Por sua vez, o viés técnico surge de restrições técnicas, a exemplo daquelas relacionadas a algoritmos descontextualizados e a erros decorrentes de imperfeições na geração de números pseudoaleatórios ou no uso indevido de números pseudoaleatórios.

Finalmente, o viés emergente surge em um contexto de uso que envolve usuários reais. Esse viés resulta, geralmente, de uma mudança de conhecimento ou de valores culturais, em especial nas aplicações que possuem interface com os usuários em determinados contextos (FRIEDMAN E NISSENBAUM, 1996).

De forma geral, as coletividades humanas estão imersas em sociedades enviesadas. Em um contexto em que grupos vulneráveis são especialmente afetados, treinar algoritmos inteligentes a partir de dados representativos do mundo e tecnicamente corretos não garante, por si só, o atingimento de resultados científicos e corretos. Isso porque, na hipótese de esses dados refletirem vieses, ainda que inconscientes, a atividade de mineração de dados empregada no treinamento dos sistemas de decisões algorítmicas tomará tais vieses como fatores relevantes para a decisão e passará a replicá-los em casos futuros. Assim, ao invés de aprimorar a qualidade da decisão, o recurso à IA pode consolidar um status quo que se deseja mudar, normalizando vieses e dando a essas decisões enviesadas um caráter de cientificidade que elas não merecem possuir.

#### 4 REGULAÇÃO DA IA NO MUNDO E NO BRASIL

Conforme ensina Wolfgang Hoffmann-Riem, todas as normas vigentes no âmbito nacional e internacional se aplicam ao uso de tecnologias digitais, ainda que não haja expressa referência a tal âmbito de aplicação daquelas normas. Todavia, ante a insuficiência em face das complexidades que decorrem da expansão da IA, é preciso a criação de estratégias de promoção da “boa governança digital”, inclusive por meio da regulação jurídica dos usos associados à tecnologia (HOFFMANN-RIEM, 2019, p. 06).

Nesse contexto, o Estado deve atuar como um agente promotor de condições da inovação e, ao mesmo tempo, garantidor de direitos fundamentais.

Hoje existe praticamente um consenso internacional de que a regulação de IA e de outras tecnologias digitais não pode mais ser feita somente por meio de técnicas de *soft law* e de autorregulação (como princípios éticos), com dispositivos genéricos, desprovidos de normatividade e coercibilidade, em alguns casos sem o *enforcement* necessário. Entre especialistas, acadêmicos, desenvolvedores da tecnologia, além de blocos regionais como a União Europeia, não há mais um debate se é necessário regular a IA, mas sim como regulá-la.

A regulação estatal é uma das formas de regulação do uso de tecnologias que envolvem o emprego de IA, que se soma a outras, como a autorregulação e a regulação híbrida. Trata-se de projetar “um sistema legal elaborado e aplicado por meio de instituições sociais ou governamentais para regular o comportamento dos agentes na infosphere” (FLORIDI, 2018, p. 03). Em relação à tarefa do Estado, deve-se atentar para a



necessidade de restringir os riscos associados ao emprego de IA, evitando o efeito contraproducente consistente na inibição da inovação (HOFFMANN-RIEM, 2019, p. 26).

Em 2019, com o propósito de promover a contenção dos riscos que envolvem a utilização de IA, a Comissão Europeia publicou Orientações Éticas para uma IA de Confiança, segundo a qual o adequado uso da IA requer, além da observância de princípios éticos (IA ética) e de sólidas regras técnicas (IA técnica), o cumprimento de toda a legislação e regulamentação aplicáveis (IA legal) (UNIÃO EUROPEIA, 2019).

No mesmo ano de 2019, a OCDE instituiu a Recomendação sobre Inteligência Artificial (OECD/LEGAL/0449), definindo um conjunto de princípios relacionados ao uso de IA, de natureza não vinculativa, sublinhando a necessidade de crescimento inclusivo; desenvolvimento sustentável e bem-estar; valores e justiça centrados no ser humano; transparência e explicabilidade; robustez, segurança, proteção e responsabilização. Também estão incluídas diretrizes para formuladores de políticas, inclusive na área de cooperação internacional, capacitação e pesquisa e investimentos. Já em 2020, no âmbito da OCDE, ocorreu o lançamento da Global Partnership on Artificial Intelligence (GPAI).

Diante do avanço da IA Generativa e da urgência de parâmetros mínimos globais voltados para a prevenção de riscos e para a precaução, a reunião do G7 de 2023, constituiu um grupo de trabalho com a adesão de outros países, reunindo especialistas e inovadores, no chamado Processo de Hiroshima. O grupo tem o objetivo de promover interoperabilidade entre estruturas de governança e sobre ferramentas para IA confiáveis, como estruturas de gerenciamento de risco, auditoria e possíveis esquemas de certificação. Especificamente, visa à criação de um código de conduta para adesão dos entes privados envolvidos com o desenvolvimento da Inteligência Artificial.

Já em novembro de 2023, o Reino Unido sediou o AI Safety Summit, uma conferência internacional para discutir padrões de segurança e regulação de IA. Vinte e oito países, incluindo os Estados Unidos, a China e o Brasil, emitiram um acordo conhecido como Declaração de Bletchley, apontando a cooperação internacional para gerenciar os desafios e riscos da inteligência artificial, com ênfase na regulamentação da chamada "IA de Fronteira".

Os Estados Unidos lançaram em 2023 uma Declaração de Direitos de IA, uma diretiva para Estrutura de Gerenciamento de Risco de IA e duas ordens executivas, a última bastante ampla, com forte ênfase nas questões de segurança, discriminação algorítmica, mais transparência e novos padrões.

Mais recentemente, em março de 2024, a União Europeia aprovou a Lei de Inteligência Artificial da União Europeia (AI Act), a primeira lei abrangente de IA do mundo. Entre os seus objetivos, o de “garantir que os sistemas de IA disponibilizados no mercado da UE sejam seguros e respeitem a legislação em vigor”. Para isso, a Lei realiza um escalonamento de riscos, definindo maior controle para os usos de IA associados aos maiores riscos (AI Act, 2024).

A UE também está trabalhando em outro projeto de lei, chamado Diretiva de Responsabilidade da IA, para atualizar as questões de responsabilidade civil, que deve ser aprovada este



ano.

Na ONU, além da Resolução de 2023 sobre a governança da Inteligência Artificial com foco nos Direitos Humanos, de ordem principiológica, foi criado um Comitê Consultivo de Alto Nível para Inteligência Artificial, ligado à Secretaria Geral, que apresentará um documento final para ser discutido e aprovado pelos países membros no âmbito da próxima Assembleia Geral, no processo de um “Pacto Digital Global”.

No Brasil, a Estratégia Brasileira para a Transformação Digital (E-Digital), aprovada em 2018 pelo Decreto nº 9.319/2018 e pela Portaria MCTIC nº 1.556/2018, apontou para a importância de tratar o tema da IA de maneira prioritária, em razão de seu amplo impacto transversal. Em 2021, foi definida a Estratégia Brasileira de Inteligência Artificial (EBIA), consistente em um conjunto de orientações para a utilização da IA, com vistas a potencializar o seu uso em proveito das necessidades nacionais, por meio, entre outros, do aumento de competitividade, de produtividade e da melhoria na prestação de serviços públicos. A EBIA está atualmente em processo de revisão. Tais iniciativas se somam a outras do Poder Executivo, com vistas à criação de um ambiente de uso seguro da IA.

No plano normativo, possuem uma relação próxima à matéria a Lei n. 12.965/2014, que institui o Marco Civil da Internet, e a Lei n. 13.709/2018, que institui a Lei Geral de Proteção de Dados (LGPD), conferindo abordagem específica ao tratamento massivo de dados (big data), frequentemente realizado por sistemas de IA. Ambas as leis tratam de regulação e proteção de direitos face a tecnologias emergentes. Merece destaque, ainda, a recente Resolução n. 23.732/2024 do TSE, que regulamenta a propaganda eleitoral no Brasil, inovando, por exemplo, ao proibir, em seu art. 9º-C, § 1º, seja para prejudicar ou para favorecer qualquer candidatura, o uso de conteúdo sintético, gerado ou manipulado digitalmente, mesmo com autorização, “para criar, substituir ou alterar imagem ou voz de pessoa viva, falecida ou fictícia (deep fake)”. Isso constitui abuso do poder político e uso indevido dos meios de comunicação social, acarretando a cassação do registro da candidatura ou do mandato (§ 2º).

Com o fim de disciplinar especificamente a IA, tramitam no Congresso Nacional, atualmente, diversos Projetos de Lei sobre o tema. Entre eles, destaca-se o PLS 2338/23, concebido como o mais abrangente e avançado, que consiste no resultado dos trabalhos da Comissão de Juristas responsável por subsidiar elaboração de substitutivo sobre Inteligência Artificial no Brasil (CJSUBIA).

Conforme sua justificativa, a proposição legislativa tem o objetivo de “proteger o elo mais vulnerável (a pessoa natural)” nas relações que envolvem o uso de IA, bem como “criar condições de segurança jurídica para a inovação e o desenvolvimento tecnológico, a partir de ferramentas de governança, fiscalização e supervisão” (SENADO FEDERAL, 2023).

Após a apresentação do Projeto de Lei 2338/23, o Senado Federal criou uma Comissão Especial para avaliar o texto. A Comissão realizou dezenas de audiências públicas e recebeu contribuições. O Relator



da Comissão, Senador Eduardo Gomes, apresentou um novo relatório na última reunião de abril do corrente ano, com um novo texto.

O novo texto apresentado mantém a estrutura do texto original, embora traga algumas mudanças significativas em relação ao projeto primitivo.

O PLS 2338/23 prevê, inicialmente, o escopo de aplicação da lei: uma norma geral que trata da “concepção, o desenvolvimento, implementação, utilização, adoção e governança responsável de sistemas de IA com o objetivo de proteger os direitos fundamentais, estimular a inovação responsável e garantir a implementação de sistemas seguros e confiáveis, em benefício da pessoa humana, do regime democrático e do desenvolvimento econômico, científico e tecnológico.”

E segue trazendo as exceções ao escopo de aplicação, que fundamentalmente são os sistemas de IA para finalidades econômicas, com foco naquilo que traz alto risco. A lei não se aplica aos sistemas empregados exclusivamente para fins de defesa nacional.

O conceito central do PLS é o de “sistema de IA”, que foi atualizado a partir do novo texto da OCDE: “sistema baseado em máquina que, com graus diferentes de autonomia e para objetivos explícitos ou implícitos, infere, a partir de um conjunto de dados ou informações que recebe, como gerar resultados, em especial, previsão, recomendação ou decisão que possa influenciar o ambiente virtual ou real;” (Art. 4º, inc. I).

O substitutivo ao PLS dispõe sobre a definição de “sistema de IA de propósito geral” e “modelos fundacionais”. Sendo sistemas de escopos mais amplos e abrangentes, eles terão medidas próprias de governança para seu desenvolvimento, para além daquelas que a proposta prevê, relacionadas aos riscos de uso das aplicações, seguindo os passos da legislação europeia.

O novo texto do PLS inclui a definição de inteligência artificial generativa e conteúdo sintético derivado, embora seja bastante tímido em relação aos requisitos próprios para esse tipo de sistema e do conteúdo gerado por ele.

Não obstante propostas de textos referentes a esse tipo de sistema tenham sido apresentadas ao relator, o texto do substitutivo ao PLS não incluiu obrigações diferenciadas, por exemplo em relação ao treinamento desse tipo de ferramenta, ou rotulagem daquilo que seria o conteúdo sintético derivado, como os Estados Unidos regulam em sua ordem executiva ou as discussões do processo de Hiroshima, que resultou na proposta de um Código de Conduta para as companhias que desenvolvem essa tecnologia.

O PLS aborda, mas sem maiores desdobramentos, o importante conceito de Integridade Informacional. O novo texto definiu integridade informacional como característica de informações que, em seu conjunto, sejam predominantemente precisas, consistentes e confiáveis. No cenário internacional, o Brasil tem sido um dos países a liderar essa discussão: a presidência brasileira do G20 tem enfatizado o destaque ao tema, que também será tratado no processo do Pacto Global Digital da ONU.

As chamadas Fake News, com impacto, por exemplo, na integridade eleitoral e no processo democrático, na cobertura vacinal e nas questões climáticas, têm preocupado governos e organismos



internacionais. O uso da Inteligência Artificial generativa, que a cada dia apresenta resultados mais surpreendentes, traz impacto exponencial a essas temáticas.

Pesquisadores afirmam que atualmente as plataformas ainda não conseguem identificar conteúdos sintéticos, que continuam circulando sem a identificação (DESINFORMANTE, 2023). Isso é relevante em um cenário em que, de acordo com relatório divulgado pela Europol, especialistas estimam que, em 2026, 90% do conteúdo online será gerado por IA (EUROPOL, 2022).

Outro conceito importante no PLS é o de discriminação indireta. Previsto em acordos internacionais internalizados pelo país, como na Convenção Interamericana contra o Racismo, é um conceito central para o Direito Antidiscriminatório. No PL original, ele aparecia mais detalhadamente no art. 12. No substitutivo, ele consta na previsão do direito à não-discriminação e à correção de vieses discriminatórios diretos, indiretos, ilegais ou abusivos (inc V. art. 8º).

O conceito jurídico tradicional de discriminação parte de um critério ou prática que acarreta uma desvantagem ou um tratamento adverso, de um membro de um grupo ou de um grupo com características protegidas como raça, gênero, sexualidade, religião (elemento objetivo), para negar direitos, somando-se ao intuito discriminatório (elemento subjetivo) (MOREIRA, 2020, DIMOULIS, 2021).

Essa discriminação choca-se contra o preceito de igualdade formal, e há uma intencionalidade no tratamento desvantajoso ou na utilização de critério proibido por lei (FREDMAN, 2016; BAROCAS, 2019).

Mas os algoritmos, a partir de características do próprio artefato tecnológico, apresentam vieses que podem levar a resultados discriminatórios, mesmo que de modo não intencional. Esses vieses podem prejudicar classes, mas, diferente de outros mecanismos de discriminação tradicional, são menos óbvios e, portanto, de difícil identificação. Assim gera-se o que chamamos de discriminação algorítmica, mencionada pela doutrina americana como *disparate impact* (BAROCAS, SELBEST, 2016; CITRON, PASQUALE, 2014).

O *disparate impact*, ou impacto desigual, ocorre quando uma disposição, critério ou prática aparentemente neutros resulta uma desvantagem significativa para um grupo protegido, apesar não usar explicitamente essa característica e mesmo quando não há necessariamente essa intencionalidade (WACHTER; MITTELSTADT, 2019).

O conceito tradicional de discriminação direta não é suficiente para dar conta dos efeitos do *disparate impact*, que é abarcado pelo conceito jurídico de discriminação indireta, central no que denominamos de discriminação algorítmica. Uma vez que o paradigma de causalidade se torna desatualizado na era algorítmica, não há que se falar de intencionalidade, e sim se o resultado de uma decisão desfavorece uma característica protegida pela nossa legislação (WACHTER; MITTELSTADT, 2019, BAROCAS, SELBEST, 2016).

Um dos maiores desafios do legislador, quando precisa codificar de forma clara garantias para que os impactos sociotécnicos de uma tecnologia não afetem os direitos fundamentais já estabelecidos, é como



fazer que princípios gerais se convertam em mecanismos capazes de traduzir, e de “operacionalizar”, normas legais como parâmetros obrigatórios para os processos de desenvolvimento e usos dessa tecnologia emergente (CELESTE, 2023)

Fundamental perquirir como adotar regras adequadas para a aplicação em um contexto no qual necessitamos de um sistema de padrões muito granular e muito dinâmico, direcionado a setores muito diferentes, já que se trata de uma legislação transversal. Outro desafio é encontrar uma ferramenta de governança capaz de cumprir o que é normativo e que ao mesmo tempo reconheça o papel dos instrumentos de *softlaw*, **que são necessários, mas de aplicação complexa** (CELESTE, 2023).

Essas regras precisam resultar na manutenção das garantias e dos processos por meio dos quais são criados, exercidos e aplicados os direitos substantivos (CELESTE, 2023).

Por isso, é importante que a proposta legislativa mantenha a abordagem mista proposta pela Comissão de Juristas instituída pelo Senado (CJSUBIA), e tenha, para além da legislação baseada em riscos (*risk based approach*), a previsão de direitos substantivos e garantias procedimentais para sua efetividade (*rights based approach*).

Essa abordagem mista, pragmática (incluindo ferramentas e parâmetros-chave) combinada com a abordagem baseada em princípios e previsão de direitos, traz um equilíbrio que pode tornar mais eficaz os objetivos da legislação (MANTELERO, 2023).

Nesse passo, é de alta relevância o capítulo do PLS sobre um conjunto de direitos titularizados pelas pessoas “afetadas por sistemas de IA”, que deverão ser informados de forma clara e acessível, como mencionamos doravante.

Do direito fundamental ao livre desenvolvimento da personalidade deriva a liberdade cognitiva, que é assegurada pelos direitos associados à informação e compreensão das decisões tomadas por sistemas de inteligência artificial (como o direito à informação prévia quanto às interações com sistemas de inteligência artificial e o direito à participação humana em decisões de sistemas de inteligência artificial), tudo conforme os incisos I e III do Artigo 8º.

Do direito fundamental à autodeterminação informativa decorre o direito à privacidade e à proteção de dados pessoais (inc. II, art. 8º).

Do devido processo legal e da autonomia humana se extraem os direitos previstos no art. 9º, como o direito à explicação sobre decisão, recomendação ou previsão tomadas por sistemas de inteligência artificial (inc.I); o direito de contestar decisões ou previsões de sistemas de inteligência artificial que produzam efeitos jurídicos ou que impactem de maneira significativa os interesses do afetado (inc. II); o direito à supervisão humana das decisões, levando-se em conta o contexto e o estado da arte do desenvolvimento tecnológico e o respectivo risco associado (inc III).

Do direito à igualdade se extrai o direito à não-discriminação algorítmica e à correção de vieses discriminatórios diretos, indiretos, ilegais ou abusivos (inc. IV, art. 8º).



O PL original da CJSUBIA também previa o direito à inferência razoável, que abrange a proteção contra decisões, recomendações ou previsões amparadas em inferências discriminatórias, irrazoáveis ou que atentem contra a boa-fé objetiva, o que infelizmente foi excluído do substitutivo ao PLS, mas é um importante direito na medida em que trocamos a lógica da causalidade pela correlação.

A proposta do Senado (PLS) segue dando ênfase à abordagem de riscos, baseada no seu escalonamento, tal qual a nova Lei de Inteligência Artificial editada pelo Parlamento Europeu (IA Act). A gestão baseada em riscos compreende o ajuste do peso da intervenção estatal de acordo com a extensão dos riscos associados à tecnologia. Deve ser feita a partir de uma autoavaliação e classificação de seu grau de risco previamente à utilização do sistema ou sua colocação no mercado.

Na abordagem de riscos prevista pelo PLS o elemento central é a avaliação de impacto: o ‘se’ vem antes do ‘como’. A automação sem impacto não é uma questão legal. Em caso de impacto, existem critérios da aceitabilidade a partir de uma abordagem regulatória na qual os riscos médios e baixos são considerados aceitáveis. A crítica a essa abordagem, similar à legislação europeia, é que falta consideração do impacto cumulativo, caso haja o uso de vários sistemas de baixo risco, que somados, podem trazer riscos maiores (MANTELERO, 2023).

O projeto de lei do Senado prevê critérios para que o SAI (Sistema de Regulação e Governança de Inteligência artificial) defina o que é alto risco (Art. 15). A partir desses critérios, seria possível a reclassificação dos usos conforme o grau de risco, levando em conta o estabelecido pela lei, a exemplo da possibilidade de o sistema de IA impactar negativamente o exercício de direitos e liberdades, ou a utilização de um serviço.

O alto risco deve ser determinado de acordo com a finalidade ou contexto de uso, incluindo: gestão e funcionamento de infraestruturas críticas; educação; relações de trabalho; avaliação de acesso a serviços privados e públicos que sejam considerados essenciais; crédito; administração da justiça; veículos autônomos; aplicações na área da saúde para auxiliar diagnósticos e procedimentos médicos, quando houver risco à integridade física das pessoas; segurança pública; investigação por autoridades; imigração e controle de fronteiras; ou sistemas que permitam disseminação de desinformação e discurso de ódio.

O Projeto do Senado também cuida sobre parâmetros para uma Avaliação de Impacto do Algoritmo em todos os sistemas de alto risco (Art. 22), incluindo requisitos de transparência, qualidade dos dados, foco em avaliação de vieses e rastreabilidade, confiabilidade, explicabilidade e ênfase na supervisão humana em caso de impacto nos direitos. Destaca-se também a natureza contextual da avaliação, que é um avanço importante para a garantia de direitos.

Como sistema de salvaguardas “ex post”, além dos direitos de transparência previstos no capítulo II, está previsto no PLS um sistema de supervisão e fiscalização do cumprimento da lei por meio do SAI (Sistema de Regulação e Governança de Inteligência artificial).

No mesmo terreno das salvaguardas, muito importante a previsão de um sistema próprio de responsabilidade, o qual advém da opacidade e natureza não verificável dos sistemas (conceito de “caixa



preta” de Pasquale), além da sua complexidade.

Por isso, merece redobrada atenção a fixação de normas de responsabilização civil por danos causados por sistemas de IA. Sublinha-se a previsão do dever de reparação integral pelo dano causado, independente do grau de autonomia do sistema (Art. 27, caput); a fixação da responsabilidade objetiva, no caso de sistemas de alto risco e risco excessivo (Art. 27, § 1º); e da responsabilidade por culpa presumida nas demais hipóteses (Art. 21, § 2º); bem como das causas excludentes de responsabilidade, a exemplo da culpa exclusiva da vítima ou de terceiros (Art. 28).

Estas regras de responsabilidade são complementadas pelo enfoque a Boas Práticas e Governança, incluindo mecanismos de autorregulação como Códigos de Conduta e criação de entidades para essa finalidade.

Merece destaque a seção do projeto (PLS) referente aos direitos autorais e direitos de personalidade conexos, que é um dos tópicos que mais se destaca no debate internacional acerca do tema, já com grandes disputas judiciais em curso, em especial quanto ao uso de material protegido por propriedade intelectual para o treinamento de sistemas de IA Generativa. A proposta prevê obrigações de transparência, remuneração e fiscalização sancionatória quanto aos dados e obras utilizadas, assim como de remuneração compensatória aos autores de obras intelectuais protegidas. Obrigações que, inclusive, devem retroagir ao período em que os conteúdos de direitos de autor e conexos foram utilizados.

O texto do PLS determina que, quando os sistemas de inteligência artificial utilizarem conteúdo protegido por direitos de autor e conexos em seu desenvolvimento, o desenvolvedor deverá informar quais obras e demais conteúdos protegidos por direitos de autor e conexos foram utilizados nos processos de treinamento dos sistemas de inteligência artificial. Esse preceito é essencial porque sem mecanismos de transparência sobre as obras utilizadas por sistemas de inteligência artificial não há qualquer possibilidade de tutela dos direitos autorais, quando tratamos de IA Generativa.

Há previsão de parâmetros para a exceção à incidência de direitos autorais, exclusivamente em benefício de instituições de pesquisa ou genuinamente sem fins lucrativos, uma vez que na legislação de direitos autorais tais atividades já gozam de certas prerrogativas, em diversas modalidades de utilização de obras protegidas.

Obviamente é necessário evitar que empresas lucrativas aproveitem indevidamente desse mecanismo e seja possível a exploração econômica a partir da mineração de dados, sem autorização nem compensação, de obras e conteúdos alheios - a exemplo do que ocorreu com a Open IA, que foi inicialmente constituída como instituição sem fins lucrativos (MINC, 2024).

O titular de direitos de autor e conexos poderá proibir a utilização dos conteúdos de sua titularidade no desenvolvimento de sistemas de inteligência artificial, consoante Art. 55 do PLS.

Lembramos que o conteúdo protegido por direitos autorais (textos literários, obras visuais, letras, músicas etc.) é o mais valioso insumo para o desenvolvimento e treinamento da IA Generativa, pois trazem



dados de melhor qualidade. Sem a regulamentação do uso de obras protegidas pelos direitos autorais para o desenvolvimento e treinamento de sistemas de IA Generativa, teremos um quadro altamente predatório entre os "resultados" gerados por IA e as obras feitas por seres humanos. Sendo que grande parte desses "resultados" são produtos da utilização de obras protegidas, utilizadas sem qualquer tipo de autorização ou remuneração em prol de seus titulares.

A compensação por eventual uso de obras, fonogramas, execuções, interpretações ou emissões no treinamento de sistemas de IA, também possui efeitos de fortalecer o respeito aos direitos autorais em sentido amplo, reafirmando a centralidade da pessoa humana, reconhecendo a capacidade criativa e promovendo a diversidade cultural (MINC, 2024).

Outro aspecto relevante é a remuneração do jornalismo, pelas mesmas razões compensatórias do uso de conteúdo artístico protegido, mais ainda como elemento central de proteção da integridade da informação. A popularização dos sistemas de IA generativa, se não acompanhada de forte compensação econômica para produtores de conteúdo jornalístico e outros tipos de informação, pode gerar um enorme prejuízo para aqueles que produzem informação primária de qualidade, com prejuízo enorme para garantias e liberdades.

Finalmente, se reforça que a governança da proteção de direitos autorais deve ter sua competência legal preservada no PLS sobre Inteligência Artificial, em se tratando de um sistema já em funcionamento com legitimidade, conhecimento específico acumulado, interlocução e participação do setor.

## 5 CONCLUSÃO

Ainda não existe um consenso sobre como avaliar ou abordar os riscos associados à IA. No entanto, tal como prevê o princípio da precaução no setor ambiental, a incerteza científica sobre os riscos não deve conduzir à paralisia da governança.

Nesse contexto, nota-se que, ainda que prevaleça o entendimento no sentido da necessidade de regulação da IA, persistem dúvidas sobre o que, de fato, deve ser feito nessa seara, ou seja, qual deve ser a exata dimensão regulatória, e em que medida tal tarefa deve ser exercida pelo Estado.

Defendemos a imprescindibilidade da regulação da IA por meio da intervenção normativa do Estado. Nesse sentido, entendemos que tanto a autorregulação como a fixação de princípios em documentos de soft law se mostram insuficientes para o disciplinamento eficaz dos elevados riscos associados à IA.

No Brasil, o caminho da regulação estatal da IA encontra-se aberto, destacando-se, nesse particular, as inovações contidas no PLS 2338/23, que tem como pontos fortes: I) a garantia do exercício de direitos fundamentais; II) a gestão do risco por escalonamento e; III) a responsabilização integral pelos danos causados pela IA.

## REFERÊNCIAS



BAROCAS, Solon; SELBST, Andrew D. **Big Data's Disparate Impact**. California Law Review. v. 104, n. 3, p. 671-732, 2016.

BRASIL. **Anteprojeto de Código Civil**. 2024.

BRASIL. Statement by Ambassador Sérgio França Danese, Permanent Representative of Brazil to the United Nations, at the UNSC Open Debate on Artificial Intelligence. July 18th, 2023

BRASIL. SENADO FEDERAL. PLS 2338/2023 - Dispõe sobre o uso da Inteligência Artificial. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 28 abr. 2024.

BRASIL. SENADO FEDERAL. **Relatório da Subcomissão de Direito Digital**. 2022. Disponível em: <https://legis.senado.leg.br/comissoes/mnas?codcol=2504&tp=4>. Acesso em: 27 abr. 2024.

CAZZANIGA AND OTHERS. 2024. “Gen-AI: Artificial Intelligence and the Future of Work.” IMF Staff Discussion Note SDN2024/001, International Monetary Fund, Washington, DC.

CUEVA, Ricardo Villas Bôas. A regulação da inteligência artificial no Projeto de Lei nº 21/2020. In. TEPEDINO, Gustavo et. al. **A evolução do direito no século XXI: seus princípios e valores (ESG, liberdade, regulação, igualdade e segurança jurídica): homenagem ao professor Arnoldo Wald**. São Paulo: IASP, v. 2, p. 1179-1193, 2022.

DESINFORMANTE. **Plataformas ainda não conseguem identificar conteúdos sintéticos**. nov. 28, 2023. Disponível em: <https://desinformante.com.br/plataformas-conteudo-sintetico/>

DIMOULIS, Dimitri. **Direito de igualdade: antidiscriminação, minorias sociais, remédios constitucionais**. Editora: Almedina, 2021.

EBERS Martin; NAVARRO Susana. **Algorithms and Law**. Cambridge: Cambridge University Press, 2019.

CELESTE, Edoardo. **The Content Governance Dilemma, Information Technology and Global Governance**, 2023

EHRHARDT JUNIOR, Marcos; FRANÇA NETTO, Milton Pereira de. O regime de responsabilidade civil aplicável à inteligência artificial no direito brasileiro. **Revista Magister de Direito Civil e Processual**



Civil, v. 19, n. 111, p. 5-36, 2022.

EUROPOL (2022), **Facing reality?** Law enforcement and the challenge of deepfakes, an observatory report from the Europol Innovation Lab, Publications Office of the European Union, Luxembourg.

FREDMAN, S. **Discrimination law**. New York: Oxford University Press, 2011.

FLORIDI, Luciano. **Soft ethics, the governance of the digital and the General Data Protection Regulation**. *Philosophical Transactions of the Royal Society*. v. 376, n. 2133, p. 1 1-11, 2018.

FRIEDMAN, Batya; NISSENBAUM, Helen. **Bias in Computer Systems**. *ACM Transactions on Information Systems*. v. 14, n. 3, p. 330 –347, 1996.

GUTWIRTH, Serge; HILDEBRANDT, Mireille. **Some Caveats on Profiling**. In: GUTWIRTH, S. et al. (eds.). **Data Protection in a Profiled World**. Springer, 2010. p. 31-41.

HILDEBRANDT, M. **Proactive Forensic Profiling: Proactive Criminalization? The Boundaries of the Criminal Law**, 2010, p. 113-137. Available at: [http://works.bepress.com/mireille\\_hildebrandt/37/](http://works.bepress.com/mireille_hildebrandt/37/).

HILDEBRANDT, M. **The Issue of Proxies and Choice Architectures. Why EU Law Matters for Recommender Systems 2022** Apr 28, n. 5, p. 789076. doi: 10.3389/frai.2022.789076. PMID: 35573902; PMCID: PMC9096719.

LAUANDE Rodrigues, Priscila; MASSIMO DE, Minicis. **Technological changes, labour market transformations, industrial and labour policies**. pp. 112-125

MANTELERO, Alessandro. **Direitos Fundamentais na Era Digital**. Palestra proferida no Seminário Internacional: Democracia e regulação da IA: perspectivas e desafios, IDP, 2023.

MOREIRA, Adilson José. **Tratado de Direito Antidiscriminatório I**. São Paulo: Editora Contracorrente, 2020.

MICKLITZ, H., POLLICINO, O.; REICHMAN, A.; SIMONCINI, A.; SARTOR, G.; DE GREGORIO, G. (ed.). **Constitutional Challenges in the Algorithmic Society**. Cambridge: Cambridge University Press, 2021.



MINC (2024). Nota técnica sobre PLS 2338/2023. Secretaria de Direitos Autorais. Ministério da Cultura. Governo do Brasil.

OECD (2024). "Explanatory memorandum on the updated OECD definition of an AI system", OECD Artificial Intelligence Papers, n. 8, OECD Publishing, Paris, Disponível em: <https://doi.org/10.1787/623da898-en>.

ONU NEWS. **Especialistas da ONU mencionam uso de inteligência artificial em ataques israelenses na Faixa de Gaza.** Disponível em: <https://news.un.org/pt/story/2024/04/1830431#:~:text=Espesialistas%20da%20ONU%20lamentaram%20o,moradias%2C%20servi%C3%A7os%20vitais%20e%20infraestrutura>. Acesso em: 27.04.2024.

OTERO, B. Gonzalez. **Evaluating the EC Private Data Sharing Principles:** Setting a Mantra for Artificial Intelligence Nirvana? 2019.

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

POLLICINO, O.; GREGORIO, G. Constitutional Law in the Algorithmic Society. *In:* MICKLITZ, H. W.; POLLICINO, O.; REICHMAN, A.; SIMONCINI, A.; SARTOR, G.; GREGORIO, G. de (ed.). **Constitutional Challenges in the Algorithmic Society.** Cambridge: Cambridge University Press, 2021. doi:10.1017/9781108914857.002.

POLLICINO, Oreste; GREGORIO, Giovanni De. **Blockchain and public law:** global challenges in the era of decentralization. Cheltenham, Northampton: Edward Elgar Pub, 2021.

POWLES, Julia; NISSENBAUM, Helen. The Seductive Diversion of ‘Solving’ Bias in Artificial Intelligence Trying to “fix” A.I. distracts from the more urgent questions about the technology. Disponível em: <https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53>. Acesso em: 27 abr. 2024.

SURDEN, H. **Machine learning and law.** Washington Law Review. v. 89, 2014, p. 89-90.

THE ECONOMIST. **Artificial intelligence and war:** as computers play a bigger role in warfare, the dangers to humans rise. Disponível em: <https://www.economist.com/leaders/2019/09/05/artificial-intelligence-and->



war?utm\_medium=cpc.adword.pd&utm\_source=google&ppccampaignID=19495686130&ppcadID=&utm\_campaign=a.22brand\_pmax&utm\_content=conversion.direct-response.anonymous&gad\_source=1&gclid=EAIAIqObChMIq\_r27NDihQMVhH5vBB3NNw-hEAAYAiAAEgKrpvD\_BwE&gclsrc=aw.ds. Acesso em: 27 abr. 2024.

UNIÃO EUROPEIA. AI Act. Disponível em: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>. Acesso em: 27 abr. 2024.

UNIÃO EUROPEIA. **Orientações Éticas para uma IA de Confiança**. Disponível em: <https://op.europa.eu/pt/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1>. Acesso em: 28 abr. 2024.

WHITTAKER, M. The Steep Cost of Capture. The steep cost of capture. *interactions* 28, 6 nov./dec. 2021, v. 28, n. 6, p. 50-55. Available at: <https://doi.org/10.1145/3488666>; <https://ssrn.com/abstract=4135581>.

WOLFGANG HOFFMANN-RIEM. **Inteligência artificial como oportunidade para a regulação jurídica**. *RDU*. v. 16, n. 90, p. 11-38, 2019.

