Contribuição LAPIN - CJSUBIA

Cynthia Picolo <cynthia.picolo@lapin.org.br>

qui 30/06/2022 18:55

Para: CJSUBIA <CJSUBIA@senado.leg.br>;

Cc: Gabriela Buarque <gabriela.buarque@lapin.org.br>; José Renato <joserenato@lapin.org.br>; tayrone marquesini <tayrone.marquesini@lapin.org.br>; Victor Mulin <victor.mulin@lapin.org.br>;

1 anexo

LAPIN - Contribuição CJSUBIA - jun 2022.pdf;

Exmo. Sr. Presidente Rodrigo Pacheco,
Exmo. Sr. Presidente da Comissão, Min. Ricardo Villas Bôas Cueva,
Exmo. Sr. Relator do Projeto de Lei, Sen. Eduardo Gomes,

E estimados membros da Comissão de Juristas,

Encaminhamos, no documento anexo, a contribuição à Comissão de Juristas responsável por subsidiar a elaboração de minuta de substitutivo para instruir a apreciação dos Projetos de Lei nº 5.051/2019, 21/2020, e 872/2021 em relação ao Marco Regulatório da Inteligência Artificial.

Agradecemos a consideração e permanecemos disponíveis para diálogo nessa construção.

Atenciosamente,

Cynthia Picolo

Diretora

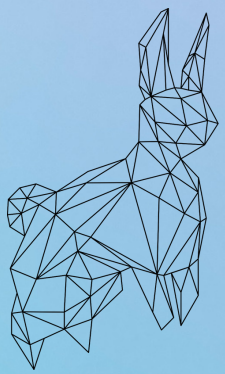
(19) 99537-0308

cynthia.picolo@lapin.org.br
lapin.org.br

2022 | JUNHO

CONTRIBUIÇÃO À COMISSÃO DE JURISTAS

responsável por subsidiar a elaboração de minuta de substitutivo para instruir a apreciação dos Projetos de Lei nº 5.051/2019, 21/2020 e 872/2021, que têm como objetivo estabelecer princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da inteligência artificial no Brasil



LAPIN

LABORATÓRIO DE POLÍTICAS PÚBLICAS E INTERNET

LABORATÓRIO DE POLÍTICAS PÚBLICAS E INTERNET

REALIZAÇÃO

Laboratório de Políticas Públicas e Internet - LAPIN

AUTORIA

Cynthia Picolo

Gabriela Buarque

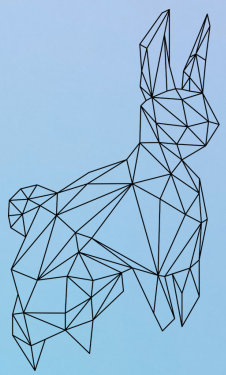
José Renato Laranjeira de Pereira

Tayrone Chiavone

Victor Mulin

REVISÃO

Cynthia Picolo



LAPIN

LABORATÓRIO DE POLÍTICAS PÚBLICAS E INTERNET

lapin.org.br

@lapin.br

/lapinbr

/lapinbr

LAPIN
LABORATÓRIO DE POLÍTICAS PÚBLICAS E INTERNET

O LAPIN

O Laboratório de Políticas Públicas e Internet (LAPIN) é um centro independente de pesquisa e ação de composição multidisciplinar e com sede na capital federal brasileira. Seu objetivo é apoiar o desenvolvimento de políticas públicas voltadas à regulação das tecnologias digitais por meio da pesquisa, articulação e da conscientização da sociedade.

Nosso trabalho consiste em:

1. **investigar, analisar e compreender os impactos** sociais, econômicos, éticos e jurídicos causados pelas tecnologias
2. **informar, incluir e ensinar** pessoas
3. **propor soluções inovadoras** para os desafios e oportunidades ao Brasil.

COMO FAZEMOS ISSO?

pesquisas interdisciplinares
desenvolvimento de projetos
ensino
comunicação acessível
articulação independente

lapin.org.br

LAPIN
LABORATÓRIO DE POLÍTICAS PÚBLICAS E INTERNET

SUMÁRIO

I. INTRODUÇÃO

O Laboratório de Políticas Públicas e Internet (LAPIN) saúda a iniciativa da instalação da Comissão de Juristas responsável por subsidiar a elaboração de minuta de substitutivo para instruir a apreciação dos Projetos de Lei n° 5.051/2019, 21/2020, e 872/2021, que têm como objetivo estabelecer princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da inteligência artificial (IA) no Brasil.

Apesar das inúmeras questões sobre diversidade e representatividade na composição da Comissão[1] e do curto tempo destinado para discussões e proposição de um substitutivo que abarque as complexidades envolvendo a IA no contexto brasileiro, o LAPIN apresenta suas contribuições e se coloca à disposição para dialogar e construir junto ao grupo de juristas e ao Senado Federal.

Esta contribuição é dividida em sete partes. A **primeira** parte se debruça sobre os atores envolvidos no contexto de uma inteligência artificial, a fim de identificar os sujeitos e seus respectivos papéis. A **segunda** versa sobre o princípio da precaução e seus desdobramentos na análise regulatória da IA. A **terceira**, traz reflexões sobre a necessidade de se garantir e como operacionalizar a transparência sobre sistemas de IA, especialmente às pessoas ou grupo de pessoas afetadas.

A **quarta** parte analisa os riscos que estão envolvidos na inteligência artificial e a aplicabilidade do princípio da precaução nesse contexto. A **quinta,** introduz noções sobre a avaliação de impacto de inteligência artificial (AIIA), apresentando seu conceito e recomendações de realização. A **sexta** seção discute a questão da responsabilidade civil e as dificuldades de restringir a sua incidência a um regime tradicional dicotômico (subjetivo ou objetivo).

Na **última,** o LAPIN sugere uma proposta de redação a ser considerada quando da elaboração do substitutivo.

1 Coalizão de Direitos na Rede. **Regulação de Inteligência Artificial: um tema transversal que exige debate multissetorial e interdisciplinar**. Disponível em: https://direitosnarede.org.br/carta-aberta-regulacao-ia/. Acesso em: 09 mai 2022.

II. ATORES

Os atores de IA são aqueles que desempenham um papel ativo em todo o ciclo de vida do sistema de IA, desde aqueles que desenvolvem, operam ou são afetados pelo sistema.[2] Pensando nisso, é possível apontar ao menos **três atores**:[3]

- **Desenvolvedor:** pessoa física ou jurídica responsável pelo desenvolvimento do sistema de IA. Como desenvolvimento, entende-se tanto **(i)** a codificação do algoritmo de aprendizagem que irá basear o sistema de IA, como também **(ii)** a seleção dos dados (de treinamento) e a subsequente criação e/ou treinamento do modelo; [4]

- **Aplicador da tecnologia:** pessoa física ou jurídica responsável pela implementação e operação de um sistema de IA para atingir um determinado objetivo. Importa mencionar que há casos em que **o desenvolvedor também poderá ser considerado como aplicador da tecnologia**, e vice versa;

- **Sujeito ou grupo afetado:** pessoa física, ou grupo de pessoas, que sejam direta ou indiretamente sujeitos ou impactados pela decisão de um sistema de IA.

Ao considerar os incontáveis modelos de negócio que a inteligência artificial pode promover, inclusive aqueles que ainda serão criados, poderão ocorrer casos específicos em que não haja compatibilidade ou identificação dos atores aqui definidos com a situação real e complexa associada ao ciclo de vida do sistema de IA. Por isso, **essa é uma definição que ainda estará em contínuo amadurecimento**, podendo haver novos atores no futuro. Para os fins da presente contribuição, será utilizada a classificação acima: desenvolvedor, aplicador da tecnologia e sujeito ou grupos afetados.

---

2 Organização para a Cooperação e Desenvolvimento Econômico (OECD). **OECD framework for the classification of AI systems**. OECD Digital Economy Papers. n. 323, Fevereiro, 2022. Disponível em: https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1652284916&id=id&accname=guest&checksum=D1B84F5562825203012134 85FA41E9CE. p. 22. Acesso em: 11 mai. 2022.

3 ANDRADE, Norberto Nuno Gomes de; KONTSCHIEDER, Verena. **AI impact assessment: a policy prototyping experimente.** Open Loop, 2021, p. 74-75. Relatório técnico. Disponível em: https://openloop.org/wp-content/uploads/2021/01/AI_Impact_Assessment_A_Policy_Prototyping_Experiment.pdf. Acesso em: 31 mai 2022.

4 De acordo com Open Loop - Facebook AI, p. 74, desenvolvedor é "a pessoa física ou jurídica que desenvolveu o sistema automatizado de tomada de decisão. Este ator pode apenas fornecer o algoritmo de aprendizagem, mas é mais provável que seja a pessoa ou organização responsável pela seleção dos dados (de treinamento) e algoritmos de aprendizagem relevantes e a subsequente criação e/ou treinamento do modelo." (tradução livre)

III. PRINCÍPIO DA PRECAUÇÃO

A partir das opiniões expressas por relatores e autores de projetos de lei para a regulação da inteligência artificial nas audiências públicas recentemente organizadas nas duas casas do Congresso Nacional sobre o tema,[5] percebe-se que um dos objetivos dos parlamentares é o de criar um quadro principiológico norteador da inteligência artificial no Brasil. Pensando nisso, foram tecidas considerações sobre princípios na Nota Técnica do Substitutivo ao PL 21/2020,[6] publicada pelo LAPIN em 28 de setembro de 2021. Nesse momento, considerando a necessidade de afunilamento e amadurecimento da discussão, passa-se a abordar o princípio da precaução, como fator de extrema relevância nas diretrizes de regulação da IA.

A regulação de riscos visa regular atividades econômicas que trazem externalidades negativas, em qualquer campo. Sendo o risco definido geralmente como a probabilidade de um evento, observa-se que se trata de um fenômeno calculável, diferentemente do conceito de incerteza, que é impossível de ser mensurado.[7]

Entretanto, a regulação de riscos está centralmente preocupada tanto com o risco quanto com a incerteza, em que pese o risco seja o tema central das discussões sociológicas. O fato é que as sociedades modernas passaram a vivenciar experiências sem precedentes no que tange à percepção de riscos, sejam eles, sociais, econômicos ou ambientais.[8] Nesta esteira, surge um extenso debate quando se fala em estratégias de regulação de riscos.

De um lado, surgem estudiosos defendendo medidas antecipatórias ao risco catastrófico, sugerindo a aplicação do princípio da precaução na regulação de riscos. De outro, alguns defendem o sopesamento entre possíveis riscos e benefícios de determinada atividade. Este tradeoff entre bens e males é comumente chamado de ***cost-benefit analysis,***[9] abordagem regulatória que opera dentro de um contexto

5 Falas nesse sentido foram feitas tanto pelo autor do projeto, Dep. Eduardo Bismarck, quanto pela relatora Dep. Luísa Canziani durante a Audiência Pública do PL 21/2020 na Câmara dos Deputados e Sessão Temática do PL 872/2021 no Senado Federal.

6 Disponível em: https://lapin.org.br/2021/09/28/nota-tecnica-pl-21-2020/. Acesso em: 15 mai 2022.

7 BALDWIN et al. **Understanding Regulation: Theory, Strategy, and Practice.** 2. ed.: Oxford University Press, 2013. p. 83-84.

8 *Ibidem.*

9 Em tradução livre, "análise de custos e benefícios".

de racionalidade limitada[10] e vieses bem documentados na tomada de decisão.[11] Assim, importa melhor delimitar as diferenciações entre as duas abordagens.

A **abordagem precaucionária** traz, de modo pragmático, uma visão de aversão ao risco, seguindo a premissa *better safe than sorry*.[12] Isso se dá na concepção dos defensores da precaução de que não há possibilidade de fazer qualquer tipo de sopesamento entre riscos e benefícios quando a atividade em questão impõe riscos catastróficos.[13] Já a ***cost-benefit analysis*** faz parte de uma abordagem técnica, que visa antecipar danos valendo-se da frequência de suas ocorrências. Esta abordagem está associada à visão de que riscos podem ser tomados com base em evidências objetivas.

Considerando a insuficiência das atuais técnicas de análise probabilística do risco no campo da IA e que esta tecnologia está em constante evolução, a melhor abordagem a se adotar para sua regulação é a precaucionária. Isso porque o princípio da precaução surge primariamente com o intuito de ressignificar os institutos da responsabilização clássica, invertendo sua lógica, de modo a criar mecanismos *ex-ante*, já que a mera indenização por um evento catastrófico se torna inócua.

Apesar de ter suas raízes no campo do Direito Ambiental, o princípio da precaução pode ser inserido na regulação de qualquer atividade disruptiva que possa trazer riscos aos direitos fundamentais, incluindo nas esferas da saúde e do meio ambiente (natural, artificial, cultural e do trabalho).

No contexto da IA, as tensões referentes à opacidade gerada por algoritmos e a insuficiência da simples abertura do código fonte para dar explicabilidade à decisão

---

10 Racionalidade limitada é um tipo de modelo de julgamento para a tomada de decisão que se baseia em evidências, seguindo a seguinte estrutura: (i) formulação do problema; (ii) estruturação do problema, visualização de uma relação entre as suas partes e criação de um modelo; (iii) montagem técnica do problema; (iv) testagem e simulação do modelo com suas possíveis soluções; (v) definição de controles sobre a situação; (vi) implementação da solução.

11 BALDWIN *et al*. Op. cit., p. 96.

12 *Ibidem.* Em tradução livre, “melhor prevenir do que remediar”.

13 A ideia de risco catastrófico de pronto remete a danos ambientais (meio ambiente natural). No entanto, para fins desta contribuição, sugerimos pensar o conceito como um evento que possa danificar um bem humano em escala global, como riscos ao ambiente democrático e digital. No contexto da atual discussão do Marco Regulatório da IA no Brasil, o conceito pode ser lido como um “risco irreparável”.

automatizada,[14] reforçam o campo de incertezas científicas o qual está inserido o desenvolvimento desta tecnologia. Assim, a inserção do princípio da precaução no rol de princípios da proposta legislativa ganha propósito.

Neste ponto, é importante fazer uma breve diferenciação entre o princípio da precaução e o princípio da prevenção. Ambos trazem mecanismos regulatórios ex-ante, todavia, operam em lógicas diversas.

O **princípio da prevenção** visa mitigar possíveis riscos previamente comprovados, ou seja, o regulador irá propor salvaguardas baseando-se na coleta de dados que evidenciem determinado risco. Noutro giro, o **princípio da precaução** atua diante de dados insuficientes para quantificar e qualificar os riscos, buscando administrar a incerteza de determinada atividade. Deste modo, conclui-se que a **prevenção atua de forma reativa aos riscos**, enquanto a **precaução atua de forma proativa**.[15]

Diante das inúmeras expressões do que seja o princípio da precaução, críticos e defensores o classificam em versões **forte, moderada** e **fraca**.[16] De modo não exaustivo, é importante exemplificá-las e diferenciá-las:

Versão fraca

nessa perspectiva, o princípio da precaução é caracterizado pela preferência na coleta de dados objetivos que justifique a ação ou inação e uma preferência pela gestão de riscos. Ou seja, a incerteza ante às consequências de determinada atividade pode justificar sua regulação se houver motivos plausíveis para crer que tal atividade seja prejudicial. Deste modo, deve haver certa ênfase na coleta de dados que justifiquem tal crença.

14 FERRARI,Isabella;BECKER, Daniel; WOLKART,Erik Navarro; E. N. **Arbitrium ex machina:: panorama, riscos e a necessidade de regulação das decisões informadas por algoritmos**. Revista dos Tribunais, São Paulo, 2018, v. 995, n. 1, p. 8. Disponível em: http://governance40.com/wp-content/uploads/2018/11/ARBITRIUM-EX-MACHINA-PANORAMA-RISCOS-E-A-NECESSIDADE.pdf. Acesso em: 01 jun 2022.

15 ARAGÃO, Alexandra. **Aplicação nacional do princípio da precaução.** Associação dos Magistrados da Jurisdição Administrativa e Fiscal de Portugal , Lisboa, v. 1, n. 1, p. 5, jan 2013.

16 PARSONS, K.GARNETT J. **Multi-Case Review of the Application of the Precautionary Principle in European Union Law and Case Law.** Risk Analysis, Cranfield, v. 1, n. 1, p. 4, mai, 2016. Disponível em: https://www.researchgate.net/publication/303313626. Acesso em: 22 fev 2022.

Versão moderada

enquanto a versão fraca orienta a regulação mesmo diante de incertezas científicas, a versão moderada indica que a precaução exige a tomada de ações que visem reduzir as incertezas. Estabelecem-se, assim, controles que possam trazer maiores margens de segurança, levando a uma extensa investigação quanto à relação causa e efeito decorrente da incerteza, definindo também quais riscos são considerados aceitáveis. Portanto, a versão moderada mostra-se mais proativa na busca pela identificação dos riscos.[17]

Versão Forte

nesta versão, o princípio da precaução não se limita às ameaças de riscos graves e irreversíveis. Ela vai além e propõe a inversão do ônus da prova ao proponente da atividade econômica, fazendo com que este seja obrigado a clarificar as incertezas trazidas pela sua atividade antes de implementá-la. Pode-se dizer que o princípio da precaução, ao inverter o ônus da prova, traz um dever de *accountability*[18] para o proponente da atividade.[19] A aplicação forte tende a ter uma lógica de prevenção de riscos e não de gestão. A incerteza por si só sobre determinada atividade pode levar a ações rigorosas, como a inversão do ônus da prova até a proibição da atividade, mesmo não havendo robustas evidências para crer que a conduta seja prejudicial.[20]

A compreensão destas classificações é importante justamente porque a variação das salvaguardas em cada expressão precaucional pode servir como referência para o nível de segurança que se pretende atingir.

17 *Ibidem.*

18 Em português “prestação de contas”.

19 SUSTEIN, Cass R. Op. cit., p. 23.

20 PARSONS, K. G.. Op. cit., p. 4.

A título exemplificativo, a proposta europeia de regulação da IA (*AI Act*)[21] acena para uma expressão moderada do princípio da precaução por alguns motivos, como **(a)** o estabelecimento de critérios de classificação de riscos inaceitáveis; **(b)** a criação de sistemas de gestão de riscos implícitos, como indicado em seu art. 9°; **(c)** a determinação de ações proativas de avaliação de conformidade; **(d)** obrigações horizontais visando garantir o gerenciamento dos riscos e conformidade da atividade; **(e)** obrigações de transparência e explicabilidade em um nível suficiente para o permitir que pessoas interessadas e afetadas pela IA interpretem o resultado do sistema. Logo, mesmo que o princípio não seja explicitamente mencionado na proposta europeia, sua operacionalização se faz de modo intrínseco.[22]

Os legisladores brasileiros podem seguir este mesmo caminho, **colocando um nível de precaução *moderado* ou *forte* na proposta legislativa,** a depender do que será considerado como *riscos concretos* mencionados no art. 6°, III do PL 21/2020.[23] Caso o Brasil se inspire no modelo europeu criando hipóteses de riscos previamente inaceitáveis, a expressão da precaução estará presente em sua forma *moderada*. Por outro lado, se os legisladores optarem por um modelo que determine que a responsabilidade seja do desenvolvedor/aplicador da tecnologia em provar que determinada atividade é segura, invertendo o ônus da prova, estaríamos diante de uma versão *forte* do princípio.

Deste modo, é importante a inserção do princípio da precaução no rol de princípios do Marco Regulatório da IA brasileiro, justamente para que se tenha um norte para estipular medidas de gerenciamento de risco e salvaguardas ante as incertezas geradas pelo desenvolvimento tecnológico.

---

21 COMISSÃO EUROPEIA. **Proposta de regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de Inteligência Artificial (Regulamento Inteligência Artificial) e altera determinados atos legislativos da União**, Brussels, v. 1, n. 1, p. 4, jun 2021. Disponível em: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF. Acesso em: 7 jun 2022.

22 *Ibidem.* p.1-25. Neste sentido, PARSONS, K. G. Op. cit., p. 3-5..

23 Art. 6° Ao disciplinar a aplicação de inteligência artificial, o poder público deverá observar as seguintes diretrizes:
(...) III – gestão baseada em risco: o desenvolvimento e o uso dos sistemas de inteligência artificial deverão considerar os riscos concretos, e as definições sobre a necessidade de regulação dos sistemas de inteligência artificial e sobre o respectivo grau de intervenção deverão ser sempre proporcionais aos riscos concretos oferecidos por cada sistema e à probabilidade de ocorrência desses riscos, avaliados sempre em comparação com:
a) os potenciais benefícios sociais e econômicos oferecidos pelo sistema de inteligência artificial; e b) os riscos apresentados por sistemas similares que não envolvam inteligência artificial, nos termos do inciso V deste caput; (...).

IV. TRANSPARÊNCIA[24]

Como vimos acima, sistemas de inteligência artificial estão presentes em múltiplos aspectos da experiência humana e seus impactos têm sido cada vez mais expressivos e notados pela sociedade.

No entanto, a complexidade de seu funcionamento faz com que poucos desses sistemas sejam compreensíveis por humanos, o que dificulta compreender o real impacto que geram. Isso inclui desde o modo e grau de como eles modificam relações de trabalho e o acesso à informação, até os impactos que causam ao meio ambiente.[25] Como consequência, tal opacidade obstaculiza ainda a responsabilização daqueles que os desenvolvem ou utilizam por erros em que incorram.[26]

Tornar esses sistemas transparentes tem sido, portanto, considerado um instrumento importante para permitir que reguladores e usuários entendam melhor como esses modelos de tomada de decisão automatizada alcançam previsões específicas e como revisá-los em caso de erros, permitindo que, de um lado, desenvolvedores e aplicadores, e, do outro, usuários, tenham maior controle sobre esses sistemas.[27]

Apesar de suas limitações e do fato de não ser, na maioria das vezes, suficiente por si só, a transparência é um passo importante para que se possa compreender até que ponto e para quais fins podemos confiar em sistemas de IA.[28]

Exemplos de sistemas de IA cujo impacto negativo foi identificado por meio de avaliações garantidas graças a mecanismos de transparência incluem a identificação de

---

24 Recomendações baseadas nas conclusões preliminares de dissertação de mestrado ainda em elaboração. LARANJEIRA DE PEREIRA, José Renato. **Unveiling Mysteries: Responsive Regulation as a Means For Enforcing Machine-Learning Transparency**. Dissertação (Mestrado em Direito) – Faculdade de Direito da Universidade de Brasília, em fase de elaboração.

25 CRAWFORD, Kate. **The Atlas of AI.** New Haven And London: Yale University Press, 2021.

26 PASQUALE, Frank. **The Black Box Society: the secret algorithms that control money and information.** Cambridge, Massachusetts: Harvard University Press, 2015.

27 KAMINSKI, Margot E. **The Right to Explanation, Explained**. University of Colorado Law Legal Studies Research Paper, n. 18-24, 15 June 2018. Disponível em: https://scholar.law.colorado.edu/articles/1227/. Acesso em: 26 jun 2022.

28 DOSHI-VELEZ, Finale & KORTZ, Mason. **Accountability of AI Under the Law.** Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper, 2017. Disponível em: https://dash.harvard.edu/bitstream/handle/1/34372584/2017-11_aiexplainability-1.pdf. Acesso em: 26 jun 2022.

discriminação em sistemas de planos de saúde;[29] em sistemas preditivos de reincidência criminal, como o caso do sistema COMPAS, em que se descobriu que categorizava erroneamente muito mais pessoas negras como de alto risco do que o fazia com pessoas brancas;[30] e falhas em carros autônomos.[31] Vale ressaltar que todos esses casos foram identificados por pesquisadores ou organizações da sociedade civil independentes, o que ressalta a importância de garantir acesso e a possibilidade de avaliação aprofundada a esse tipo de ator a sistemas de IA.[32]

Arya *et al.*[33] identificaram que há uma lacuna entre o que a comunidade técnica está produzindo sobre transparência e o que os reguladores e a sociedade como um todo exigem desses sistemas. Uma razão para essa lacuna é a falta de uma definição precisa de como essas informações devem ser fornecidas, algo que se deve especialmente ao fato de que pessoas diferentes em ambientes diversos podem exigir diferentes tipos de explicações.

A **inteligibilidade de sistemas de IA não deve consistir necessariamente em uma descrição precisa e detalhada de como os algoritmos funcionam.**[34] Tal forma de fornecimento de informações pode levar, em vários contextos, a um excedente informacional que pode ser inútil ou até prejudicial, levando ao que Ananny e Crawford chamam de "opacidade estratégica".[35]

Um exemplo em que o acesso ao código de programação não teria muita utilidade seria sua entrega a um usuário de uma plataforma de mídia social que recebe desinformação e que pretende aprender mais sobre como as informações são direcionadas para sua conta, especialmente se for leigo em ciência da computação.[36]

---

29 OBERMEYER, Z. & MULLAINTHAN, S. Dissecting Racial Bias in an Algorithm that Guides Health Decisions for 70M People. **Proceedings of the Conference on Fairness, Accountability, and Transparency [online].** 2019. Disponível em: https://dl.acm.org/doi/10.1145/3287560.3287593. Acesso em: 26 jun 2022.

30 LARSON, J. *et al.*. How We Analyzed the COMPAS Recidivism Algorithm (2016). **ProPublica**. Disponível em: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm. Acesso em: 26 jun 2022.

31 WILSON, B., HOFFMAN, J. *et al.* **Predictive Inequity in Object Detection**. Arxiv. Disponível em: https://arxiv.org/pdf/1902.11097.pdf. Acesso em: 26 jun 2022.

32 ALÌ, Gabriele Spina; YU, Ronald. Artificial Intelligence between Transparency and Secrecy: from the EC white paper to the AYA and beyond (2021).. **European Journal Of Law And Technology,** v. 12, n. 3, p. 1-25. Disponível em: https://www.ejlt.org/index.php/ejlt/article/view/754. Acesso em: 26 jun 2022.

33 ARYA, Vijay et al. One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques (2019). Arvix. Disponível em: https://arxiv.org/abs/1909.03012. Acesso em: 26 jun 2022.

34 BUCHER, Taina. **If... then**: **algorithmic power and politics.** Oxford University Press, 2019.

35 ANANNY, Mike; CRAWFORD, Kate. **Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability.** New Media & Society, v. 20(3), pp. 973–989, 2016, p. 979. Disponível em: https://journals.sagepub.com/doi/10.1177/1461444816676645. Acesso em: 26 jun 2022.

36 BUCHER, Taina. Op. cit.

Diferentemente, um especialista em auditoria de um sistema de reconhecimento facial provavelmente teria interesse em entender melhor este código ou mesmo acessar os bancos de dados que o sistema utiliza para aprendizado, a fim de analisar se tal banco de dados é discriminatório ou não. Considerando que um nível ótimo de transparência depende da pessoa e do ambiente em que a decisão automatizada ocorre, é importante ter em mente que os níveis e meios de fornecer informações variam de caso para caso.[37]

Isso demonstra o **caráter contextual** da transparência em sistemas de IA.

Diante disso, como proceder a uma regulamentação desses sistemas que permita sua compreensão não só por indivíduos e grupos impactados, como também entes reguladores, auditores, pesquisadores, ativistas e as próprias desenvolvedoras e aplicadoras desses sistemas, de modo a garantir o exercício de direitos e a utilização ética e lícita de aplicações de IA?

Uma norma que reja esse tema deve garantir flexibilidade suficiente para que informações sejam fornecidas de modo que se adequem a diferentes respostas às seguintes questões:[38]

(1) Quem é o destinatário das informações? Um regulador? Um advogado? Um indivíduo ou comunidade afetada pelo sistema? Uma organização da sociedade civil? Um especialista na área específica em que um sistema está sendo implantado? Um engenheiro/cientista de computação? Disso dependerá uma compreensão do tipo de informação a ser prestada e seu grau de complexidade/tecnicidade, e é importante que certo nível de acesso a informações seja garantido a todos esses atores.[39]

37 HAMON, Ronan *et al.* Bridging the Gap Between AI and Explainability in the GDPR: towards trustworthiness-by-design in automated decision-making (2022). **IEEE Computational Intelligence Magazine**, pp. 72-85. Disponível em: https://ieeexplore.ieee.org/document/9679770/. Acesso em: 26 jun 2022.

38 Perguntas baseadas nas conclusões preliminares de dissertação de mestrado ainda em elaboração. LARANJEIRA DE PEREIRA, José Renato. **Unveiling Mysteries: Responsive Regulation as a Means For Enforcing Machine Learning Transparency.** Op. cit. Vide também DOSHI-VELEZ, Finale & KIM, Been. Considerations for Evaluation and Generalization in Interpretable Machine Learning. In **Explainable and Interpretable Models in Computer Vision and Machine Learning**; Springer, Berlin, Germany, 2018, pp. 3–17; HAMON, Ronan et al. Bridging the Gap Between AI and Explainability in the GDPR: towards trustworthiness-by-design in automated decision-making. **IEEE Computational Intelligence Magazine,** pp. 72-85. Feb. 2022.

39 HAMON, Ronan *et al.* Bridging the Gap Between AI and Explainability in the GDPR: towards trustworthiness-by-design in automated decision-making (2022). **IEEE Computational Intelligence Magazine**, pp. 72-85. Disponível em: https://ieeexplore.ieee.org/document/9679770/. Acesso em: 26 jun 2022.

(2) Como as informações devem ser fornecidas? É necessário que o sistema seja auditado ou uma explicação é suficiente? No caso deste último, as informações podem ser transmitidas usando vocabulário especializado ou de uma forma que um leigo deva entender?

(3) Para que fins? É para montar um desafio legal? Para consertar um *bug?* Para avaliar a imparcialidade ou o viés de uma saída? Para avaliar as razões do resultado de um sistema que alega que um paciente tem câncer?

(4) Que tipo de informações são necessárias para atingir o objetivo pretendido? É suficiente avaliar o conjunto de dados de treinamento, informações gerais sobre o funcionamento do sistema ou, indo além do próprio sistema, seus gastos energéticos ou as escolhas econômicas e políticas que levaram ao desenvolvimento da aplicação? É importante saber como os dados foram coletados, tais como diretamente por humanos, sensores automatizados ou ambos? Além disso, é necessário saber como o sistema como um todo ou como ele conseguiu uma decisão específica, como, por exemplo, por que ele recusou crédito a um indivíduo?

(5) Que tipo de sistema de IA está sendo avaliado? Seria uma aplicação de reconhecimento facial? De *scoring* de crédito? Ou um sistema de personalização de conteúdo?

Isso inclui o fornecimento de informações como:

- Os dados que estão sendo usados para **treinamento** desses sistemas (*training data,* em inglês);
- Os dados de uma pessoa específica que foram utilizados (***input*)** para que o sistema emitisse determinada informação, predição, decisão, etc **(*output***, em inglês);
- Informações sobre como o **sistema funciona**, em termos gerais. Aqui, expressões como “lógica envolvida”, ou “critérios e parâmetros” envolvidos no funcionamento do sistema têm sido usadas em regulações como o Regulamento Geral de Proteção de Dados europeu (RGPD) e a LGPD;

- Se a **mudança** de determinado fator modificaria a decisão. Isso permitirá a uma pessoa afetada por sistema de IA entender se um fator específico foi determinante para a predição emitida, técnica a que estudiosos muitas vezes se referem como “explicações contrafactuais”;[40]
- Se o sistema pode tomar **decisões diferentes em casos semelhantes**, de modo a se entender se uma mesma variável teria pesos diferentes em casos semelhantes. Esse tipo de informação poderia, por exemplo, ajudar a detectar problemas de acurácia ou até vieses discriminatórios em sistemas;
- Qual o **gasto de energia** incorrido para seu treinamento e operacionalização e, no caso de sistemas que dependam de dispositivos específicos, a origem de seus componentes e planos para descarte, de modo a compreender seus impactos ambientais;
- Em última instância, considerando não só o grau de sensitividade enquanto segredo industrial, mas também a complexidade desse tipo de informação e sua inefetividade em muitos (ou talvez a maior parte dos) contextos, o **código do sistema.**

Tudo isso dependerá, mais uma vez, do caso concreto e das respostas às questões de 1 a 5 colocadas acima. Esse esforço deve ser direcionado de modo a garantir o que Frank Pasquale chama de “transparência qualificada” (*qualified transparency*). Na maioria das vezes, **não é ter acesso ao código de um sistema que nos ajudará** a resolver um problema relacionado ao funcionamento ou ao ambiente (práticas comerciais abusivas, custos ambientais) que circunda um sistema de IA, **mas, em vez disso, ter acesso a "revelações limitadoras, a fim de respeitar todos os interesses envolvidos em uma determinada informação"**.[41]

Além disso, faz sentido que essa prestação de informações seja garantida em pelo menos quatro esferas:

---

40 WACHTER, Sandra; MITTELSTADT, Brent and RUSSEL Chris. Counterfactual Explanations Without Opening The Black Box: Automated Decisions and the GDPR. **Harvard Journal Of Law & Technology**, [s. l], v. 31, n. 2, p. 841-887, Spring 2018. Disponível em: https://jolt.law.harvard.edu/assets/articlePDFs/v31/Counterfactual-Explanations-without-Opening-the-Black-Box-Sandra-Wachter-et-al.pdf. Acesso em: 26 jun 2022.

41 PASQUALE. Op.cit. p 142.

1. Por meio de **registro**, pelo desenvolvedor e/ou aplicador do sistema de IA em uma **base de dados de acesso público a ser mantida pelo Estado** que deve conter todos os sistemas aplicados no Brasil ou sobre indivíduos localizados no país. A inspiração para essa sistemática pode ser encontrada no Artigo 51 e Anexo VIII do *AI Act* europeu, mas desde que traga também mais informações sobre o funcionamento do sistema. Isso inclui:
    a. informações sobre **consumo de energia do sistema**, bem como quais medidas foram tomadas para reduzir seu impacto ambiental;
    b. informações sobre **tipos de dados** tratados, sobre as **bases de treinamento**, sobre os **parâmetros** utilizados pelo sistema e **outros dados** relevantes sobre seu funcionamento (como é a supervisão humana, como dados são categorizados, critérios, lógica do sistema, etc);
    c. informações sobre o **perfil das pessoas** envolvidas na categorização de dados, desenvolvimento e aplicação do sistema, o que inclui nível de escolaridade e potencialmente informações relativas a grupos minoritários aos quais essas pessoas possam pertencer. Aqui, cabe reflexão sobre de que forma isso pode ser garantido de modo a auxiliar na identificação do nível de diversidade envolvido no desenvolvimento e aplicação da tecnologia sem, ao mesmo tempo, incorrer em violações à proteção de dados pela abertura de dados sensíveis dessas pessoas; e
    d. demais informações a serem incluídas podem ser encontradas também no *AI Act*, em seu Anexo IV, relativo à documentação técnica a ser fornecida.

2. Elaboração de **relatórios de impacto** como o descrito nesta contribuição, que incluam a avaliação de potenciais riscos a direitos fundamentais, e com sua disponibilização ao público em versão que proteja informações confidenciais. Exemplos para esse tipo de restrição de conteúdo encontram-se, por exemplo, em processos do Conselho Administrativo de Defesa Econômica (CADE), em que informações sensíveis de empresas são apresentadas por estas em versões confidenciais de processos administrativos dada a confiança de que serão devidamente protegidas pela autarquia;

3. **Explicações dirigidas** a pessoas ou grupos sobre os quais possam repercutir efeitos, que não se restringem aos jurídicos, a seus interesses advindos do uso de sistema de IA. A linguagem e o formato da explicação devem levar em conta as cinco questões apresentadas anteriormente, com especial atenção ao nível de alfabetização digital do potencial receptor e o conhecimento que tem sobre IA e outras tecnologias digitais, bem como as particularidades relativas à origem regional, cultural, social da pessoa ou do grupo e o nível de especialização no campo em que o sistema é aplicado. A explicação também deve ser suficiente para que seu destinatário possa contestar e prestar contas da produção do sistema ou da decisão que ele influenciou, inclusive através do exercício do direito da pessoa de reclamar a uma autoridade supervisora (direito à transparência/explicabilidade);

4. **Mecanismos de obtenção de informações mais aprofundadas** por parte de reguladores, auditores, pesquisadoras, organizações da sociedade civil ou mesmo pessoas ou grupos interessados, havendo a possibilidade, inclusive, de realização de auditoria pelo ente regulador ou, mediante sua autorização, pelos outros grupos mencionados.

Ainda, vale ressaltar **a importância de que sistemas de IA utilizados pelo poder público sejam por padrão transparentes e explicáveis**, de forma similar a como o determina, por exemplo, a lei francesa n. 2016-1321, a *Loi pour une République Numérique*.[42]

O nível atual de opacidade de sistemas de IA vai além do que indústrias já consolidadas e que possuem interesses comerciais de valores altíssimos, como a farmacêutica, possuem. Por isso, a complexidade e o fato de guardar segredos industriais não podem ser aspectos que motivem a falta de provimento de informações significativas e adequadas sobre sistemas de IA, e isso exige uma mudança de postura regulatória.

Como afirmado por Asghari *et al*, "ninguém deveria se sujeitar a normas e instituições que não podem se justificar [ou explicar] baseadas em razões que não podem contestar".[43] Por isso, t**ransparência deve ser um imperativo** a ser garantido em uma regulação sobre qualquer tema que cause impactos negativos à sociedade e ao planeta como um todo. Sistemas de IA não são exceção.

42 Disponível em: https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000033202746/. Acesso em: 26 jun 2022.

43 ASGHARI, Hadi et al. **What to explain when explaining is difficult?**: an interdisciplinary primer on XAI and meaningful information in automated decision-making. Berlin: Alexander von Humboldt Institute For Internet And Society, 2021, p. 9. Disponível em: https://www.hiig.de/publication/what-to-explain-when-explaining-is-difficult-an-interdisciplinary-primer-on-xai-and-meaningful-information-in-automated-decision-making. Acesso em: 26 jun 2022.

V. RISCOS DA IA

Apesar dos aspectos benéficos do desenvolvimento e utilização de tecnologias de inteligência artificial, é preciso ter cuidado e atenção aos inúmeros riscos que esses sistemas podem provocar aos direitos fundamentais.

Dentre os principais direitos fundamentais em risco, pode-se apontar o **direito à privacidade** (art. 5º, X, CF/88), à **proteção de dados pessoais** (art. 5º, LXXIX, CF/88), à **autodeterminação pessoal**, à **liberdade de expressão** (art. 5º, IX, CF/88), à **liberdade de informação** (art. 5º, XXXIII, CF/88), à **liberdade de reunião pacífica** (art. 5º, XVI, CF/88), à **igualdade e à não discriminação** (art. 5º, caput, CF/88) e ao **devido processo legal** (art. 5º, LIV, CF/88).

O **risco à privacidade** reside no potencial que as tecnologias de inteligência artificial têm de coletar, processar e armazenar dados em larga escala, como ocorre, por exemplo, nas tecnologias de reconhecimento facial. Essas tecnologias de vigilância permitem monitorar qualquer pessoa que transite naquele espaço, indistintamente, e com quem se relacionam. O Alto Comissariado de Direitos Humanos das Organizações das Nações Unidas também aponta o perigo da possibilidade de países e empresas utilizarem estas tecnologias para monitorar, analisar e até mesmo prever o comportamento de seus cidadãos ou clientes.[44]

Já a **proteção de dados pessoais** é ameaçada, por exemplo, pelo risco de uso e compartilhamento indevido de dados coletados através de inteligência artificial sem o consentimento do titular, sem transparência sobre o tratamento e sem análise aprofundada da proporcionalidade entre necessidade de uso dos dados e a proteção aos direitos humanos.[45]

Outro risco ao direito de proteção de dados pessoais está ligado ao uso de técnicas de perfilização para categorizar pessoas em grupos específicos em sistemas de *credit score.* A partir da categorização, o sistema pode permitir ou limitar o acesso de uma

44 ORGANIZAÇÃO DAS NAÇÕES UNIDAS. Alto Comissariado de Direitos Humanos. **O direito à privacidade na era digital**. A/HRC/39/29 (03 de agosto de 2018), p. 2. Disponível em: https://undocs.org/A/HRC/39/29. Acesso em: 04 mai 2022.

45 *Ibidem*. p. 4-5.

pessoa a serviços e direitos, como concessão de empréstimos, financiamentos e até mesmo aposentadoria.[46]

O uso de técnicas de perfilização em conjunto com as de *microtargeting* pode também provocar **riscos à autodeterminação pessoal** dos indivíduos mediante manipulação de comportamento. A utilização de *microtargeting*, baseada em uma economia comportamental, permite que uma empresa identifique vulnerabilidades em grupos de indivíduos, se valendo dessas informações para oferecer produtos e serviços em momentos em que estejam mais propensos a uma tomada de decisão favorável à empresa.[47] Da mesma forma, essa tecnologia pode ser utilizada para influenciar o comportamento político de um eleitor.

Outro grande direito em **risco é a liberdade de expressão**. Inteligências artificiais para fins de moderação de conteúdo online, se utilizadas sem a devida diligência em relação aos direitos fundamentais, podem se transformar em instrumentos de censura,[48] [49] tanto pela iniciativa privada quanto pelo poder público.

Mais um aspecto da violação da liberdade de expressão, sendo também um **risco à liberdade de informação**, relaciona-se com o modo como as ferramentas de busca online e redes sociais são desenhadas. Tais sistemas definem as informações a que os usuários terão acesso de acordo com prévios interesses e buscas realizadas e, consequentemente, acabam selecionando/restringindo quais conteúdos serão acessados.

Segundo a organização de sociedade civil Access Now, os algoritmos por trás de tais sistemas têm sido responsáveis por criar "***echo chambers***" (câmaras de eco), espaços em que se propaga apenas uma mesma ideia ou perspectiva, colocando em risco a pluralidade de ideias e a diversidade de opiniões no ambiente online.[50]

---

46 VINAYAK, Vrinda. The Human Rights Implications of China's Social Credit System. **Oxford Human Rights Hub**. Disponível em: https://ohrh.law.ox.ac.uk/the-human-rights-implications-of-chinas-social-credit-system/. Acesso em: 04 jun 2022.

47 EBERS, Martin. **Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges**. Op. cit., p.32.

48 ACCESS NOW. **Human rights in the age of human intelligence**. 2018, p. 22. Relatório técnico. Disponível em: https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf. Acesso em: 29 jun. 2022.

49 LATONERO, Mark. **Governing artificial intelligence: upholding human rights & dignity**. Data&Society, 2018 , p. 14-15. Disponível em: https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf.Acesso em: 29 jun. 2022.

50 ACCESS NOW. Op. cit. p. 23.

A utilização de tecnologias de IA para moderação de conteúdo online também pode implicar **violações à liberdade de reunião pacífica**. Redes sociais e sítios eletrônicos são meios utilizados por movimentos sociais, grupos e coletivos para organizar reuniões e protestos. A inteligência artificial pode ser usada para retirar do ar total ou parcialmente grupos, páginas ou conteúdos direcionados a organizar encontros e colaborações de grupos.[51]

Além disso, tem sido igualmente debatido como o uso de tecnologias de reconhecimento facial (TRF) em espaços publicamente acessíveis restringe indevidamente oo direito à reunião pacífica. O uso do reconhecimento facial em contextos de protestos ou reuniões pode ter um "***chilling effect***", ou seja, um efeito dissuasivo.[52] Isto é, pessoas podem se sentir menos confortáveis em participar de protestos e reuniões se souberem que poderão ser identificadas. Esse desconforto pode levá-las a deixarem de se reunirem, limitando o seu exercício da liberdade de reunião e afetando o espaço de discussão democrática.

A controvérsia a respeito do uso de tecnologias de reconhecimento facial não se restringe ao impacto negativo à liberdade de reunião pacífica. Estas tecnologias vêm sendo discutidas principalmente pelo seu potencial discriminatório. Joy Buolamwini e Timmit Gebru demonstraram em 2018 que, apesar de as TRF desenvolvidas pela IBM, Microsoft e Face++ possuírem taxas de acerto expressivas, elas performaram diferente em relação a distintos grupos demográficos.[53] As pesquisadoras identificaram que todas as tecnologias possuíam uma taxa de acerto maior para homens brancos e performavam com menos acertos quando as pessoas avaliadas eram negras e do sexo feminino. As inteligências artificiais utilizadas para processos de recrutamento de novos empregados também estão sob a mira de críticos por viabilizarem, consciente e inconscientemente, **práticas discriminatórias.**

Por fim, os sistemas de inteligência artificial também podem provocar **riscos ao devido processo legal** através do uso de algoritmos decisórios quando não atendem aos critérios de transparência e explicabilidade em decisões tomadas exclusivamente de forma automatizada.

---

51 *Ibidem.*

52 *Ibidem.*

53 BUOLAMWINI, Joy; GEBRU, Timnit. **Gender Shades: Intersectional Accuracy Disparities in Commencial Gender Classification**. Conference on Fairness, Accountability, and Transparency. Proceedings of Machine Learning Research, 81:1–15, 2018. Disponível em: https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf, p. 1-15. Acesso em: 15 mai 2022.

Um caso conhecido refere-se ao **COMPAS** (*Correctional Offender Management Profiling for Alternative Sanctions*), um algoritmo utilizado no setor da justiça criminal Americano que realiza uma avaliação de risco sobre cada condenado com a finalidade de prever a potencial probabilidade de reincidência criminal desse indivíduo. O sistema, portanto, interfere diretamente na possibilidade de liberdade condicional do preso sem que houvesse qualquer intervenção humana ou que apresentasse justificativa ou fundamentação para a decisão.[54]

Vislumbra-se, então, que o uso de tecnologias de inteligência artificial pode ter sérias implicações no gozo de direitos fundamentais.

a) Formas de delimitação de riscos (gradação de riscos)

Atualmente, ainda existe uma ampla discussão a nível internacional sobre como diferenciar, de modo preciso, o que seria uma atividade de baixo risco de uma de alto risco. Esse debate culminou em duas propostas, sendo uma delas **prescritiva** e a outra **procedimental.**[55]

A **proposta prescritiva** é aquela ancorada em classificações rígidas que indicam previamente o tipo de utilização e o tipo de setor em que a aplicação poderá ser considerada de alto risco, como, por exemplo, a utilização de sistemas de IA para identificação de padrões de doenças no setor da saúde.[56] Essa classificação demanda a observação de um **requisito duplo e cumulativo** para caracterizar um sistema de IA como de alto risco. No caso em questão, tanto o setor quanto a utilização pretendida devem representar riscos significativos relativos à proteção da segurança, aos direitos dos consumidores e aos direitos fundamentais em setores de saúde, transportes e energia, por exemplo.[57]

---

54 ANGWIN, J. et al. **Machine Bias.** ProPublica, 2016. Disponível: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Acesso em: 15 maio 2022.

55 ANDRADE, Norberto Nuno Gomes de; KONTSCHIEDER, Verena. Op. cit. p. 11-12.

56 COMISSÃO EUROPEIA. **Livro Branco sobre a Inteligência Artificial: Uma abordagem europeia virada para a excelência e a confiança** (2020). Disponível em: https://op.europa.eu/pt/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1, p.20. Acesso em: 07 jun 2022.

57 *Ibidem*.

Por exigir uma definição prévia e prescritiva a respeito dos setores em que um sistema de IA pode ser considerado de alto risco, pode-se abrir lacunas de interpretação e aplicação da lei. Isso porque sempre haverá evolução da tecnologia, e as novas aplicações desenvolvidas, mesmo que de alto risco, não serão legalmente identificadas dessa forma.

Por outro lado, a **proposta procedimental** defende que a definição de alto risco depende de um conjunto de etapas e indagações destinadas a alcançar uma identificação do risco mediante diálogo, reflexão e uma análise qualitativa de informações associadas ao sistema de IA em si, e não somente ao setor e tipo de utilização a que se propõe.[58] A definição do risco, portanto, não se consagra aprioristicamente, e depende assim de uma análise do caso concreto.

Há um alto risco quando há uma chance significativa de que as decisões automatizadas tomadas pelo sistema de IA possam resultar ou influenciar em efeitos negativos nos direitos e liberdades dos sujeitos afetados. **A compreensão do que é risco inclui**, mas não se limita a, efeitos adversos que podem causar: [59]

interferência nos direitos fundamentais, como os direitos à igualdade e à não discriminação, à privacidade, e à liberdade de expressão

potenciais danos à saúde ou segurança do sujeito, como perda da vida ou danos corporais

potenciais danos psicológicos, como a autocensura, a perda de autoestima e a perda de autonomia pessoal

potenciais danos sociais ou econômicos, como danos financeiros, perda de propriedade ou restrição a serviços públicos ou privados

potenciais danos reputacionais ou de estigmatização

potenciais preconceitos ou discriminações injustas em relação ao sujeito, incluindo discriminação de preços, discriminação no emprego ou acesso diferenciado ou discriminatório aos serviços

potenciais danos coletivos, como uma perda de liberdade ou instabilidade econômica ou política

58 ANDRADE, Norberto Nuno Gomes de; KONTSCHIEDER, Verena. Op. Cit., p. 9.

59 *ibidem* pp. 30, 35-36 e 75-76.

Ressalte-se que os **sistemas de IA de médio ou baixo risco também devem ser objeto de regulação**, cumprindo parâmetros mínimos de transparência e segurança. Isso porque esses sistemas também têm o condão de afetar os interesses envolvidos, sendo necessário que possam ser efetivamente contestados. Para isso, é imprescindível que haja mínimo nível de transparência e mecanismos de diálogo com o sujeito afetado.

O direito de revisão de decisões automatizadas, nesse ponto, é fundamental para operacionalizar uma participação ativa do indivíduo na atividade da IA. No Brasil, esse direito é corroborado pelo **art. 20 da LGPD**,[60] que trata da solicitação de revisão de decisões tomadas com base em tratamento automatizado de dados pessoais, bem como pela eficácia horizontal dos direitos fundamentais, que estende o direito ao devido processo legal (art. 5º, LIV, CF/88) para as relações travadas entre os sujeitos afetados e os desenvolvedores e aplicadores de IA

Essa perspectiva parte da premissa de que, atualmente, as ameaças de violações aos direitos fundamentais também podem partir das relações privadas, sendo imprescindível que a eficácia constitucional cubra essas situações. Dessa forma, uma premissa fundamental no diálogo é a de que o direito ao devido processo legal não pode ser negligenciado como um dano colateral à adoção de tecnologias algorítmicas cada vez mais sofisticadas.[61]

Propõe-se, portanto, a **utilização equilibrada das duas correntes**, ou seja, que o Marco Regulatório para a Inteligência Artificial indique **(i)** uma **lista** de sistemas de inteligência artificial que representam um alto risco aos direitos e garantias fundamentais, notadamente os que representem um risco grave e/ou irreparável, como também **(ii)** determine **etapas** que permitam analisar aplicações de IA caso a caso, impedindo assim uma lacuna no ordenamento jurídico.

---

60 Art. 20. O titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade.

61 VILLASENOR, John; FOGGO, Virginia. **Algorithms and sentencing: What does due process require?** Disponível em: https://www.brookings.edu/blog/techtank/2019/03/21/algorithms-and-sentencing-what-does-due-process-require/. Acesso em: 28 jun 2022.

b) Hipóteses de riscos inaceitáveis

Além dos riscos acima exemplificados, existem ainda usos de sistemas de IA que violam direitos fundamentais, como aqueles que vulnerabilizam de sobremaneira a subsistência, a segurança e a saúde física e mental das pessoas. Nesses casos, os riscos são difíceis de evitar, mitigar ou compensar, razão pela qual se caracterizam como **expressivos e desproporcionais.** Estes são classificados como riscos inaceitáveis, devendo **seu uso ser banido da sociedade**.

Dito isso, a **primeira hipótese de risco inaceitável,** em que o uso de sistemas de IA deve ser definitivamente banido, está ligada às tecnologias de **reconhecimento biométrico e facial** para promoção de vigilância em massa no setor de segurança pública. A baixa acurácia atrelada aos vieses raciais encontrados nessa tecnologia aumentam a ocorrência da discriminação algorítmica.[62] Além disso, a coleta de dados biométricos e seu uso em ferramentas de estatísticas tendem a criar um ciclo retroalimentativo de marginalização de grupos vulneráveis, já que sistemas de IA se alimentam de dados que são historicamente enviesados pelo racismo estrutural.

Além disso, "enquanto as pessoas em espaços acessíveis ao público puderem ser instantaneamente identificadas, destacadas ou rastreadas, seus direitos humanos serão minados"[63], uma vez que a vigilância constante e desmotivada (sem justa causa em relação à pessoa que está sendo identificada) inibe que o indivíduo exerça livremente seus direitos e, inclusive, desenvolva sua personalidade.

Nessa mesma linha, pode-se ainda citar o **policiamento preditivo,** cujo propósito seria o de usar e analisar dados para monitoramento e identificação de situações suspeitas. Por serem sistemas baseados em conjuntos de dados originariamente opacos, racistas e inconsistentes, seu uso, além de não haver comprovação alguma de benefício para a segurança pública, da mesma forma aumenta a vigilância, práticas discriminatórias e desigualdades socioeconômicas.

---

62 ACCESS NOW. **Carta aberta para banimento global de usos de reconhecimento facial e outros reconhecimentos biométricos remotos que permitam vigilância em massa, discriminatória e enviesada (2021)**. Disponível em: https://www.accessnow.org/cms/assets/uploads/2021/06/BanBS-Portuguese.pdf, p. 3. Acesso em: 07 jun 2022.

63 *Ibidem*.

Assim, o uso dessas tecnologias deve ser banido, pois, caso contrário, estaremos promovendo ferramentas que fomentam a vigilância, a cultura do encarceramento de populações negras e vulneráveis e do punitivismo do sistema penal.

A **segunda hipótese de risco inaceitável** é o investimento e o desenvolvimento de **armas autônomas** baseadas em inteligência artificial e com potencial de causar morte em um confronto armado. Além de serem sistemas cuja finalidade por si só não é neutra, colocando o direito fundamental mais caro em risco, há mínima interferência de um ser humano no poder decisório e impossibilidade de exigir total responsabilização de quem os opera.

A **terceira hipótese** está ligada ao uso de sistemas de IA que se valem de vulnerabilidades físicas, emocionais e psicológicas para distorcer e manipular o comportamento de indivíduos ou grupo de indivíduos,[64] como o **reconhecimento de emoções**. Além dos riscos inerentes ao uso da IA para este fim, há falta de fundamento científico de que seja possível identificar emoções somente com base em expressões faciais.

Por fim, a **quarta hipótese** de risco inaceitável refere-se ao uso de sistemas de IA que valoram a confiança de um indivíduo ou de um grupo de indivíduos mediante análise de conduta social ou de características pessoais ou de personalidade (conhecidas ou preditivas), causando um tratamento prejudicial ou desfavorável capaz de injustamente limitar ou impossibilitar o regular exercício de direitos, como o **crédito social** (*social scoring*).[65]

Assim, pode-se citar como sistemas de IA que geram riscos inaceitáveis, devendo ser **banidos no ordenamento jurídico brasileiro,** as aplicações para reconhecimento facial em espaços públicos, policiamento preditivo, armas autônomas, reconhecimento de emoções e crédito social (*social scoring).* No entanto, vale ressaltar que outros casos os quais o uso deva ser banido podem surgir com o tempo. Sugere-se, portanto, que a futura legislação preveja a possibilidade de adição de novos casos pela autoridade reguladora competente.

64 COMISSÃO EUROPEIA. **Proposta de regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de Inteligência Artificial (Regulamento Inteligência Artificial) e altera determinados atos legislativos da União** (2021), Artigo 5º. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:52021PC0206. Acesso em: 08 jun 2022.

65 *ibidem.*

VI. AVALIAÇÃO DE IMPACTO DE INTELIGÊNCIA ARTIFICIAL (AIIA)

Avaliações, relatórios e diagnósticos de impacto são instrumentos que têm ganhado cada vez mais importância em uma sociedade na qual as ações humanas e empresariais podem provocar riscos de difícil ou impossível reparação. Atualmente, no Brasil existem ao menos três avaliações de impacto setoriais definidas por lei, sendo elas **(i)** a avaliação de impacto ambiental, **(ii)** a avaliação de impacto regulatório e **(iii)** o relatório de impacto à proteção de dados (RIPD).

No campo da IA, a **Avaliação de Impacto de Inteligência Artificial (AIIA)** é vista como um **instrumento de governança** que possibilita ao desenvolvedor ou aplicador da tecnologia identificar e reduzir possíveis riscos que determinado sistema de IA possa causar aos direitos e liberdades fundamentais.

Apesar do art. 5º, XVII, da Lei Geral de Proteção de Dados Pessoais dispor acerca do Relatório de Impacto à Proteção de Dados Pessoais, a AIIA se diferencia na medida em que o **RIPD** é uma avaliação de impacto que se concentra apenas nos **riscos que uma atividade de tratamento de dados** pode causar a uma pessoa física ou a um conjunto de pessoas cujos dados pessoais estão sendo tratados, ao passo em que **a AIIA é uma avaliação mais abrangente**.

A AIIA pode ser considerada um instrumento mais amplo porque sua condução exige que se protejam **direitos e liberdades fundamentais** que não se restringem ao tratamento de dados pessoais, mas que também envolvem a atividade do próprio algoritmo, da programação e da imprevisibilidade da máquina, englobando estes outros elementos que se unem para caracterizar uma inteligência artificial.

Como um instrumento de governança, a AIIA ajuda a manter a **transparência** e a **confiança** dos usuários na tecnologia, permitindo expor as capacidades e a finalidade do sistema de IA a todos aqueles que sejam por ela afetados. Além disso, a AIIA permite a gestão e o tratamento dos riscos que uma atividade pode causar, possibilitando a definição de técnicas e processos para mitigá-los ou evitá-los.

Assume, portanto, especial relevância em uma sociedade com forte tradição consumerista, onde **acesso à informação** figura como direito básico.

Também compete sublinhar o Princípio 19, apontado pela **Organização das Nações Unidas (ONU)** na carta de **Princípios Orientadores sobre Empresas e Direitos Humanos** de 2011, que afirma que para prevenir e mitigar os impactos negativos sobre os direitos humanos, as empresas devem integrar as conclusões de suas avaliações de impacto em funções e processos internos relevantes e tomar as medidas apropriadas.[66]

Sob a perspectiva econômica e de mercado, o projeto Open Loop recentemente confirmou os **benefícios da realização de AIIAs nos setores europeu e estadunidense de empresas e *startups***.[67] O estudo constatou que todos os participantes que conduziram uma AIIA se tornaram mais capazes de identificar e mitigar os riscos associados às suas aplicações, assim como de incorporar as melhores práticas e salvaguardas no design de seus produtos,[68] resultando em "maior eficiência e entrega mais rápida ao mercado, reduzindo os custos e riscos de interrupção posterior".[69]

---

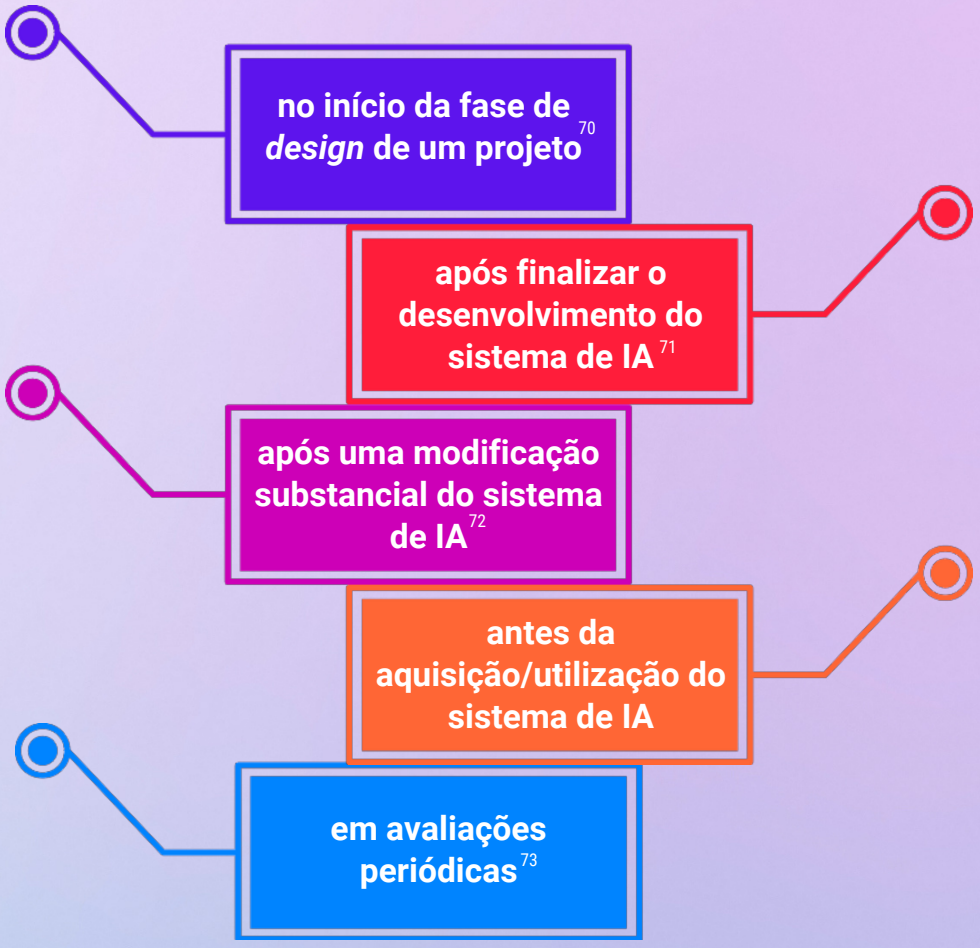
66 Escritório do Alto Comissariado das Nações Unidas para os Direitos Humanos. **Princípios Orientadores sobre Empresas e Direitos Humanos: Implementando os parâmetros 'Proteger, Respeitar e Reparar' das Nações Unidas.** (2011). Disponível em: https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf, p. 20. Acesso em: 10 jun 2022.

67 O Open Loop é um programa apoiado pelo Facebook AI, denominado "AI Impact Assessment: A Policy Prototyping Experiment". Trata-se de relatório que apresenta os resultados do programa de prototipagem de leis realizado com dez organizações europeias a respeito das avaliações de impacto de decisão automatizada. O Open Loop realizou um experimento de realização de AIIA com ao menos 10 empresas e startups que desenvolvem e operam sistemas de IA, como: Allegro.ai (Irlanda), RiAtlas (Itália), NAIX Technology (Alemanhã),Evo Pricing (Reino Unido), Keepler Data Tech (Espanha),Unbabel (Portugal), Feedzai (Estados Unidos), Rogervoice (França), Irida Labs (Grécia) e Reface (Ucrânia). Ver: ANDRADE, Norberto Nuno Gomes de; KONTSCHIEDER, Verena. Op. cit., p. 21-22.

68 CENTRE FOR INFORMATION POLICY LEADERSHIP. **Recommendations on Adopting a Risk-Based Approach to Regulating Artificial Intelligence in the EU** (2021). Disponível em: https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_risk-based_approach_to_regulating_ai__22_march_2021_.pdf, p. 3-4. Acesso em: 05 mai 2022.

69 *Ibidem*. Conclusão semelhante pode se extrair de BRYNJOLFSSON, Eric; MCAFEE, Andrew. The business of Artificial Intelligence. **Harvard Business Review**, edição especial Artificial Intelligence, Boston, jul 2017. Disponível em: https://hbr.org/2017/07/the-business-of-artificial-intelligence. Acesso em: 05 mai 2022.

a) Realização da AIIA durante o ciclo de vida da IA



1. No início da fase de *design* de um projeto: permite que os desenvolvedores incluam considerações técnicas e legais ainda no *design* da aplicação, aumentando a viabilidade de desenvolvimento, de comercialização do produto e reduzindo custos, uma vez que posteriormente pode se tornar impossível ou muito custoso fazer qualquer alteração;[74]

70 CANADÁ. **Algorithmic Impact Assessment Tool**. Disponível em: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html#toc3-1. Acesso em: 05 mai 2022.

71 *ibidem*.

72 EPC | Platform for the Information Society. **Artificial Intelligence Impact Assessment** (2018). Disponível em: https://ecp.nl/wp-content/uploads/2019/01/Artificial-Intelligence-Impact-Assessment-English.pdf, p. 59-61. Acesso em: 07 jun 2022. REISMAN, Dillon; SCHULTZ, Jason; CRAWFORD, Kate; WHITTAKER, Meredith. **Algorithmic Impact Assessment in the public sector** (2018). AI Now. Disponível em: https://ainowinstitute.org/aiareport2018.pdf, p-10. Acesso em: 31 mar 2022.

73 REISMAN, Dillon; SCHULTZ, Jason; CRAWFORD, Kate; WHITTAKER, Meredith. Op. cit. p. 10.

74 EPC | Platform for the Information Society. Op. cit. p. 29.

2. Após finalizar o desenvolvimento do sistema de IA: a realização da AIIA uma segunda vez, agora ao final da fase de desenvolvimento, permitirá que o desenvolvedor e sua equipe validem os resultados do sistemas de IA, isto é, verifiquem se eles refletem exatamente aquilo que foi projetado e desenvolvido sem que traga efeitos negativos, planejados ou não, sobre direitos.[75]

3. Após uma modificação substancial do sistema de IA: é recomendável que se realize uma nova AIIA sempre que houver:

a) uma modificação substancial acerca da forma de operação do sistema de IA ou nas bases de dados de treinamento do sistema;

b) uma mudança/ampliação da sua finalidade originalmente planejada;[76] ou

c) caso o sistema de IA atue com aprendizado de máquina de modo que a sua tomada de decisão possa ser influenciada por novos dados e informações ao longo do tempo.[77]

4. Antes da aquisição/utilização do sistema de IA: é recomendável que o aplicador da tecnologia realize uma AIIA quando, ao adquirir um sistema de IA, o operacionalize a partir da sua própria base de dados e informações. Essa é uma medida que promove transparência e faz com a que empresa tenha ganho reputacional perante os parceiros comerciais e frente ao próprio consumidor final.

5. Avaliação periódica: independentemente se o responsável é o desenvolvedor ou o aplicador da tecnologia do sistema de IA, é recomendável que se defina um intervalo de tempo para avaliar periodicamente a aplicação. Há recomendações para que uma avaliação seja conduzida anualmente ou a cada dois anos.[78]

---

75 CANADÁ. **Algorithmic Impact Assessment Tool**. Disponível em: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html#toc3-1. Acesso em: 05 mai 2022.

76 EPC | Platform for the Information Society. Op. cit. p. 61.

77 REISMAN, Dillon; SCHULTZ, Jason; CRAWFORD, Kate; WHITTAKER, Meredith. Op. cit. p. 10.

78 *Ibidem.* Ver também ANDRADE, Norberto Nuno Gomes de; KONTSCHIEDER, Verena. Op. cit. p. 76.

Ao identificar a necessidade da AIIA, recomenda-se que a sua realização tenha o aconselhamento de uma equipe jurídica especializada e o apoio de especialistas independentes de áreas como TI, sociologia ou ética,[79] de modo a identificar os riscos a partir de uma perspectiva multidisciplinar.

Também é importante permitir que **terceiros externos à organização participem da AIIA** quando as aplicações de IA, a depender da amplitude,[80] da finalidade[81] ou do risco[82] associado, possam afetar significativamente a sociedade, ainda que em grupos específicos ou minorias.

A participação da sociedade civil, da academia e dos próprios sujeitos afetados pela aplicação pode ser viabilizada, por exemplo, por meio de *webinars,* audiências públicas, fóruns, listas de discussão, emissão de parecer e nota técnica ou qualquer outra via que permita o amplo debate de ideias.

É recomendável, ainda, optar por publicizar um resumo ou o resultado de uma AIIA. Isso permite que a sociedade compreenda o funcionamento e os riscos do sistema, demonstrado que a organização que desenvolve e opera a tecnologia atua com transparência e responsabilidade social. Nesses casos, resguardadas informações sensíveis e segredos comerciais,[83] é recomendável que se apresente à sociedade apenas um resumo qualificado da AIIA ou sua respectiva conclusão, contendo informações mínimas para que quem a analise compreenda os riscos envolvidos no sistema.[84]

---

79 INFORMATIONa COMMISSIONER'S OFFICE (ICO). **Guidelines on Data protection impact assessment**. Maio, 2018. Versão 1.0.124. Disponível em: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/, p. 34. Acesso em: 05 mai 2022.

80 Por exemplo, um sistema de IA privado que pretende utilizar reconhecimento facial em locais públicos, como o metrô ou outra zona de grande movimento de pessoas.

81 Por exemplo, o desenvolvimento de *hardwares* ou *softwares* baseados em inteligência artificial que possibilitem grande automação de processos e procedimentos e que, consequentemente, possam causar impacto no mercado de trabalho.

82 Por exemplo, o desenvolvimento de um sistema de IA voltado para moderação de conteúdo de redes sociais.

83 O ordenamento jurídico brasileiro protege o segredo comercial e industrial, existindo diversos artigos da própria LGPD que colocam tais segredos como um limite à própria transparência e ao direito à informação do titular de dados, como o artigo 6°, VI, LGPD.

84 GRUPO DE TRABALHO DO ARTIGO 29º PARA A PROTEÇÃO DE DADOS. **Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD) e que determinam se o tratamento é «suscetível de resultar num elevado risco» para efeitos do Regulamento (UE) 2016/679** (2017). Disponível em: https://www.cnpd.pt/media/f0ide5i0/aipd_wp248rev-01_pt.pdf, p. 21. Acesso em: 05 mai 2022.

b) Recomendações metodológicas

A estrutura a seguir foi inspirada em iniciativas, relatórios e artigos científicos que defendem o uso do RIPD em sistemas de IA, bem como nas propostas de avaliações de impacto voltadas para decisões automatizadas. Vale ressaltar, no entanto, que a proposta de metodologia a seguir é uma estrutura orientadora, uma vez que organizações podem adaptar o modelo face às peculiaridades de seu cotidiano de trabalho.

Proposta de metodologia para Avaliação de Impacto de Inteligência Artificial

ETAPA 1
Descrição detalhada do sistema de IA

ETAPA 2
Avaliação da legalidade da finalidade do sistema de IA

ETAPA 3
Avaliação da necessidade de participação de terceiros

ETAPA 4
Identificação e avaliação dos riscos envolvidos

ETAPA 5
Definição de salvaguardas, medidas de segurança e medidas de proteção dos direitos e liberdades dos sujeitos

ETAPA 6
Avaliação da proporcionalidade entre finalidade e resultado

ETAPA 7
Riscos residuais e a necessidade de consulta prévia a uma Autoridade de Supervisão

ETAPA 8
Documentação e publicação

Etapa 1. Descrição detalhada do sistema de IA

A descrição do sistema de IA, etapa prevista em diversas propostas de AIIA,[85] deve responder a pelo menos seis questões:

- qual o **contexto** no qual o sistema de IA será utilizado?
- qual a respectiva **finalidade** do sistema de IA?
- qual a **lógica de operação** do sistema de IA?
- qual **tipo de dado** será utilizado e como será realizado o seu **treinamento?**
- quais são os **sujeitos afetados** de forma direta, indireta, intencional ou não intencional pelo sistema de IA?
- quais são os **benefícios** do sistema de IA?
- quais as **leis** que o sistema de IA deve respeitar?

Etapa 2. Avaliação da legalidade da finalidade do sistema de IA

Com base nas informações já apresentadas, é importante que se identifique a legalidade ou ilegalidade da finalidade do sistema e da maneira pela qual essa finalidade será alcançada.[86] Essa análise pode recair tão somente na questão legal, o que inclui o Código de Defesa do Consumidor, a Lei de Proteção de Dados Pessoais e leis e regulamentações específicas de cada setor, como saúde, trabalho, financeiro ou educação, por exemplo. Contudo, recomenda-se fortemente que também se realize uma avaliação ética do sistema de IA. Para isso, pode-se levar em consideração quadros de princípios éticos internacionalmente reconhecidos, como o da Organização para a Cooperação e Desenvolvimento Econômico (OCDE),[87] da Comissão Europeia,[88] da UNESCO,[89] entre outros.

---

85 INFORMATION COMMISSIONER'S OFFICER (ICO). **Guidance on AI and data protection**. Versão 0.0.22, jul 2020. Disponível em: https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection-0-0.pdf, p. 17. Acesso em: 31 mar 2022; ECP | Platform for the Information Society. **Artificial Intelligence Impact Assessment.** Op. cit. p. 42-43; ANDRADE, Norberto Nuno Gomes de; KONTSCHIEDER, Verena. Op. cit. p. 81.

86 *Ibidem,* p. 48.

87 ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **Recommendation of the Council on Artificial Intelligence** (2019). OECD/LEGAL/0449**.** Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Acesso em: 04 mai 2022.

88 UNIÃO EUROPEIA. Grupo de Peritos de Alto Nível sobre a Inteligência Artificial. **Orientações éticas para uma IA de confiança.** Disponível em: https://op.europa.eu/pt/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-pt. Acesso em: 05 mai 2022.

89 UNESCO. **Recomendação sobre a ética da inteligência artificial.** Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137_por. Acesso em: 28 jun 2022.

Etapa 3. Avaliação da necessidade de participação de terceiros

Poderão existir casos em que a amplitude e a finalidade de um sistema de IA desperte um interesse na sociedade e em outros agentes externos à organização. Por isso, sugere-se que se analise a possibilidade de participação da sociedade civil, da academia, do poder público e dos próprios sujeitos afetados na realização da AIIA, mesmo que de forma restrita ou somente opinativa.[90]

É possível que a avaliação de impacto precise de avaliadores, especialmente cientistas sociais e outros pesquisadores, que há muito estudam como a compreensão de raça, gênero e outras identidades sociais minoritárias estão ligadas às desigualdades dos sistemas.[91]

Caso o responsável pela a AIIA entenda ser adequada a participação de terceiros, é importante que se documente as opiniões e manifestações apresentadas. Por outro lado, caso o responsável pela AIIA chegue a uma conclusão contrária, recomenda-se que essa decisão seja devidamente fundamentada e documentada junto com a AIIA.

Etapa 4. Identificação e avaliação dos riscos envolvidos

É fortemente aconselhado que se adote uma abordagem em que os direitos fundamentais são reconhecidos como os valores objeto de proteção,[92] identificando, a partir deles, os riscos aos sujeitos afetados pela decisão de um sistema de IA. Nessa perspectiva, pode-se citar, **de forma não exaustiva**, ao menos quinze valores que podem sofrer efeitos negativos provenientes de sistemas de IA, dentre eles:

---

90 GRUPO DE TRABALHO DO ARTIGO 29º PARA A PROTEÇÃO DE DADOS. **Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD) e que determinam se o tratamento é «suscetível de resultar num elevado risco» para efeitos do Regulamento (UE) 2016/679**, abril de 2017, p. 8. Disponível em: https://www.cnpd.pt/media/f0ide5i0/aipd_wp248rev-01_pt.pdf. Acesso em: 15 mai 2022.

91 DATA & SOCIETY. **Assembling Accountability: Algorithmic Impact Assessment for the Public Interest.** Disponível em: https://datasociety.net/wp-content/uploads/2021/06/Assembling-Accountability.pdf. Acesso em: 15 mai 2022.

92 GETTING THE FUTURE RIGHT – **Artificial intelligence and fundamental rights. European Union Agency for Fundamental Rights.** 2020, p. 87. Disponível em: https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights; JANSSEN, Heleen. Detecting new approaches for a fundamental rights impact assessment to automated decision-making. **International Data Privacy Law**, Oxford, v. 10, n. 1, p. 76-106, fevereiro 2020. Disponível em: https://doi.org/10.1093/idpl/ipz028. Acesso em: 15 mai 2022.

autonomia humana	liberdade de reunião
dignidade humana	liberdade de pensamento
igualdade e equidade	emprego
privacidade e proteção de dados	Estado de Direito e democracia
saúde física e mental	devido processo legal
educação	asilo
liberdade de expressão	meio ambiente

A depender da finalidade do sistema de IA, e dos respectivos dados utilizados, estes são alguns valores que podem sofrer efeitos negativos relacionados às decisões automatizadas. Nesse sentido, ao perceber que um desses valores está em risco, o responsável pela condução da AIIA deverá identificar, respectivamente e de forma concreta, os possíveis impactos negativos que o sistema de IA pode causar ao determinado valor. Após essa identificação, **o responsável pela AIIA deverá avaliar a respectiva probabilidade de ocorrência e de gravidade do dano caso tal evento adverso efetivamente ocorra**.[93]

Etapa 5. Definição de salvaguardas, medidas de segurança e medidas de proteção dos direitos e liberdades dos sujeitos

A fim de definir medidas adequadas a mitigar ou eliminar os riscos, o responsável pela condução da AIIA deverá, primeiramente, registrar a fonte de cada risco identificado na etapa anterior. A partir do momento em que se identifica o risco e sua respectiva origem, deve-se analisar e implementar medidas adequadas para sua eliminação ou mitigação a um nível aceitável.

Etapa 6. Avaliação da proporcionalidade entre finalidade e resultado

A presente etapa é uma inspiração lógica do princípio legal da proporcionalidade, onde a sua essência está diretamente relacionada com valores e idéias como justiça, equidade, bom senso, prudência, moderação, justa medida e proibição de excesso.[94] Nesse contexto, a avaliação da proporcionalidade[95] irá exigir que responsável pela AIIA identifique **(i)** se a forma como o sistema de IA foi planejado definitivamente possibilita e contribui no alcance dos propósitos pretendidos; e **(ii)** se não há um meio alternativo, razoável e menos intrusivo de se conseguir os mesmo resultados.[96]

---

93 IEEE. **Recommended practice for assessing the impact of autonomous and intelligent systems on human well-being.**IEEE 7010-2020, p. 22. Disponível em: https://standards.ieee.org/standard/7010-2020.html. Acesso em: 05 mai 2022; ANDRADE, Norberto Nuno Gomes de; KONTSCHIEDER, Verena. Op. cit.,, p. 79-82.

94 LENZA, Pedro. **Direito constitucional esquematizado**. 12ª ed. rev., atual. e amp. São Paulo: Saraiva, 2008, p. 75.

95 A busca pela proporcionalidade pode ser encontrada no Regulamento Geral de Proteção de Dados europeu (RGPD), em seu artigo 35(7)(b).

96 ECP | Platform for the Information Society. **Artificial Intelligence Impact Assessment.** Op. cit. p. 47.

Etapa 7. Riscos residuais e a necessidade de consulta prévia a uma Autoridade de Supervisão

Quando riscos residuais forem considerados de nível aceitável,[97] o responsável pela condução da AIIA deverá identificá-los, fundamentar por que o risco residual é aceito e indicar quais medidas serão tomadas para limitar e/ou reparar possíveis danos se o risco se manifestar.[98] Caso o risco residual ainda resulte em um alto risco para os direitos e liberdades dos sujeitos afetados, não podendo, portanto, ser mitigado, além de ser devidamente identificado na AIIA, é recomendável que o responsável pela condução da AIIA consulte previamente uma autoridade supervisora antes de implementar e operar o respectivo sistema.[99]

Conforme afirmamos em audiência pública, consideramos necessário existir um regulador central que trabalhe como coordenador das tantas agências reguladoras setoriais que dividem competências sobre a miríade de setores em que se aplica IA, de modo a evitar decisões regulatórias difusas e por vezes contraditórias sobre esse tema.

Essa autoridade terá, como competências primárias, a conciliação desses atores, bem como a interpretação e o desenho de políticas comuns a setores diversos, o que inclui tanto atores governamentais quanto entidades auditoras, certificadoras e de padronização.[100]

Etapa 8. Documentação e publicação

A documentação é considerada uma etapa de encerramento da AIIA, onde se registram as respostas das etapas 1 a 6, de modo a manter registro que comprove a forma responsável que a empresa ou organização lida com a tecnologia. Ela será útil caso o responsável pelo sistema de IA precise apresentar explicações relativas ao modo de funcionamento ou aos possíveis efeitos provocados pela respectiva aplicação.

Esse também é o momento em que se deve deliberar sobre a conveniência e oportunidade de se publicar um resumo da AIIA, protegendo informações sigilosas e segredos comerciais. Essa decisão é recomendável no sentido de demonstrar maior responsabilidade social e criar um sentimento de confiança e transparência com parceiros, clientes e utilizadores.

---

97 Não há uma definição clara e objetiva do que será considerado risco aceitável. Por se tratar de uma avaliação procedimental, cada empresa ou organização chegará a sua própria conclusão sobre o que é ou não um elevado risco, assim como o que será um risco aceitável. Como exemplo pode-se citar a tecnologia de reconhecimento facial. Isso porque, enquanto empresas como a IBM reconheceram a gravidade dos riscos atrelados ao uso dessa tecnologia, e por isso interromperam o seu desenvolvimento, outras empresas continuam a apostar comercialmente em sua expansão.

98 ECP | Platform for the Information Society. **Artificial Intelligence Impact Assessment**. Op cit. p. 57-58.

99 ANDRADE, Norberto Nuno Gomes de; KONTSCHIEDER, Verena. Op. cit. p. 55.

100 Audiência pública - Comissão de Juristas do Senado Federal. Modelos de Regulação e Abordagem (28/04/2022).

VII. RESPONSABILIDADE CIVIL

O **inciso VI** do art. 6° do PL 21-A/2020 determina que a responsabilidade dos agentes que atuam na cadeia de desenvolvimento e operação de sistemas de inteligência artificial, salvo disposição legal em contrário, seja subjetiva, levando em consideração a efetiva participação desses agentes, os danos específicos que se deseja evitar ou remediar e a forma como esses agentes podem demonstrar adequação às normas aplicáveis, por meio de esforços razoáveis compatíveis com os padrões internacionais e as melhores práticas de mercado.

Primeiramente, ao restringir a responsabilidade à esfera subjetiva, o dispositivo desconsidera que a avaliação da responsabilidade civil como subjetiva ou objetiva depende do caso concreto,[101] notadamente quando se trata de inteligência artificial, cuja aplicação pode ocorrer nas formas e nos contextos mais distintos possíveis. As diferentes características da inteligência artificial trazem diferentes riscos e desafios regulatórios, o que se reflete também nos regimes de responsabilização.

A restrição trazida nesse dispositivo não parece favorável à reparação da vítima e ao adequado atendimento de seus direitos fundamentais, uma vez que **a centralidade no conceito de culpa é insuficiente** para lidar com danos causados por entes compostos por sistemas de IA.[102] Afinal, o **alto grau de autonomia, imprevisibilidade e aprendizagem desses sistemas** dificulta a identificação da fronteira entre danos que resultam de erro humano direto e aqueles que são desencadeados pela atividade regular do algoritmo, que também devem ser objeto de responsabilização, independente de comprovação de culpa de quem estiver em controle do sistema.

O comportamento imprevisível derivado de sistemas baseados em *deep learning,* por exemplo, cujo funcionamento é extremamente complexo de ser compreendido por seus próprios desenvolvedores,[103] torna difícil, ou quase impossível, para a vítima comprovar se eventual dano adveio de uma conduta negligente do ser humano.

---

101 MORAES, Maria Celina Bodin de. A constitucionalização do direito civil e seus efeitos sobre a responsabilidade civil (2006). **Direito, Estado e Sociedade**. v.9 (29) , p. 233 - 258, p. 239.

102 BARBOSA, Mafalda Miranda. Responsabilidade Civil pelos danos causados por entes dotados de inteligência artificial. In: **Direito Digital e Inteligência Artificial**, Editora Foco, 2021, p. 160.

103 MALGIERI, Gianclaudio. **"Just" Algorithms: Justification (beyond explanation) of automated decisions under the GDPR.** Law and Business, vol 2021, issue 1 1, p. 6. Disponível em: https://www.gianclaudiomalgieri.eu/2020/12/14/just-algorithms/. Acesso em: 09 mai 2022.

Assim como há muitos casos em que pode existir culpa como, por exemplo, as hipóteses de danos causados pela não realização das atualizações de *software,* ou situações de quebra de deveres de cuidado que permitem que terceiros interfiram no sistema, também **existem situações em que os danos sejam oriundos de atividade autônoma e inesperada do algoritmo, fruto de sua aprendizagem.**[104] Isto é, situações em que a máquina assume tamanha autonomia e conduz posturas que não eram exatamente esperadas pelo desenvolvedor do algoritmo.

Um exemplo clássico é o do **robô *Gaak***, desenvolvido na Inglaterra, no *Magna Science Center*. Tratou-se de experimento que atribuía aos robôs os papéis de "caçador" e "presa", colocando-os em uma arena apenas para que, respectivamente, caçassem e fugissem. O experimento visava verificar a aplicabilidade do princípio da sobrevivência do mais apto aos robôs dotados de inteligência artificial, bem como verificar se eles poderiam se beneficiar do conhecimento adquirido. O robô *Gaak,* no entanto, fora deixado sem vigilância por 15 minutos, conseguindo fugir da arena, atravessar o muro da sede e encontrar uma saída, sendo posteriormente atingido por um carro no estacionamento. Tudo isso sem que tivesse sido programado para tanto.[105]

Também a título exemplificativo, se um detector de fumaça em um ambiente doméstico inteligente não dispara um alarme em razão de algum erro de *software*, a comprovação desse defeito não é simples para o usuário, principalmente porque requer uma análise cuidadosa do código e de sua adequação para os componentes de *hardware*.[106]

Ademais, na sociedade contemporânea, as possibilidades de prejuízos vão além dessa hipótese abstrata. No contexto de políticas públicas que utilizam reconhecimento facial para identificação de suspeitos, tem sido debatido profundamente acerca de conflitos entre a noção de justiça, autonomia humana, privacidade e os deveres estatais de proteção e segurança, máxime tendo em vista que sua atividade tem apresentado elevado índice de erro e acarretado detenções indevidas[107], [108], exacerbando, ainda, o racismo

---

104 BARBOSA, Mafalda Miranda. Responsabilidade Civil pelos danos causados por entes dotados de inteligência artificial. Op. cit. p. 160.

105 THE GUARDIAN. **Robot fails to find a place in the sun.** Disponível em: https://www.theguardian.com/uk/2002/jun/20/engineering.highereducation. Acesso em: 09 mai 2022.

106 THE EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES - NEW TECHNOLOGIES FORMATION, EUROPEAN COMISSION. **Liability for artificial intelligence and other emerging digital technologies.** Disponível em: https://op.europa.eu/en/publication-detail/-/publication/1c5e30be-1197-11ea-8c1f-01aa75ed71a1/language-en/format-PDF. Acesso em: 09 mai 2022.

107 UOL. **Técnicas de vigilância como identificação fácil ainda são falhas** (2019)**.** Disponível em: https://www.uol.com.br/tilt/noticias/redacao/2019/05/27/tecnicas-de-vigilancia-como-identificacao-facial-ainda-sao-falhas.htm. Acesso em: 09 mai 2022.

108 FOLHA DE SÃO PAULO. **Nos erros de reconhecimento facial, um "caso isolado" atrás do outro.** Disponível em: https://piaui.folha.uol.com.br/nos-erros-de-reconhecimento-facial-um-caso-isolado-atras-do-outro/. Acesso em: 09 mai 2022.

estrutural existente no país.[109] Acidentes com veículos autônomos também não são incomuns[110] e já foram constatadas situações de morte de pacientes após falha de inteligência artificial em cirurgias.[111]

No campo patrimonial, um grupo que compra e vende ações para promover liquidez no mercado, suportou um prejuízo de milhões de dólares após a eclosão de um erro operacional em um *software* de negociações de valores mobiliários.[112] Também cabe mencionar a atuação do robô da *Microsoft* chamada *Tay*, que, em menos de 24 horas de interação, passou a proferir termos racistas no *Twitter*.[113]

A sociedade de risco acentua a mitigação da culpa como único requisito de responsabilização civil e incorpora papel primordial na estruturação de bases teóricas que elegem vias alternativas de parametrização.[114] O regime de responsabilidade baseado exclusivamente na culpa passou por transformações no decorrer do século XX em razão de sua clara insuficiência.[115] Mitiga-se o paradigma exclusivo da culpa, que designava a ideia de reprovação moral da conduta,[116] dando espaço para a reparação com fulcro no risco da atividade.

No mesmo sentido, **a verificação do defeito de desenvolvimento do sistema nem sempre é fácil**. Isso se dá principalmente considerando a existência de múltiplas cadeias de produção e, ainda, a constatação de que a máquina poderá naturalmente apresentar resultados incomuns ou, ainda, comportar-se conforme o esperado e, mesmo assim, desencadear danos.[117]

---

109 G1. **Sistema de reconhecimento facial da PM do RJ falha e mulher é detida por engano**. Disponível em: https://g1.globo.com/rj/rio-de-janeiro/noticia/2019/07/11/sistema-de-reconhecimento-facial-da-pm-do-rj-falha-e-mulher-e-detida-por-engano.ghtml; Acesso em: 06 mai 2022. http://www.startse.com/noticia/ecossistema/reconhecimento-facial-policia-londres. Acesso em: 09 mai. 2022.

110 BBC. **Tesla: acidente com carro "sem motorista" mata 2 pessoas nos EUA.** Disponível em: https://www.bbc.com/portuguese/internacional-56806154. Acesso em: 09 mai 2022.

111 EXTRA. **Paciente morre após erro de robô em cirurgia cardíaca.** Disponível em: https://extra.globo.com/noticias/page-not-found/paciente-morre-apos-erro-de-robo-durante-cirurgia-cardiaca-23216846.html. Acesso em: 09 mai 2022.

112 THE NEW YORK TIMES. **Knight Capital Says Trading Glitch Cost It $440 Million** (2012)**.** Disponível em: https://dealbook.nytimes.com/2012/08/02/knight-capital-says-trading-mishap-cost-it-440-million/. Acesso em: 09 mai 2022.

113 Maiores informações sobre este caso podem ser encontradas em: REVISTA VEJA. **Exposto à internet, robô da Microsoft vira racista em 1 dia**. Disponível em: https://veja.abril.com.br/tecnologia/exposto-a-internet-robo-da-microsoft-vira-racista-em-1-dia/. Acesso em: 09 mai 2022.

114 PORTO, Uly de Carvalho Rocha. **A responsabilidade civil extracontratual por danos causados por robôs autônomos.** 2018**.** Dissertação (Mestrado em Ciências Jurídico-Civilistas), Faculdade de Direito da Universidade de Coimbra, Coimbra, p. 35.

115 BUSATTA, Eduardo Luiz. Proteção de dados pessoais e antijuridicidade. **Direito Civil e Tecnologia: Tomo II.** Ehrhardt Júnior, Marcos; CATALAN, Marcos; MALHEIROS, Pablo. Belo Horizonte: Fórum, 2021, p. 301.

116 PEREIRA, Alexandre Pimenta Batista. Os confins da responsabilidade objetiva nos horizontes da sociologia do risco. **Revista de Informação Legislativa.** v. 43, n. 170, p. 181-189, abr./jun. 2006, p. 4.

117 MAGRANI, Eduardo; SILVA, Priscilla; VIOLLA, Rafael. Novas perspectivas sobre ética e responsabilidade de inteligência artificial. In: FRAZÃO, Ana. MULHOLLAND, Caitlin (coord.). **Inteligência artificial e direito: ética, regulação e responsabilidade**. São Paulo: Thomson Reuters Brasil, 2019, p. 134.

É imprescindível destacar, ainda, que **a responsabilidade civil, dentro do panorama da inteligência artificial**, se desenvolve a partir do paradigma da solidariedade social, da reparação integral do dano e da cláusula geral de tutela da pessoa humana. Isso afasta a ideia de que os prejuízos oriundos dessa atividade são fatos não indenizáveis.

Nesse sentido, **não parece adequado consagrar, de antemão, que a responsabilidade será sempre subjetiva ou objetiva**, uma vez que o caso concreto pode apontar soluções diversas, a depender se trata de uma relação civil, consumerista, empresarial ou trabalhista, além dos aspectos inerentes ao tipo de inteligência artificial e conduta do desenvolvedor. É necessário pensar, assim, em um **sistema múltiplo de responsabilidades, que considere a tipologia e a autonomia da IA, bem como os sujeitos envolvidos e a natureza da relação jurídica posta em apreciação.**

O **parágrafo terceiro do mesmo artigo** determina, ainda, que quando a utilização do sistema de IA envolver relações de consumo, o agente responderá independentemente de culpa pela reparação dos danos causados, no limite de sua participação efetiva no evento danoso.

Apesar disso, dado que o regime da responsabilidade objetiva nas relações consumeristas já está consagrado no Código de Defesa do Consumidor (CDC), o PL, ao repetir tal disposição, acaba por acrescentar a limitação da responsabilidade do fornecedor às fronteiras de sua participação efetiva no evento danoso. Além disso, é importante evidenciar que o dispositivo não mitiga os problemas advindos da previsão da responsabilidade subjetiva, que, nos termos do inciso VI, subsistiria para os outros casos, motivo pelo qual é **integralmente problemática a proposta de responsabilidade veiculada no projeto de lei**.

Ainda no campo do CDC, é importante mencionar que a noção de defeito como pressuposto de responsabilização parece insuficiente, pois os sistemas de inteligência artificial podem apresentar resultados inesperados. Isto é, resultados adversos oriundos do comportamento imprevisível da máquina que, como se sabe, é um fenômeno comum nesse tipo de programação.

É por isso que se tem defendido que a lógica tradicional da responsabilidade pelo fato do produto não daria mais conta de explicar danos causados por uma ação autônoma da máquina, pois esse agir muitas vezes não seria propriamente um defeito imputável aos seus fabricantes, mas uma decorrência da autonomia crescente da inteligência artificial.[118]

Da mesma forma, o **parágrafo quarto** se limita a reproduzir o exato teor do art. 37, § 6º da Constituição Federal, que assegura a responsabilidade objetiva do Estado e, portanto, também não mitiga os problemas oriundos da previsão de responsabilidade subjetiva para as demais hipóteses. Trata-se, portanto, de um dispositivo com pouca força normativa.

Vale ressaltar que **o argumento de que a responsabilidade objetiva seria um fator de inibição à inovação tecnológica já vem sendo desconstruído**. O desenvolvimento histórico tem demonstrado que os modelos de culpa presumida ou de responsabilidade objetiva não limitaram a proliferação de novas tecnologias. Ao contrário, eles têm assegurado o pleno desenvolvimento industrial ao passo em que os custos de tais modelos foram incorporados pelo mercado sem prejuízo do ressarcimento das vítimas de danos injustos, **implementando-se um modelo solidarista de responsabilidade.**[119]

No mesmo sentido, tampouco a adoção da responsabilidade objetiva implica no esvaziamento dos deveres de cuidado a serem tomados pelos responsáveis. Isso porque a tendência contemporânea de responsabilização se materializa na **constatação da proatividade, em que se torna necessário ir além do mero cumprimento da lei, demonstrando, também, a tomada de medidas proativas para prevenção do dano**.[120] Essa constatação parte do pressuposto da insuficiência de um modelo de responsabilidade meramente dicotômico que se divide entre responsabilidade objetiva e subjetiva, dando espaço para o desenvolvimento de teorias de proatividade.

---

118 MEDON, Filipe. **Inteligência Artificial e Responsabilidade Civil: Autonomia, Riscos e Solidariedade**. Salvador: Editora JusPodivm, 2020, p. 348.

119 MORAES, Maria Celina Bodin de. LGPD: um novo regime de responsabilização dito "proativo". **Civilistica**. A. 8, n. 3, 2019. Disponível em: http://civilistica.com/lgpd-um-novo-regime-de-responsabilizacao-civil-dito-proativo/. Acesso em: 09 mai 2022.

120 *Ibidem*.

Ainda nessa temática, a **Resolução do Parlamento Europeu de 3 de maio de 2022** sobre a inteligência artificial na era digital salientou desafios consideráveis na eficácia das disposições do quadro de responsabilidade devido às características dos sistemas de IA, como **(i)** a sua complexidade, **(ii)** conectividade, **(iii)** opacidade, **(iv)** vulnerabilidade, **(v)** possibilidade de sofrer alterações por meio de atualizações, **(vi)** capacidade de autoaprendizagem e potencial autonomia, e a **(vii)** multiplicidade de intervenientes envolvidos na sua criação, implantação e utilização. Considera, por conseguinte, que, embora **não haja necessidade de proceder a uma revisão completa dos regimes de responsabilidade funcionais**, é necessário proceder a ajustamentos específicos e coordenados dos regimes de responsabilidade para evitar que as pessoas que sofrem danos acabem não sendo indenizadas.[121]

É possível readaptar as categorias clássicas da responsabilidade civil, acomodando-as aos cânones constitucionais e aos contextos sociais contemporâneos, não sendo necessária ou pertinente a referida alteração legislativa no contexto atual, especialmente tendo em vista a necessidade de maturação do debate sobre regulação de IA.[122]

Sendo assim, recomenda-se a **exclusão do inciso IV e parágrafos 3° e 4° do PL 21/2020**, de modo que as normas sobre responsabilidade por danos oriundos da atividade de IA **levem em consideração a tipologia da IA, o risco gerado e o grau de autonomia em relação ao ser humano, além da natureza dos agentes envolvidos**, a fim de se determinar, no caso concreto, qual o regime de responsabilidade civil será aplicável.[123]

---

121 PARLAMENTO EUROPEU. **Resolução do Parlamento Europeu, de 3 de maio de 2022, sobre a inteligência artificial na era digital.** Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_PT.html. Acesso em: 11 mai 2022.

122 TEPEDINO, Gustavo. SILVA, Rodrigo da Guia. Desafios da inteligência artificial em matéria de responsabilidade civil. **Revista Brasileira de Direito Civil.** Belo Horizonte, V. 21, p. 61-86, jul/set, 2019, p. 11.

123 CONJUR. **Especialistas criticam a responsabilidade subjetiva prevista no PL do Marco da IA.** Disponível em: https://www.conjur.com.br/2021-out-27/especialistas-questionam-artigo-pl-marco-legal-ia. Acesso em: 09 mai 2022.

VIII. SUGESTÕES DE REDAÇÃO

Conscientes de que o debate a respeito da regulação da IA ainda pode avançar no Legislativo brasileiro e que outros elementos que ainda não identificamos podem ser incorporados a essa redação, propomos as seguintes sugestões:

DA TRANSPARÊNCIA[124] [125]

Art. 1° Decisões tomadas exclusivamente por, ou com a assistência de um sistema de inteligência artificial que produza efeitos legais em relação a um indivíduo ou grupo afetado, ou que de alguma forma afete seus interesses, deve ser acompanhada por uma explicação significativa e relevante de pelo menos:

I - qual foi o papel do sistema de inteligência artificial no processo de tomada de decisão e a medida em que o resultado produzido por esse sistema influenciou a decisão neste caso;
II - critérios, principais parâmetros utilizados e seu peso relativo, bem como a lógica geral de funcionamento do sistema;
III - quais informações relativas ao sujeito ou grupo afetado foram utilizados pelo sistema de inteligência artificial na tomada de decisão, incluindo a indicação de seus dados pessoais, e cada um dos parâmetros com base nos quais a decisão foi tomada;
I**V -** se aplicável, a categoria ou grupo no qual o sujeito a inteligência artificial foi classificado;
V - se a mesma decisão foi tomada em relação a outras pessoas em circunstâncias similares e, se não, uma explicação por que o sujeito foi tratado de maneira diferente, sem prejuízo da proteção de dados pessoais;
VI - em quais circunstâncias a decisão tomada poderia ter sido diferente;
VII - informações sobre os direitos da pessoa previstos nesta lei, incluindo o direito de apresentar uma reclamação a uma autoridade supervisora.

---

124 Uma potencial e preliminar redação para abranger o ponto **(iii) na seção relativa à Transparência** abordada nesta contribuição pode ser inspirada nas propostas de emenda ao *AI Act* europeu sob números 1149, 1151, 1152 e 1153. PARLAMENTO EUROPEU. **Amendments 774 - 1189**. Draft Report. Harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts. Bruxelas, 13 jun. 2022. Disponíveis em: https://www.europarl.europa.eu/doceo/document/CJ40-AM-732837_EN.pdf. Acesso em: 26 jun. 2022.

125 Ponto (iii): Explicações dirigidas a pessoas ou grupos sobre os quais possam repercutir efeitos.

§1º. A explicação prevista no caput deve ser fornecida por padrão ao mesmo tempo em que a decisão é comunicada ao sujeito afetado pelo sistema de inteligência artificial, devendo ser feita de forma clara, facilmente compreensível e acessível a pessoas com deficiências.

§2º. Para que as informações especificadas no inciso III sejam significativas para seu recipiente, elas devem incluir uma descrição facilmente compreensível das inferências retiradas de outros dados.

DA AVALIAÇÃO DE IMPACTO DE INTELIGÊNCIA ARTIFICIAL

Da realização da Avaliação de Impacto de Inteligência Artificial

Art. 1° Quando um determinado uso de sistema de inteligência artificial for suscetível de provocar um alto risco para o exercício de direitos e liberdades fundamentais, o desenvolvedor ou aplicador da tecnologia deverá realizar avaliação de impacto de inteligência artificial.
Parágrafo único. As avaliações de impacto de inteligência artificial deverão ser realizadas:
I - na fase de design do sistema de inteligência artificial;
II - após finalização do desenvolvimento do sistema de inteligência artificial;
III - após uma modificação substancial do sistema de inteligência artificial;
IV - antes da aquisição/utilização do sistema de inteligência artificial;
V - de forma periódica, em intervalos não superiores a um ano, após o início de sua utilização por seu aplicador.

Dos casos obrigatórios de Avaliação de Impacto de Inteligência Artificial

Art. 2° A realização de uma avaliação de impacto de inteligência artificial será obrigatória quando o uso de um sistema de inteligência artificial atender cumulativamente a pelo menos uma condição geral e um risco específico:

I - condição geral:

a) uso do sistema para avaliação sistemática de pessoas físicas com base em tratamento automatizado, incluindo a definição de perfis, em que decisões adotadas a partir dessa avaliação produzam efeitos jurídicos relativamente a indivíduos ou grupos ou que os afete significativamente de forma similar;

b) ocorrência de operações de tratamento em larga escala de dados pessoais sensíveis a que se refere o art. 5°, II, da Lei 13.709/2018, ou de dados pessoais relacionados a infrações e condenações penais, observadas as leis e normas específicas;

c) controle sistemático de zonas acessíveis ao público.

II - risco específico:

a) interferência nos direitos fundamentais, como os direitos à igualdade e à não discriminação, à privacidade, e à liberdade de expressão;

b) potenciais danos à saúde ou segurança do sujeito, como perda da vida ou danos corporais;

c) potenciais danos psicológicos, como a autocensura, a perda de autoestima e a perda de autonomia pessoal;

d) potenciais danos sociais ou econômicos, como danos financeiros, perda de propriedade, manipulação econômica ou restrição a serviços públicos ou privados;

e) potenciais danos reputacionais ou de estigmatização;

f) potenciais preconceitos ou discriminações injustas, incluindo discriminação de preços, discriminação no mercado de trabalho ou acesso diferenciado ou discriminatório a serviços;

g) potenciais danos coletivos, como perda de liberdade ou instabilidade econômica ou política.

§ 1°. As autoridades reguladoras competentes poderão indicar novos riscos específicos a liberdades e direitos fundamentais que ensejam a necessidade de realização de uma avaliação de impacto de inteligência artificial.

§ 2°. É recomendado que autoridades reguladoras competentes elaborem e tornem pública uma lista dos usos de sistemas de inteligência artificial aos quais não é obrigatória uma avaliação de impacto de inteligência artificial. A referida lista poderá ser alterada periodicamente a depender de nova compreensão sobre os riscos dos sistemas em questão e a necessidade de se conduzir uma avaliação de impacto.

Requisitos mínimos para uma Avaliação de Impacto de Inteligência Artificial

Art. 3° A avaliação de impacto deve incluir, pelo menos:
I - descrição detalhada do sistema de inteligência artificial, contendo o contexto de utilização, a respectiva finalidade, a lógica de operação da tecnologia, o tipo de dado utilizado e como será realizado o respectivo treinamento, os sujeitos ou grupos afetados de forma direta, indireta, intencional e não intencional, os benefícios da aplicação e quais leis o sistema deve respeitar;
II - avaliação da legalidade da finalidade do sistema de inteligência artificial;
III - observações de terceiros interessados e/ou afetados pelo sistema;
IV - identificação e avaliação dos riscos para os direitos e liberdades das pessoas afetadas;
V - definição e indicação das salvaguardas, medidas de segurança e medidas de proteção dos direitos e liberdades fundamentais dos sujeitos e grupos afetados;
VI - avaliação da proporcionalidade entre a finalidade e resultado, identificando se não há um meio alternativo, razoável e menos invasivo de alcançar o mesmo objetivo.

Art. 4° Identificados riscos residuais na fase de desenvolvimento da tecnologia ou em momento posterior capazes de resultar um alto risco para as liberdades e direitos fundamentais, e não sendo possível mitigar os seus efeitos adversos, o desenvolvedor ou o aplicador da tecnologia deverá deverá suspendê-la até que apresente resultados, atestados pela autoridade reguladora competente.

Art. 5° O desenvolvedor ou o aplicador do sistema de inteligência artificial deverá disponibilizar a avaliação de impacto de inteligência artificial, em repositório digital, de acesso público, a ser gerido pela autoridade reguladora competente.

§1°. Informações contidas na avaliação de impacto de inteligência artificial que consistam em segredo comercial ou industrial poderão ser classificadas como confidenciais e ter seu acesso restrito ao público.

§2°. As avaliações de impacto de inteligência artificial deverão ser divulgadas integralmente quando dizerem respeito a sistema desenvolvido ou utilizado por órgão ou entidade da Administração Pública, observadas as limitações e salvaguardas relativas à proteção de dados pessoais, nos termos da Lei nº 13.709/2018.

IX. BIBLIOGRAFIA

ACCESS NOW. **Carta aberta para banimento global de usos de reconhecimento facial e outros reconhecimentos biométricos remotos que permitam vigilância em massa, discriminatória e enviesada.** 2021. Disponível em: https://www.accessnow.org/cms/assets/uploads/2021/06/BanBS-Portuguese.pdf.

ACCESS NOW. **Human rights in the age of human intelligence**. Access Now, 2018. Relatório técnico. Disponível em: https://www.accessnow.org/cms/assets/uploads/2018/11/AI-and-Human-Rights.pdf.

ANANNY, Mike; CRAWFORD, Kate. **Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability.** New Media & Society, v. 20(3), pp. 973–989, 2016. Disponível em: https://journals.sagepub.com/doi/10.1177/1461444816676645.

ANDRADE, Norberto Nuno Gomes de; KONTSCHIEDER, Verena. **AI impact assessment: a policy prototyping experimente.** Open Loop, 2021. Disponível em: https://openloop.org/wp-content/uploads/2021/01/AI_Impact_Assessment_A_Policy_Prototyping_Experiment.pdf

ANGWIN, J. et al. **Machine Bias.** ProPublica, 2016. Disponível: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

ARAGÃO, Alexandra. **Aplicação nacional do princípio da precaução**. Associação dos Magistrados da Jurisdição Administrativa e Fiscal de Portugal , Lisboa, v. 1, n. 1, jan./2013.

ARYA, Vijay *et al.* **One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques** (2019). Arvix. Disponível em: https://arxiv.org/abs/1909.03012.

ASGHARI, Hadi et al. **What to explain when explaining is difficult?**: an interdisciplinary primer on XAI and meaningful information in automated decision-making. Berlin: Alexander von Humboldt Institute For Internet And Society, 2021. Disponível em: https://www.hiig.de/publication/what-to-explain-when-explaining-is-difficult-an-interdisciplinary-primer-on-xai-and-meaningful-information-in-automated-decision-making.

BALDWIN et al. **Understanding Regulation: Theory, Strategy, and Practice.** 2. ed: Oxford University Press, 2013.

BARBOSA, Mafalda Miranda. Responsabilidade Civil pelos danos causados por entes dotados de inteligência artificial. In: **Direito Digital e Inteligência Artificial**, Editora Foco, 2021, p. 160.

BBC. **Tesla: acidente com carro "sem motorista" mata 2 pessoas nos EUA.** Disponível em: https://www.bbc.com/portuguese/internacional-56806154.

BERGUER FILHO, Aírton Guilherme. **Regulação e Governança dos Riscos das Nanotecnologias**. 1. ed. Belo Horinzonte: Arraes, 2018.

BONINI, Paulo Rogério. **Responsabilidade civil por ato lícito.** Disponível em: https://www.tjsp.jus.br/download/EPM/Publicacoes/ObrasJuridicas/rc6.pdf?d=636680468024086265.

BRYNJOLFSSON, Eric; MCAFEE, Andrew. The business of Artificial Intelligence. **Harvard Business Review**. Edição especial Artificial Intelligence, Boston, jul 2017. Disponível em: https://hbr.org/2017/07/the-business-of-artificial-intelligence.

BUCHER, Taina. **If... then**: **algorithmic power and politics.** Oxford University Press, 2019.

BUOLAMWINI, Joy; GEBRU, Timnit. Gender Shades: Intersectional Accuracy Disparities in Commencial Gender Classification. **Proceedings of Machine Learning Research**. 2018, Nova York, v. 8. Disponível em: http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf.

BUSATTA, Eduardo Luiz. Proteção de dados pessoais e antijuridicidade. **Direito Civil e Tecnologia: Tomo II.** Ehrhardt Júnior, Marcos; CATALAN, Marcos; MALHEIROS, Pablo. Belo Horizonte: Fórum, 2021.

CANADÁ. **Algorithmic Impact Assessment Tool**. Disponível em: https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/algorithmic-impact-assessment.html#toc3-1.

CENTRE FOR INFORMATION POLICY LEADERSHIP (CIPL). **Recommendations on Adopting a Risk-Based Approach to Regulating Artificial Intelligence in the EU**. CIPL, 2021. Disponível em: https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_risk-based_approach_to_regulating_ai__22_march_2021_.pdf.

CERKA, Paulius; GRIGIENE, Jurgita; SIRBIKYTÈ, Gintarè. Liability for damages caused by artificial intelligence. **Computer Law and Security Review.** United Kingdom, v. 31, p. 376-389, 2015.

COMISSÃO EUROPEIA. **Proposta de regulamento do Parlamento Europeu e do Conselho que estabelece regras harmonizadas em matéria de Inteligência Artificial (Regulamento Inteligência Artificial) e altera determinados atos legislativos da União** (2021). Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:52021PC0206.

COMISSÃO EUROPEIA. **Livro Branco sobre a Inteligência Artificial: Uma abordagem europeia virada para a excelência e a confiança** (2020). Disponível em: https://op.europa.eu/pt/publication-detail/-/publication/ac957f13-53c6-11ea-aece-01aa75ed71a1.

CONJUR. **Especialistas criticam a responsabilidade subjetiva prevista no PL do Marco da IA.** Disponível em: https://www.conjur.com.br/2021-out-27/especialistas-questionam-artigo-pl-marco-legal-ia.

CRAWFORD, Kate. **The Atlas of AI.** New Haven And London: Yale University Press, 2021.

DATA&SOCIETY. **Assembling Accountability: Algorithmic Impact Assesment for the Public Interest.** Disponível em: https://datasociety.net/wp-content/uploads/2021/06/Assembling-Accountability.pdf.

DOSHI-VELEZ, Finale & KORTZ, Mason. **Accountability of AI Under the Law**: Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper, 2017. Disponível em: https://dash.harvard.edu/bitstream/handle/1/34372584/2017-11_aiexplainability-1.pdf

EBERS, Martin. **Chapter 2: Regulating AI and Robotics: Ethical and Legal Challenges** (April 17, 2019). Martin Ebers and Susana Navas Navarro (eds.), Algorithms and Law, Cambridge, Cambridge University Press, 2019.

ECP | PLATFORM FOR THE INFORMATION SOCIETY. **Artificial Intelligence Impact Assessment** (2018). Disponível em: https://ecp.nl/wp-content/uploads/2019/01/Artificial-Intelligence-Impact-Assessment-English.pdf.

EDWARDS, Lilian & VEALE, Michael. Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For (2017). **Duke Law & Technology Review**. Vol. 16, n. 1, pp. 1-65. Disponível em: https://ssrn.com/abstract=2972855.

ESCRITÓRIO DO ALTO COMISSARIADO DAS NAÇÕES UNIDAS PARA OS DIREITOS HUMANOS. **Princípios Orientadores sobre Empresas e Direitos Humanos: Implementando os parâmetros 'Proteger, Respeitar e Reparar' das Nações Unidas** (2011). Disponível em: https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf.

EXTRA. **Paciente morre após erro de robô em cirurgia cardíaca.** Disponível em: https://extra.globo.com/noticias/page-not-found/paciente-morre-apos-erro-de-robo-durante-cirurgia-cardiaca-23216846.html.

FERRARI; BECKER; WOLKART, E. N. **Arbitrium ex machina:: panorama, riscos e a necessidade de regulação das decisões informadas por algoritmos**. Revista dos Tribunais: subtítulo da revista, São Paulo, v. 995, n. 1, p. 8, set./2018. Disponível em: http://governance40.com/wp-content/uploads/2018/11/ARBITRIUM-EX-MACHINA-PANORAMA-RISCOS-E-A-NECESSIDADE.pdf

FOLHA DE SÃO PAULO. **Nos erros de reconhecimento facial, um "caso isolado" atrás do outro.** Disponível em: https://piaui.folha.uol.com.br/nos-erros-de-reconhecimento-facial-um-caso-isolado-atras-do-outro/.

G1. **Sistema de reconhecimento facial da PM do RJ falha e mulher é detida por engano**. Disponível em: https://g1.globo.com/rj/rio-de-janeiro/noticia/2019/07/11/sistema-de-reconhecimento-facial-da-pm-do-rj-falha-e-mulher-e-detida-por-engano.ghtml**.**

GETTING THE FUTURE RIGHT – **Artificial intelligence and fundamental rights. European Union Agency for Fundamental Rights** (2020) Disponível em: https://fra.europa.eu/en/publication/2020/artificial-intelligence-and-fundamental-rights

GRUPO DE TRABALHO DO ARTIGO 29 PARA A PROTEÇÃO DE DADOS**. Orientações relativas à Avaliação de Impacto sobre a Proteção de Dados (AIPD) e que determinam se o tratamento é «suscetível de resultar num elevado risco» para efeitos do Regulamento (UE) 2016/679**, abril de 2017. Disponível em: https://www.cnpd.pt/media/f0ide5i0/aipd_wp248rev-01_pt.pdf.

HAMON, Ronan et al. Bridging the Gap Between AI and Explainability in the GDPR: towards trustworthiness-by-design in automated decision-making (2022). IEEE Computational Intelligence Magazine. Disponível em: https://ieeexplore.ieee.org/document/9679770/.

INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS (IEEE). **Recommended practice for assessing the impact of autonomous and intelligent systems on human well-being.** IEEE SA. IEEE 7010-2020. Disponível em: https://standards.ieee.org/standard/7010-2020.html.

JANSSEN, Heleen. Detecting new approaches for a fundamental rights impact assessment to automated decision-making. **International Data Privacy Law**, Oxford, v. 10, n. 1, p. 76-106, fevereiro de 2020. Disponível em: https://doi.org/10.1093/idpl/ipz028.

KAMINSKI, Margot E. **The Right to Explanation, Explained**. University of Colorado Law Legal Studies Research Paper, n. 18-24, 15 June 2018. Disponível em: https://scholar.law.colorado.edu/articles/1227/.

LARANJEIRA DE PEREIRA, José Renato. **Unveiling Mysteries:** Responsive Regulation as a Means For Enforcing Machine-Learning Transparency**.** Dissertação (Mestrado em Direito) – Faculdade de Direito da Universidade de Brasília, em fase de elaboração.

LARSON, J. *et al.*. How We Analyzed the COMPAS Recidivism Algorithm (2016). **ProPublica**. Disponível em: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.

LATONERO, Mark. **Governing artificial intelligence: upholding human rights & dignity**. Data&Society, 2018. Disponível em: https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf.

LENZA, Pedro. **Direito constitucional esquematizado**. 12ª ed. rev., atual. e amp. São Paulo: Saraiva, 2008.

MAGRANI, Eduardo; SILVA, Priscilla; VIOLLA, Rafael. Novas perspectivas sobre ética e responsabilidade de inteligência artificial. In: FRAZÃO, Ana. MULHOLLAND, Caitlin (coord.). **Inteligência artificial e direito: ética, regulação e responsabilidade**. São Paulo: Thomson Reuters Brasil, 2019.

MALGIERI, Gianclaudio. **"Just" Algorithms: Justification (beyond explanation) of automated decisions under the GDPR.** Law and Business, vol 2021, issue 1 1, p. 6. Disponível em: https://website.sciendo.com/wp-content/uploads/2021/06/Malgieri.pdf.

MEDON, Filipe. **Inteligência Artificial e Responsabilidade Civil: Autonomia, Riscos e Solidariedade**. Salvador: Editora JusPodivm, 2020.

MORAES, Maria Celina Bodin de. A constitucionalização do direito civil e seus efeitos sobre a responsabilidade civil (2006). **Direito, Estado e Sociedade**. v. 9 (29).

MORAES, Maria Celina Bodin de. LGPD: um novo regime de responsabilização dito "proativo". **Civilistica**. A. 8, n. 3, 2019. Disponível em: http://civilistica.com/lgpd-um-novo-regime-de-responsabilizacao-civil-dito-proativo/.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. **AI Risk Management Framework.** Disponível em: https://www.nist.gov/system/files/documents/2022/03/17/AI-RMF-1stdraft.pdf.

OBERMEYER, Z. & MULLAINTHAN, S. Dissecting Racial Bias in an Algorithm that Guides Health Decisions for 70M People. **Proceedings of the Conference on Fairness, Accountability, and Transparency [online].** 2019. Disponível em: https://dl.acm.org/doi/10.1145/3287560.3287593.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OECD). OECD framework for the classification of AI systems. **OECD Digital Economy Papers.** n. 323, Fevereiro, 2022. Disponível em: https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1652284916&id=id&accname=guest&checksum=D1B84F556282520301213485FA41E9CE.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). Alto Comissariado de Direitos Humanos**, O direito à privacidade na era digital**. A/HRC/39/29 (03 de agosto de 2018).Disponível em: https://undocs.org/A/HRC/39/29.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE)**. Recommendation of the Council on Artificial Intelligence** (2019)**.** OECD/LEGAL/0449**.** Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

PARLAMENTO EUROPEU. **Resolução do Parlamento Europeu, de 3 de maio de 2022, sobre a inteligência artificial na era digital.** Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2022-0140_PT.html.

PARSONS, K.GARNETT ;. J. **Multi-Case Review of the Application of the Precautionary Principle in European Union Law and Case Law**. Risk Analysis, Cranfield, v. 1, n. 1, maio, 2016. Disponível em: https://www.researchgate.net/publication/303313626.

PASQUALE, Frank. **The Black Box Society: the secret algorithms that control money and information.** Cambridge, Massachusetts: Harvard University Press, 2015.

PEREIRA, Alexandre Pimenta Batista. **Os confins da responsabilidade objetiva nos horizontes da sociologia do risco.** Revista de Informação Legislativa. v. 43, n. 170, p. 181-189, abr./jun. 2006.

PORTO, Uly de Carvalho Rocha. **A responsabilidade civil extracontratual por danos causados por robôs autônomos.** 2018**.** Dissertação (Mestrado em Ciências Jurídico-Civilistas), Faculdade de Direito da Universidade de Coimbra, Coimbra.

REINO UNIDO. Information Commissioner 's Office (ICO). **Guidelines on Data protection impact assessment**. Versão 1.0.124, maio de 2018. Disponível em: https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments/.

REINO UNIDO. Information Commissioner's Officer (ICO). **Guidance on AI and data protection**. Versão 0.0.22, julho de 2020. Disponível em: https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-dp-themes/guidance-on-ai-and-data-protection-0-0.pdf.

REISMAN, Dillon; SCHULTZ, Jason; CRAWFORD, Kate; WHITTAKER, Meredith**. Algorithmic Impact Assessment in the public sector**. AI Now, 2018. Disponível em: https://ainowinstitute.org/aiareport2018.pdf.

REVISTA VEJA. **Exposto à internet, robô da Microsoft vira racista em 1 dia**. Disponível em:https://veja.abril.com.br/tecnologia/exposto-a-internet-robo-da-microsoft-vira-racista-em-1-dia/.

SUSTEIN, Cass R.. **Para além do princípio da precaução: Beyond the Precautionary Principle**. Revista de Direito Administrativo, Rio de Janeiro, v. 1, n. 1, abril, 2012.

TEPEDINO, Gustavo. SILVA, Rodrigo da Guia. Desafios da inteligência artificial em matéria de responsabilidade civil. **Revista Brasileira de Direito Civil.** Belo Horizonte, V. 21, p. 61-86, jul./set. 2019.

THE EXPERT GROUP ON LIABILITY AND NEW TECHNOLOGIES - NEW TECHNOLOGIES FORMATION, EUROPEAN COMISSION. **Liability for artificial intelligence and other emerging digital technologies.** Disponível em: https://op.europa.eu/en/publication-detail/-/publication/1c5e30be-1197-11ea-8c1f-01aa75ed71a1/language-en/format-PDF.

THE GUARDIAN. Robot fails to find a place in the sun. Disponível em: https://www.theguardian.com/uk/2002/jun/20/engineering.highereducation.

THE NEW YORK TIMES. Knight Capital Says Trading Glitch Cost It $440 Million (2012). Disponível em: https://dealbook.nytimes.com/2012/08/02/knight-capital-says-trading-mishap-cost-it-440-million/.

UNIÃO EUROPÉIA. Grupo de Peritos de Alto Nível sobre a Inteligência Artificial. **Orientações éticas para uma IA de confiança.** Disponível em: https://op.europa.eu/pt/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-pt.

UNESCO. **Recomendação sobre a ética da inteligência artificial.** Disponível em: https://unesdoc.unesco.org/in/documentViewer.xhtml?v=2.1.196&id=p::usmarcdef_0000381137_por&file=/in/rest/annotationSVC/DownloadWatermarkedAttachment/attach_import_9946263c-f568-4674-8cd0-154ab58eba98%3F_%3D381137por.pdf&locale=en&multi=true&ark=/ark:/48223/pf0000381137_por/PDF/381137por.pdf#%5B%7B%22num%22%3A55%2C%22gen%22%3A0%7D%2C%7B%22name%22%3A%22XYZ%22%7D%2C0%2C774%2C0%5D.

UOL. **Técnicas de vigilância como identificação fácil ainda são falhas.** Disponível em: https://www.uol.com.br/tilt/noticias/redacao/2019/05/27/tecnicas-de-vigilancia-como-identificacao-facial-ainda-sao-falhas.htm.

VILLASENOR, John; FOGGO, Virginia. **Algorithms and sentencing: What does due process require?** Disponível em: https://www.brookings.edu/blog/techtank/2019/03/21/algorithms-and-sentencing-what-does-due-process-require/.

VINAYAK, Vrinda. The Human Rights Implications of China's Social Credit System. **Oxford Human Rights Hub**. Disponível em: https://ohrh.law.ox.ac.uk/the-human-rights-implications-of-chinas-social-credit-system/.

WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. **International Data Privacy Law**, Oxford, v. 7, n. 2, p. 76–99, maio de 2017. Disponível em: https://academic.oup.com/idpl/article/7/2/76/3860948.

WACHTER, Sandra; MITTELSTADT, Brent and RUSSEL Chris. Counterfactual Explanations Without Opening The Black Box: Automated Decisions and the GDPR. **Harvard Journal Of Law & Technology**, [s. l], v. 31, n. 2, p. 841-887, Spring 2018. Disponível em: https://jolt.law.harvard.edu/assets/articlePDFs/v31/Counterfactual-Explanations-without-Opening-the-Black-Box-Sandra-Wachter-et-al.pdf.

WILSON, B., HOFFMAN, J. *et al.* **Predictive Inequity in Object Detection**. Arxiv. Disponível em: https://arxiv.org/pdf/1902.11097.pdf.