TEMAS CENTRAIS NA REGULAÇÃO DE IA:

O Local, o regional e o global na busca da interoperabilidade regulatória

Bruno Bioni Marina Garrote Paula Guedes

DataPrivacyBR Research

Ficha técnica

A Associação Data Privacy Brasil de Pesquisa é uma organização da sociedade civil, sem fins lucrativos, que promove a proteção de dados pessoais e outros direitos fundamentais diante da emergência de novas tecnologias, desigualdades sociais e assimetrias de poder. Conta com uma equipe multidisciplinar de diferentes regiões brasileiras que desenvolve pesquisas de interesse público, notas técnicas, textos de análise sobre assuntos emergentes, formações com agentes decisórios e com a sociedade de um modo geral.

A Associação acredita que a proteção de dados pessoais é um dos fundamentos da democracia e que precisa ser vista a partir da perspectiva da justiça social e assimetrias de poder. Assim, trabalha para a promoção de uma cultura de proteção de dados e para que os direitos digitais sejam direitos fundamentais de todas e todos, conduzindo pesquisas abertas ao público, orientadas por um forte compromisso social e com financiamento ético.

Para mais informações sobre a organização, impacto de seus projetos e como pesquisas são apoiadas, visite **www.dataprivacybr.org**.

Diretores

Bruno Bioni, Rafael Zanatta e Mariana Rielli

Coordenação

Carla Rodrigues, Jaqueline Pigatto, Pedro Saliba e Victor Barcellos

Incidência

Vinicius Silva

Pesquisadores

Eduardo Mendonça, Gabriela Vergili, Júlia Mendonça, Horrara Moreira, Louise Karczeski, Marina Meira, Paula Guedes e Nathan Paschoalini

Comunicação

Alicia Lobato, João Paulo Vicente, Rafael Guimarães, Rafael Regatieri e Roberto Junior

Adm e Financeiro

Elisa Bayón e Matheus Arcanjo

Licença

Creative Commons

É livre a utilização, circulação, ampliação e produção de documentos derivados desde que citada a fonte original e para finalidades não comerciais.

Imprensa

Para esclarecimentos sobre o documento e entrevistas, entrar em contato com a Associação pelo e-mail imprensa@dataprivacybr.org

Como citar esse documento

BIONI, Bruno; GARROTE, Marina; GUEDES, Paula. *Temas centrais na Regulação de IA: O local, o regional e o global na busca da interoperabilidade regulatória*. São Paulo: Associação Data Privacy Brasil de Pesquisa, 2023.

Sumário

TEMAS CENTRAIS NA REGULAÇÃO DE IA: O LOCAL, O REGIONAL E O GLOBAL NA BUSCA DA INTERO-PERABILIDADE REGULATÓRIA[1]

Bruno Bioni[2]

Marina Garrote[3;4]

Paula Guedes[5]

1 Esta publicação é resultado do projeto onde canta o sabIA: governança e regulação de inteligência artificial no Brasil com financiamento da Fundação Luminate, Eko, e Heinrich Böll Stiftung. Para mais informações, acesse: https://www.dataprivacybr.org/projeto/onde-canta-o-sabia-governanca-e-regulacao-de-inteligencia-artificial-a-partir-do-brasil/.

2 Doutor em Direito Comercial e Mestre em Direito Civil na Faculdade de Direito da Universidade de São Paulo. Foi study visitor do Departamento de Proteção de Dados Pessoais do European Data Protection Board/EDPB e do Conselho da Europa, pesquisador visitante no Centro de Pesquisa de Direito, Tecnologia e Sociedade da Faculdade de Direito da Universidade de Ottawa. É autor do livro Proteção de Dados Pessoais: a função e os limites do consentimento. É membro da Rede Latino-Americana de Estudos sobre Vigilância, Tecnologia e Sociedade/LAVITS, e também da International Association of Privacy Professionals – IAPP, com Certificação CIPP/E. É diretor fundador do Data Privacy Brasil, um espaço de intersecção entre uma escola de cursos e uma associação de pesquisa na área de privacidade e proteção de dados.

3 Advogada. Mestranda em Direito pela Escola de Direito da New York University. Mestre em Processo Civil pela Universidade de São Paulo. Especialista em Gênero e Sexualidade pelo Centro Latino-Americano em Sexualidade e Direitos Humanos do Instituto de Medicina Social da Universidade Estadual do Rio de Janeiro.

4 Marina Garrote, co-autora desta publicação, se envolveu na estruturação e metodologia da pesquisa como um todo e contribuiu com escrita e pesquisa, durante os meses de maio e junho de 2023, especialmente no eixo 1.

5 Advogada. Doutoranda em Direito e Inteligência Artificial pela Universidade Católica Portuguesa – Centro Regional do Porto e Mestre em Direito Internacional e Europeu pela mesma instituição; especialista em Direito Digital pelo ITS-Rio em parceria com a UERJ. Membro do Núcleo de Pesquisa em Direito e Tecnologia Legalite da PUC-Rio. Pesquisadora da Associação Data Privacy Brasil de Pesquisa.

1. Sumário Executivo e apontamentos metodológicos

O objetivo primário deste position paper é organizar conceitos e referenciais teóricos básicos sobre três temas estruturais de qualquer proposta regulatória sobre inteligência artificial (IA), verificando-se como eles foram cobertos pelas iniciativas legislativas no Brasil. Especialmente o projeto de lei 2338/2023, comparando-o com leis, projetos de regulação e documentos de soft law[6] de países e entidades internacionais.

Não se trata de um trabalho exaustivo e que esgotará todas as questões que surgem em cada um dos temas, mas a finalidade central é informar aos possíveis interessados o estado da arte atual em termos de regulação de IA, especialmente no curso do processo legislativo no Brasil.

Somado a isso, o objetivo secundário deste estudo é mapear o nível de convergência das propostas brasileiras ao de outros países e organismos multilaterais e internacionais. Uma análise qualitativa que captura qual é a racionalidade regulatória de intersecção de um movimento global em governança de IA, mas, ao mesmo tempo, sem perder de perspectiva as nuances do que está acontecendo no Brasil (em especial o PL 2338/2023). Ao fim e ao cabo, (a)o leitor(a) terá coordenadas para avaliar em que medida a discussão brasileira é interoperável[7] e quais suas particularidades na direção contrária de um transplante legal acrítico e até mesmo de colonização[8] frente ao que vem sendo discutido fora do país.

As principais conclusões deste estudo, que podem auxiliar na organização do debate regulatório sobre o tema, são:

6 *Soft law* é entendido como regras de conduta com conteúdo normativo, porém, sem força vinculante formal e, portanto, geram efeitos práticos no comportamento de indivíduos e instituições, a partir de condutas de autorregulação dos atores privados; DA SILVA, Paula Guedes Fernandes. Inteligência Artificial na União Europeia: formas de regular a tecnologia que já nos regula. *In*: MENDES, Gilmar Ferreira; DE MORAIS, Carlos Blanco. Governance da Ordem Jurídica em Transformação. Anais do X Fórum Jurídico de Lisboa, 2022, p. 589. Disponível em: https://www.forumjuridicodelisboa.com/2023-anais; TRUBEK, David M.; COTRELL, Patrick; NANCE, Mark. "Soft Law," "Hard Law," and European Integration: Toward a Theory of Hybridity. Legal Studies Research Paper Series, Winsconsin, n. 1002, p. 1–42, nov. 2005. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=855447.

7 Colin Bennet, em seus estudos sobre o campo da proteção de dados pessoais, fala sobre "convergência regulatória" nesta seara. Em outras palavras, apesar de o Direito ser tradicionalmente reconhecido por variar de país para país, há certos cenários em que é possível observar uma padronização regulatória ao existir regulações nacionais com elementos fundantes semelhantes; BENNETT, Colin J.; RAAB, Charles D., Revisiting the governance of privacy: Contemporary policy instruments in global perspective. Regulation & Governance,Vol. 14, Issue 3, p. 447–464, 2018.

8 A formação da cultura jurídica na América Latina, em especial no Brasil, se deu pela importação, muitas vezes acrítica, de regramentos estrangeiros, principalmente vindos da União Europeia, para nossa realidade. Essa realidade faz com que certas provisões legais regulatórias não sejam pautadas de acordo com as necessidades brasileiras enquanto país de sul global, mas como mera importação de sistemas que voltam-se para interesses e necessidades de uma sociedade diferente da nossa em muitos termos; FERRAZZO, Débora; DUARTE, Francisco Carlos. Colonização jurídica na América Latina. Disponível em: www.publicadireito.com.br/artigos/?cod=f376b8ae6217d18c.

(1) **Como se regular e navegar entre o geral e o setorial:** o contínuo surgimento de novas regulações direcionadas à AI, seja por meio de projetos de lei/regulamento ou de documentos internacionais de atores globais de relevância, revela a tendência global em que não se discute mais se, mas como se regular o uso desta tecnologia. Pela continuidade de produção de externalidades negativas de forma transversal, regulações setoriais não são suficientes. Isso, contudo, não afasta a necessidade de um arranjo de governança que navegue entre o geral e o específico justamente para traduzir normas de governança gerais às particularidades de um determinado contexto. Dito de outra forma, uma lei geral não exclui, mas, muito pelo contrário, abre espaço para que a regulação setorial floresça a partir de fundações comuns a diferentes setores da economia;

(2) **Inovação responsável e resiliente socioeconomicamente:** não se deve buscar qualquer tipo de progresso tecnológico, mas um que seja responsável socioeconomicamente. O trade-off não está entre inovação e a proteção de direitos e liberdades fundamentais, mas sim sobre qual tipo de inovação: se de reforço ou corrosiva ao estado democráctico de direito. Por isso, tem se adjetivado o termo: "inovação responsável". A partir desta premissa, propostas regulatórias – em especial transversais e não apenas setoriais – têm o potencial de catalisar o desenvolvimento tecnológico, econômico e social. Em especial com o surgimento das chamadas IAs fundacionais (e.g., generativas) que terão diversos usos e em variados contextos (downstream applications);

(3) **Alvo regulatório plástico e uma regulação dinâmica e equilibrista (regulação assimétrica com base no risco):** justamente por ser um objeto regulatório que se esparrama por diversos setores e contextos, é impossível ter uma resposta homogênea. Por essa razão, verifica-se como grande ponto comum, dentro de um variado leque de opções, o modelo de regulação assimétrica baseada no risco. A ideia é calibrar o peso da regulação – a intensidade de obrigações, direitos e deveres de um determinado agente regulado – de acordo com o nível do risco em um determinado contexto. Isso faz com que os esforços regulatórios e as obrigações de governança não sejam iguais para todos os casos de uso, mesmo que em um mesmo setor, nem mesmo para todos atores da cadeia de IA. Tal escolha regulatória ganhou destaque com a proposta de regulamento de IA da União Europeia, mas já consta de diferentes outras fontes, vindas da OCDE, UNESCO, Canadá, Conselho da União Europeia e até mesmo dos EUA. Essa abordagem é vista como positiva para

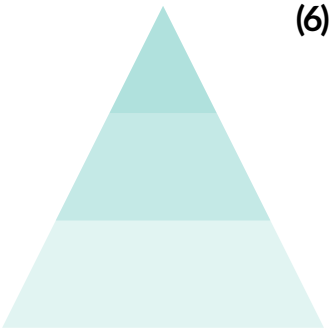
o estímulo à inovação, já que dosa proporcionalmente o grau de intervenção regulatória de acordo com o nível de risco, não criando carga excessiva de obrigações. No cenário brasileiro, o projeto de lei que mais dialoga com a tendência internacional é o PL 2338 de 2023, já que, diferentemente dos demais, busca procedimentalizar minimamente uma classificação dinâmica e equilibrista de acordo com o risco contextual de IA.

(4) **Existem vários modelos de regulação de risco:** a técnica regulatória baseada em risco não é monolítica, mas, muito pelo contrário, tem diversas variações e até mesmo extremos que vão de um monopólio estatal (regulação de comando e controle) ou privado (autorregulação) na tarefa de gerenciamento de riscos. Ainda, é possível ter, inclusive, modelos híbridos como é o caso da grande maioria de propostas de regulação de IA que apostam em um modelo de corregulação em que há alocação de recursos estatais e incentivos para que os próprios agentes econômicos se juntem em uma espécie de parceria público-privada. No entanto, mesmo em tais modelos híbridos, há nuances importantes como de um modelo com maior "supervisão democrática" em que o risco é objeto de maior escrutínio público e controle social. O melhor exemplo disto, e com particular aderência para a cultura jurídico-legal brasileira, é o da seara ambiental. Neste contexto, há diferentes formas nas quais a sociedade civil colabora, a exemplo da participação na formulação e na execução de políticas ambientais, seja mediante a atuação de representantes da sociedade civil em órgãos colegiados dotados de poderes normativos, seja na possibilidade de participação em audiências públicas no âmbito dos estudos de impacto ambiental, ou até mesmo pela participação nos conselhos municipais do meio ambiente. Ao final e ao cabo, a grande tensão de um modelo de regulação baseada em risco consiste justamente em um método de maior ou menor porosidade social cujas distorções notam-se historicamente nos mais diferentes setores regulados. Ao mesmo tempo que uma regulação de IA pode acirrar essas assimetrias e a regulação de risco ser menos democrática e mais tecnocrática, pode, por outro lado, ser paradoxalmente uma janela de oportunidade para fins de equalização e, por conseguinte, maior legitimidade na produção regulatória com mais engajamento social. No Brasil, o PL 2338/23 avança em um modelo de supervisão de risco democrático, mas pode ser aperfeiçoado tendo, por exemplo, um capítulo mais programático a esse respeito ao lado de previsões já existentes – e que podem ser reforçadas – de participação pública na avaliação, classificação e gerenciamento de riscos associados à IA.

(5) Variados degraus da escada do risco: o modelo de regulação de riscos de regulação assimétrica é constituído de faixas, que podem variar de acordo com a metodologia escolhida. No cenário internacional, observa-se a tendência pela definição do que seriam os riscos inaceitáveis e altos, deixando os demais (baixos e médios) à título residual, como acontece no EU AI Act e no Projeto de Lei 15869-19 do Chile. A nomenclatura para cada uma dessas faixas de risco pode variar.

Utilizando-se de metáforas, a gradação de riscos pode ser associada à imagem de uma pirâmide, em que a base constitui os casos de menor risco (sem grandes deveres), o meio representa riscos altos (há imposição de devereres para que a implementação de tecnologia seja permitida) e o topo são os riscos excessivos/muito altos/inaceitáveis. Neste último caso, trata-se de uma intervenção regulatória significativa ao impedir o uso da tecnologia por entender que ela traz mais riscos do que benefícios. Para definição de cada um desses níveis, é importante que sejam estabelecidos elementos qualitativos. Em outras palavras, ao invés de apenas definir os graus de risco (por exemplo, baixo/médio/alto risco) de forma generalista, é indispensável ter critérios mínimos para identificação dos sistemas em cada um desses níveis.

Especificamente no caso brasileiro, o PL 2338/23 é o único que realiza tal divisão ao criar as categorias de risco excessivo e alto, além da categoria residual dos sistemas não classificados pelos dois primeiros níveis. Cada uma dessas categorizações vai desengatilhar obrigações distintas, mais ou menos intensas, o que calibra também os recursos regulatórios.



(6) Risco enquanto elemento dinâmico: além de estabelecer os níveis de risco, é imprescindível que sejam também definidos critérios mínimos para identificação dos sistemas em cada um desses níveis, a partir de elementos qualitativos e quantitativos. Pela experiência internacional, exemplos desses critérios podem ser: contexto, escopo, nível de automação, grau de explicabilidade, potencial de pessoas afetadas, quantidade de dados tratados, entre outros. A definição de critérios dá segurança jurídica para os agentes regulados ao evitar o excesso de generalismo regulatório. Há certo padrão de direcionamento da maior carga regulatória para os sistemas de IA de alto risco que, apesar de não serem proibidos *ex-ante*, devem cumprir com uma gama de obrigações para seu desenvolvimento e

utilização. Geralmente são apresentadas por um rol de casos exemplificativos que são complementados por critérios qualitativos e quantitativos para a atualização destas hipóteses, como ocorre no PL 2338/23 (Brasil), AIDA (Canadá), AI Act (União Europeia) e PL 15869-19 (Chile). Esta previsão de critérios para atualização dos casos de IA de alto risco permite que a legislação se mantenha viva e não fadada ao decurso do tempo.

(7) **A difícil conciliação de uma regulação baseada em risco e em direitos - taxonomia de risco como um dos possíveis indicadores (*proxy*):** assim como AIDA (Canadá), AI Act (União Europeia) e PL 15869-19 (Chile), diretrizes da OCDE e da UNESCO, o PL 2338/23 é explícito ao considerar que uma regulação baseada em risco deve servir como medida de reforço e não de esvaziamento dos direitos de pessoas e grupos que são afetados pela IA. Nesse sentido, o referido projeto de lei brasileiro optou por sistematizar tais direitos e, simetricamente, os deveres correlatos. Com isso, há uma estrutura topográfica normativa que se mostra coerente com tal promessa de harmonização. Principalmente com a previsão de que alguns direitos aplicam-se independentemente do risco do sistema de IA. Um outro indicador dessa possível conciliação é como tais propostas regulatórias articulam a taxonomia de riscos excessivos (inaceitáveis) e alto. Enquanto algumas propostas têm optado, por exemplo, pelo banimento de dados biométricos para fins de persecução penal, outras vão na direção de uma moratória até que tal prática seja regulada. Ainda, há uma significativa variação do rol exemplificativo de IAs de alto risco, bem como dos critérios quantitativos e qualitativos para uma taxonomia dinâmica a esse respeito. Ou seja, o apetite regulatório de intervenção não é o mesmo ao prever de forma ex ante em quais contextos certos direitos e liberdades são inegociáveis, bem como a dilatação das situações nas quais a carga regulatória seria intensa para proteção das pessoas ou dos grupos afetados. Essa imbricação da lógica da classificação de riscos com direitos é um possível indicador em torno da referida conciliação.

(8) **Avaliações de Impacto Algorítmico (AIA) públicas, inclusivas e sobre direitos sociais e não apenas individuais:** para ser uma efetiva ferramenta de *accountability*, diversas propostas regulatórias mapeadas optaram por uma procedimentalização mínima da AIA a partir de um tripé. O primeiro ponto é a publicidade – ao menos uma versão divulgável – para que o resultado das avaliações de gerenciamento do risco seja compartilhado com toda a sociedade, que passa a ser também agente de fiscalização. Inclusive, algumas das

propostas de regulação, como a brasileira, a europeia e a norte-americana, preveem a criação de uma base de dados pública sobre sistemas de IA de alto risco. Um segundo elemento é a participação pública multissetorial significativa de indivíduos e comunidades potencialmente afetados, especialmente dos mais vulneráveis e invisibilizados, o que garante que o processo e o resultado final da avaliação sejam o mais justo e acurado possível para a realidade em que será aplicado. Isso gera maior legitimidade e supervisão democrática de sistemas de IA, caminhando em direção a uma governança multiparticipativa e à co-geração da tecnologia em todo o seu ciclo. Dado o histórico brasileiro consolidado de governança multissetorial na Internet e disposições do PL 2338/23 afirmativas desse direito e dever de co-deliberação sobre os riscos aceitáveis de IA, o Brasil poderia vir a ter experiências regulatórias bem sucedidas no futuro. O terceiro é a variedade dos riscos e benefícios a serem avaliados, já que ainda há uma predominância de efeitos adversos a direitos fundamentais individuais em comparação a direitos sociais. Diferentemente do que prevê as recomendações da UNESCO, o PL 2338/23 não prevê expressamente direitos difusos e coletivos – como ao trabalho e ao meio ambiente – como parte da sua matriz.

(9) Uma regulação atenta aos aspectos sócio-técnicos-econômicos locais: da mesma forma que o Brasil e países do sul global devem traçar estratégias nacionais aderentes aos desafios e oportunidades locais que não são as mesmas do norte-global, é fundamental que a regulação de IA não seja uma importação acrítica de modelos regulatórios de outros contextos. Dos projetos de lei hoje em tramitação no Congresso Nacional, apenas o PL 2338/2023 avança nesse tipo de tropicalização do debate. Em primeiro lugar, o projeto reconhece que o Brasil é permeado por assimetrias e desigualdades estruturais ao incluir, dentre as definições do art. 4º, os conceitos de discriminação direta e indireta da Convenção Interamericana contra o Racismo, adotada pelo Brasil em 2022 com status constitucional. Ainda, traz uma lógica normativa de participação de grupos vulneráveis na avaliação e no gerenciamento dos riscos de IAs que os afetam, bem como de regras específicas para adoção de IA no setor público diante do cenário de que os socioeconomicamente mais vulneráveis serão os mais afetados positivamente ou negativamente. Contudo, o PL é ainda bastante reativo ao não prever, como fizeram outros marcos regulatórios (e.g., MCI e LGPD), um capítulo mais programático e combinado com arranjos institucionais multissetoriais.

(10) IAs Generativas e teste de stress das propostas de regulações de IA: em novembro de 2022, com o lançamento do Chap GPT pela OpenAI, a discussão sobre IAs generativas ganhou destaque. Apesar de uma narrativa alarmista de que sua disrupção impediria a regulação pelas leis e propostas de regulação existentes, a verdade é que essas IAs também precisam e podem ser governadas, com as devidas adequações. Um primeiro desafio enfrentado é sua própria definição, já que há diferentes nomenclaturas utilizadas, como IA generativa, modelos fundacionais, modelos grandes de linguagem, modelos grandes de IA generativa, dentre outros. Porém, em razão de características comuns, é possível equipará-las, uma vez que as considerações acerca de sua regulação são semelhantes.

Um segundo – e talvez o maior – desafio das IAs generativas é que, por se prestarem a diferentes finalidades (nem sempre previsíveis), tensionam o modelo regulatório baseado no risco, atualmente predominante no campo da IA, já que é inerentemente contextual. Em uma tentativa de mitigar tal desafio, é possível a inclusão, dentro do modelo de risco, tanto a ideia de "IA de propósito geral" como a inclusão da análise acerca dos riscos que podem razoavelmente ser esperados, conjuntamente com os conhecidos e previsíveis. Isso faz com que, mesmo não tendo total ciência sobre a existência de alguns riscos, medidas de prevenção sejam tomadas, como já ocorre no atual texto do PL 2338/2023 e na versão do Parlamento Europeu do EU AI Act. Outra solução é o melhor desenvolvimento dos atores envolvidos na cadeia produtiva de sistemas de IA Generativa para que haja destrinchamento das obrigações de cada um deles – como fez a última versão da proposta europeia ao propor uma cooperação entre esses agentes.

2. Contexto de regulação de inteligência artificial no Brasil

O Brasil, seguindo os processos internacionais, inclina-se em direção à governança de IA há alguns anos. Em abril de 2021, foi publicada a Estratégia Brasileira de Inteligência Artificial (EBIA) pelo Ministério da Ciência, Tecnologia e Inovação (MCTI), que estabeleceu nove eixos temáticos centrais para o desenvolvimento de sistemas de IA no Brasil. Dentre eles, destaca-se o eixo de "Legislação, regulação e uso ético", cujo objetivo é o de buscar um equilíbrio entre proteção de direitos fundamentais, desenvolvimento tecnológico e criação de parâmetros legais para estabelecimento de segurança jurídica quanto à responsabilidade dos agentes envolvidos na cadeia de valor de IA[9].

Porém, a EBIA recebeu muitas críticas em razão de sua abordagem genérica e carente de planejamento ao, por exemplo: a) não indicar os atores responsáveis pela governança; b) não aprofundar a análise de métodos aplicáveis para questões críticas (como transparência e explicabilidade); e (iii) não refletir criticamente sobre o uso de IA em contextos altamente arriscados, como segurança pública[10].

Em paralelo à EBIA, desde antes de 2021, foram protocolados diferentes projetos de lei no Congresso Nacional sobre o tema de regulação da IA. Dentre eles, destacou-se a proposição do Projeto de Lei 21/2020 (PL 21/20), apresentado expressamente como um esforço de materializar a faceta da legislação sobre o uso de sistemas de IA no Brasil[11] e cuja urgência e aprovação foram deliberadas na Câmara dos Deputados no mesmo ano.

O texto final do PL 21/20 aprovado pela Câmara dos Deputados em setembro de 2021, sem esgotar todos os mecanismos de participação pública significativa[12], estabelecia fundamentos e princípios gerais para o desenvolvimento e aplicação de IA no Brasil, trazendo apenas algumas diretrizes especificamente para o poder público, de forma a manter o modelo de autorregulação setorial da tecnologia, não prevendo também rol de direitos e deveres. Ademais, o texto aprovado carecia de densidade normativa, de instrumentalização de ferramentas de governança efetivas e de direcionar riscos específicos

9 Ministério da Ciência, Tecnologia e Inovações (MCTI). Estratégia Brasileira de Inteligência Artificial – EBIA. Julho de 2021. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-documento_referencia_4-979_2021.pdf. p. 16.

10 GASPAR; Walter B.; DE MENDONÇA, Yasmin Curzi. A Inteligência Artificial no Brasil ainda precisa de uma estratégia. Relatório do Centro de Tecnologia e Sociedade da FGV Direito Rio. Maio de 2021. Disponível em: https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/30500/EBIA%20pt-br.pdf?sequence=3&isAllowed=y.

11 Associação Data Privacy Brasil de Pesquisa. Nota Técnica – Contribuições do Data Privacy Brasil ao Projeto de Lei nº 21, de 04 de fevereiro de 2020. 2021. Disponível em: https://www.dataprivacybr.org/wp-content/uploads/2021/09/dpbr_notatecnica_pl21.pdf. p. 3.

12 No episódio 78 do Podcast Democracia, Bruno Bioni ressalta que o processo de aprovação do PL 21/20 na Câmara dos Deputados em 2021 poderia ter se utilizado de outras ferramentas de participação pública disponíveis, que não apenas audiências públicas, como é o caso de consultas públicas, o que daria espaço de colaboração para pessoas e grupos que não tiveram voz durante as audiências realizadas, dando maior legitimidade ao processo legislativo. Data Privacy Brasil. Dadocracia – Ep. 78 – Marco Legal da IA. Dadocracia, publicado em nov. 2021. Disponível em: https://open.spotify.com/episode/15BWzRa4cWVRoOjtGGPm4T?si=v7X-iVnWQ3eeIArlGmKaUg.

que o desenvolvimento e utilização de IA no Brasil poderiam desencadear[13].

Após a aprovação da tramitação em regime de urgência, o PL 21/20 foi objeto de críticas, vindas especialmente da comunidade acadêmica e da sociedade civil, o que gerou mobilização social clamando por maior debate e participação pública sobre a referida proposta[14]. O texto, caso aprovado, acabaria por posicionar o Brasil em uma situação de descompasso frente a como o panorama regulatório internacional vinha se desenvolvendo[15].

Nesse contexto, em fevereiro de 2022, o Senador Rodrigo Pacheco, presidente do Senado Federal, instaurou uma Comissão de Juristas responsável por subsidiar a elaboração de um substitutivo sobre inteligência artificial no Brasil (CJSUBIA). A Comissão desenvolveu por 240 dias trabalho intenso na elaboração do substitutivo de lei. Além de seminários internacionais e a realização de audiências públicas com mais de 90 (noventa) pessoas ouvidas, ainda foi aberta consulta pública que possibilitou que qualquer indivíduo e entidade pudesse colaborar com o debate[16].

Em dezembro do mesmo ano, foi publicado o Relatório Final das atividades da CJSUBIA com mais de 900 páginas, o que incluiu, além do histórico de suas atividades e os processos de participação pública externalizados nas contribuições escritas, audiências públicas e seminário internacional, a minuta de substitutivo aos aos Projetos de Leis nºs 5.051/2019, 21/2020 e 872/2021.

Formada por 45 artigos, a nova proposta de texto pretendeu desmistificar o pretenso trade-off existente entre uma regulação que garanta direitos e o desenvolvimento eco-

13 DA SILVA, Paula Guedes Fernandes; GARROTE, Marina Gonçalves. Insuficiência dos princípios éticos para normatização da Inteligência Artificial: o antirracismo e a anti-discriminação como vetores da regulação de IA no Brasil. POLITICS, setembro de 2022. Disponível em: https://politics.org.br/edicoes/insufici%C3%AAncia-dos-princ%C3%ADpios-%C3%A9ticos-para-normatiza%C3%A7%C3%A3o-da-intelig%C3%AAncia-artificial-o.; Data Privacy Brasil. Dadocracia – Ep. 78 – Marco Legal da IA. Dadocracia, publicado em nov. 2021. Disponível em: https://open.spotify.com/episode/15BWzRa4cWVRoOjtGGPm4T?si=v7X-iVnWQ3eeIArlGmKaUg; Data Privacy Brasil. Dadocracia – Ep. 80 – Mais Marco Legal da IA. Dadocracia, publicado em dez. 2021. Disponível em: https://open.spotify.com/episode/0t4Rr07EwIjrdpmvzht79Z?si=0iOyUXc0T5-kH0nr6qhzKA&nd=1; Estadão. "Mais importante lei de recnologia no Brasil não está sendo debatida", diz especialista. Bruno Romani, publicado em 07 dez. 2021. Disponível em: https://www.estadao.com.br/link/cultura-digital/mais-importante-lei-de-tecnologia-no-brasil-nao-esta-sendo-debatida-diz-especialista/. Folha de São Paulo. Brasil apressa lei para inteligência artificial, dizem especialistas. Amanda Lemos, publicado em 18 jul. 2021. Disponível em: https://www1.folha.uol.com.br/mercado/2021/07/brasil-apressa-lei-para-inteligencia-artificial-dizem-especialistas.shtml.

14 Coalizão Direitos na Rede. Inteligência Artificial não pode ser regulada a toque de caixa. Publicado em 23 de setembro de 2021. Disponível me: https://direitosnarede.org.br/2021/09/23/inteligencia-artificial-nao-pode-ser-regulada-a-toque-de-caixa/; Coalizão Direitos na Rede. Brasil não está pronto para regular inteligência artificial. Publicado em 07 de dezembro de 2023. Disponível em: https://direitosnarede.org.br/2021/12/07/brasil-nao-esta-pronto-para-regular-inteligencia-artificial/.

15 Data Privacy Brasil Research. Nota Técnica – Contribuições do Data Privacy Brasil ao Projeto de Lei nº 21, de 04 de fevereiro de 2020. Disponível em: https://www.dataprivacybr.org/wp-content/uploads/2021/09/dpbr_notatecnica_pl21.pdf; Da Silva, Paula Guedes Fernandes; Garrote, Marina. Insuficiência dos princípios éticos para normatização da Inteligência Artificial: o antirracismo e a anti-discriminação como vetores da regulação de IA no Brasil. PoliTICS, set. 2022. Disponível em: https://politics.org.br/edicoes/insufici%C3%AAncia-dos-princ%C3%ADpios-%C3%A9ticos-para-normatiza%C3%A7%C3%A3o-da-intelig%C3%AAncia-artificial-o.

16 Para mais informações sobre as atividades desenvolvidas pela CJSUBIA, acessar: https://legis.senado.leg.br/comissoes/comissao?codcol=2504.

nômico e a inovação, a partir do estabelecimento de uma abordagem baseada em riscos e em direitos por meio de regulação assimétrica, isto é, aumentando a carga obrigacional dos agentes regulados conforme o nível de risco de seu sistema de IA. De acordo com a exposição de motivos:

> "Seu objetivo normativo é conciliar uma abordagem baseada em riscos com uma modelagem regulatória baseada em direitos. Ao mesmo tempo em que se preveem instrumentos de governança para que sejam prestadas contas e seja premiada a boa-fé dos agentes econômicos que gerenciam de forma eficaz os riscos em torno da concepção e implementação de sistemas de inteligência artificial, também há uma forte carga obrigacional para florescimento do escrutínio individual e social a seu respeito.[17]"

Em maio de 2023, o anteprojeto de lei (APL) foi convertido pelo próprio presidente da casa em um novo projeto de lei, numerado 2338/2023. Atualmente, o PL é objeto de análise na Comissão Temporária Interna sobre Inteligência Artificial no Brasil (CTIA), recém instaurada no âmbito do Senado Federal para examinar, no prazo de 120 dias, o referido PL, bem como eventuais novos projetos que tratem da matéria de IA.

Assim, apesar de existirem outros projetos de lei suscetíveis de análise, o presente relatório, quando referir-se ao contexto brasileiro de regulação de IA, se centrará nos projetos que foram mais discutidos publicamente, com grande foco no PL 2338/23 e o PL 21/20.

17 Comissão de Juristas responsável por subsidiar elaboração de substitutivo sobre inteligência artificial no Brasil (CJSUBIA). Relatório Final. Senado Federal, dez. 2022. p. 10 e 11.

3. Escopo e Premissas Metodológicas

Os três principais temas/eixos escolhidos para análise foram:

(i) regulação baseada no risco;
(ii) avaliações de impacto algorítmico; e
(iii) IA Generativa

A escolha por estes temas não é aleatória. Os eixos I e II foram selecionados por serem questões centrais para um adequado equilíbrio entre regulação baseada em riscos e em direitos. O item III representa um dos temas que mais trouxeram dúvidas sobre a abordagem regulatória nos últimos tempos. Por fim, foi adicionado também um item específico para abordar as particularidades nacionais para a regulação de IA à brasileira.

Como dito no tópico 1, o objetivo primordial é mapear as principais discussões dentro dos eixos escolhidos para informar aos leitores a respeito das discussões que ocorrem atualmente e, assim, emitir um diagnóstico do estado da arte em termos de regulação da matéria. Para isso, será feita uma comparação entre as principais iniciativas legislativas no Brasil e no mundo para que o leitor possa visualizar as possíveis escolhas a serem encampadas pelo legislador brasileiro e internacional, possibilitando-lhe uma percepção crítica frente ao leque de opções disponíveis.

Ao se comparar os projetos de leis e outras iniciativas regulatórias, utilizar-se-á bastante de tabelas para destacar as semelhanças e diferenças entre eles. Pretende-se, com isso, dar maior concretude às implicações que as diferentes escolhas em jogo são capazes de produzir, reduzindo-se, em última análise, a assimetria informacional para quem queira se engajar nesse debate legislativo e regulatório sobre IA. Ainda, se notará que majoritariamente as leis comparadas são do norte global ou de organismos internacionais – com exceção da lei chilena e das próprias propostas brasileiras. A escolha foi intencional em razão dessas propostas estarem sendo as mais cobertas pela mídia brasileira e com o objetivo de verificar que há nuances entre elas e entre as propostas brasileiras que são significativas, de sorte a disparar um movimento de colonização normativa.

Ao mesmo tempo, pensa-se que esta pesquisa é apenas um primeiro passo. Posteriormente, objetiva-se colaborar coletivamente com nossos pares da parte majoritária global em pesquisas comparativas nas quais estes sejam protagonistas-sujeitos e não coadjuvante-objetos de análise. Não pretendemos reproduzir um padrão comum em que o sul-global é apenas entrevistado-analisado, mas não são autores da produção intelectu-

al[18].

Para fins metodológicos e para o melhor entendimento do leitor, resta necessário esclarecer que, como o termo "inteligência artificial" é amplo e abarca diferentes tecnologias, como um termo guarda-chuva, nem todos os projetos, leis e documentos comparados terão exatamente o mesmo escopo. Por exemplo, enquanto algumas iniciativas falam em IA, outras abarcam apenas decisões automatizadas, assim como alguns documentos mencionam avaliação de impacto com foco em diferentes aspectos (direitos fundamentais, democracia, Estado de Direito, direitos humanos, entre outros) enquanto outros se referem a outros instrumentos correlatos, como avaliação de risco.

18 Nesse sentido, ver a contribuição conjunta da Southern Alliance para o Global Digital Compact, cujas autoras são todas entidades da América do Sul, África e Índia; Southern Alliance for the Global Digital Compact: contribution for the promotion of digital human rights.2023. Disponível em: https://www.dataprivacybr.org/documentos/southern-alliance-for-the-global-digital-compact/.

DOCUMENTOS E NORMATIVAS QUE SERÃO ANALISADOS

Nacionais			
Material	Abreviação	Explicação	Por que comparar–analisar?
Projeto de Lei 21/2020	PL 21/20	Proposta protocolada na Câmara dos Deputados pelo deputado Eduardo Bismarck em 2020 para estabelecer princípios, direitos e deveres para o uso de inteligência artificial no Brasil. O projeto foi aprovado com urgência na Câmara dos Deputados em setembro de 2021 em forma de substitutivo da deputada Luisa Canziani.	Foi o primeiro PL sobre IA no Brasil que avançou no Congresso, trazendo uma primeira gramática de princípios e objetivos abordagem de soft-law, sem grandes instrumentos de governança.
Projeto de Lei 2338/2023	PL 2338/23	Proposta protocolada no Senado Federal em maio de 2023 pelo senador Rodrigo Pacheco. O texto da proposta é fruto de meses de trabalho de uma Comissão de Juristas estabelecida com o intuito de subsidiar a criação de um substitutivo para projetos de lei sobre IA no Brasil que pendiam análise do Senado, como o PL 21/20.	Fruto de 8 meses (240 dias) de trabalho de uma Comissão de Juristas (CJSUBIA). Nesse período, a Comissão possibilitou diferentes formas de participação: audiências e consulta públicas e seminário internacional, todos com contribuições de especialistas nos temas relacionados à IA. Todos os processos, etapas e estudos realizados pela CJSUBIA foram sintetizados em seu relatório final entregue ao presidente do Senado Federal em dezembro de 2022.
Projeto de Lei 759/2023	PL 759/23	Proposta protocolada na Câmara dos Deputados em fevereiro de 2023 pelo deputado Lebrão com o intuito de regulamentar os sistemas de IA no Brasil e de criar uma obrigação para que o Poder Executivo defina uma Política Nacional de Inteligência Artificial.	Mais um projeto de lei sobre IA no Brasil que serve de exemplo de abordagem generalista e de soft-law, o que pode ser negativo para a governança de sistemas de IA no Brasil.
Projeto de Lei 872/21	PL 872/21	Proposta protocolada em 2021 pelo senador Veneziano Vital do Rêgo com o objetivo de tratar sobre os marcos éticos e as diretrizes para o desenvolvimento e o uso da Inteligência Artificial no Brasil.	Assim como o PL anterior, outro exemplo de projeto de lei generalista que visa regular a IA no Brasil.
Projeto de Lei 5051/19	PL 5051/19	Proposta protocolada em 2019 pelo senador Styvenson Valentim para regulamentar o uso de IA no Brasil.	Da mesma forma, outro projeto de lei cujo objeto central é o de regulamentar o uso de IA no Brasil com abordagem principiológica sem instrumentos de governança efetivos.

Lei Geral de Proteção de Dados (Lei 13.709/2018)	LGPD	Lei brasileira que dispõe sobre o tratamento de dados pessoais por pessoa natural ou por pessoa jurídica de direito público ou privado.	É uma lei transversal, com abordagem baseada em risco e em direitos, que guarda bastante proximidade com algumas da técnicas regulatórias para o cenário de IA.
Anteprojeto de Lei de Proteção de Dados para segurança pública e persecução penal	LGPD Penal	O Anteprojeto de Lei de Proteção de Dados para segurança pública e investigação criminal foi elaborado por uma Comissão de Juristas instituída por Ato do Presidente da Câmara dos Deputados em novembro de 2019. Esse anteprojeto buscou criar um texto de lei para suprir a exceção de aplicação da LGPD, que excepciona sua aplicação para o tratamento de dados para segurança pública e investigação criminal. Atualmente, porém, segue sem sua transformação em projeto de lei no Congresso Nacional.	É um projeto de lei que dispõe de ferramentas regulatórias similares às propostas de IA, tais como a análise de impacto regulatório e o relatório de impacto à proteção de dados pessoais. O anteprojeto de lei é um exemplo de regulação baseada no risco específica para o Estado no contexto de segurança pública e persecução penal. Como trata-se do poder público, há maior cuidado na procedimentalização das ferramentas de avaliações de impacto (regulatório e das aplicações tecnológicas em si) e dos processos de abertura de dados).

Estrangeiras				
Material / Organização	Abreviação	Local	Explicação	Por que comparar-analisar?
Proposta de Regulamento de Inteligência Artificial da União Europeia (EU AI Act Proposal)	EU AI Act	União Europeia	Proposta de regulamento criada pela Comissão Europeia para regular a IA no âmbito da União Europeia. A primeira versão foi publicada em abril de 2021 e atualmente está na última fase de discussão, dependendo da aprovação dos Estados-Membros do Parlamento Europeu (MEPs).	Proposta de Regulação da União Europeia para sistemas de IA. É referência mundialmente em termos de regulação de IA baseada no risco e provavelmente resultará em um novo efeito Bruxelas[19].

19 O Efeito Bruxelas (*Brussels Effect*) se refere à capacidade unilateral da União Europeia de regular os mercados globais ao criar suas regras que elevam os padrões de todo mundo, pois, mesmo não coercitivas para outras regiões do globo, acabam tornando-se referência mundial pelas forças do mercado, uma vez que as empresas multinacionais alargam voluntariamente estas regras para governar suas operações globais. Isso aconteceu no campo da proteção de dados, saúde e segurança dos consumidores, a proteção ambiental, antitruste e discurso de ódio online; BRADFORD, Anu. The Brussels Effect: How the European Union Rules the World. Nova York: Columbia Law School, mar. 2020.

Regulamento Geral de Proteção de Dados da União Europeia	GDPR	União Europeia	Regulamento de 2016 relativo à proteção de pessoas naturais no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados na União Europeia.	Apesar de o escopo ser a proteção de dados pessoais, é um exemplo de regulação que pode servir de inspiração para pensarmos em modelos regulatórios para o cenário de IA também, já que apresenta um modelo baseado no risco e prevê também instrumentos de governança, como a avaliação de impacto de proteção de dados pessoais.
Proyecto de Ley 15869/19	Projeto de lei chileno	Chile	Projeto de lei chileno introduzido na Câmara dos Deputados em 24 de abril de 2023 com o objetivo de regular os sistemas de inteligência artificial, robótica e tecnologias conectadas, em seus diferentes ambientes de aplicação.	O projeto de lei se inspira no EU AI Act, determinando também uma abordagem baseada no risco e um plano de gestão de riscos obrigatório para os sistemas de alto risco, mas traz diferenciações substanciais, como a necessidade de autorização prévia da autoridade competente para que sistemas de IA possam ser aplicados. Por isso,deve ser analisada como outro exemplo de proposta de regulação de IA inspirada no modelo europeu no contexto da América Latina.
NYC Bias Audit Law (Local Law 144)	NY Bias Audit	Estados Unidos (Nova York)	Lei no Estado de Nova York, aprovada em abril de 2023, que exige que uma auditoria de viés seja realizada nas ferramentas automatizadas utilizadas para tomada de decisão de emprego antes de seu efetivo uso.	Uma lei estadual, representa uma importante iniciativa vinda dos Estados Unidos para regulação de IA em sua forma de ferramentas automatizadas, com foco em auditorias.
The Washington DC Algorithms Law (B25-0114)	-	Estados Unidos (Washington)	O projeto de lei "Pare a Discriminação de Algoritmos" foi reintroduzida no Distrito de Columbia em fevereiro de 2023. O objetivo é proibir que os usuários de decisões automatizadas utilizem essas decisões por meio de critérios de elegibilidade discriminatórios. Dentre as obrigações propostas, está a realização obrigatória de auditoria anual e requerimentos de transparência.	Projeto de lei estadual, com preocupação específica para o potencial discriminatório de decisões automatizadas, que deve ser comparado em razão de sua particularidade para lidar com casos de discriminação nesse contexto.

Assembly Bill 331 on Automated Decision Tools	-	Estados Unidos (Califórnia)	O projeto de lei exige, entre outras coisas, que agentes de decisão automatizada realizem anualmente uma avaliação de impacto para qualquer ferramenta de decisão.	Projeto de lei do Estado da Califórnia direcionado para o uso de IA em decisões automatizadas que reforça a importância de que estas ferramentas passem por uma avaliação de impacto com periodicidade anual.
AI Disclosure Act of 2023 (federal USA)	-	Estados Unidos (federal)	De acordo com o projeto, todo o material gerado pela tecnologia de inteligência artificial teria que incluir um aviso expresso de que foi gerado por IA.	Diante das discussões decorrentes das IAs generativas, esse projeto de lei estadunidense propõe maior transparência deste uso da IA, o que pode ser um exemplo para outras iniciativas regulatórias no campo de IA.
Algorithmic Accountability Act EUA	AAA	Estados Unidos (federal)	Projeto de lei reintroduzido no congresso norte-americano em fevereiro de 2022. Caso aprovado, o projeto será vinculante e obrigará as empresas a avaliarem o impacto de sistemas automatizados em termos de vieses e efetividade.	O único projeto de lei federal do contexto estadunidense no tópico de IA, com foco específico em garantir que os sistemas de IA passem por mecanismos de prestação de contas, como a realização de uma avaliação de impacto.
Canada's Artificial Intelligence and Data Act	AIDA	Canadá	AIDA é parte do Projeto de Lei C-27, Lei de Implementação da Carta Digital, 2022. A AIDA representa um marco importante na implementação da Carta Digital e na garantia de que os canadenses possam confiar nas tecnologias digitais, garantindo que a concepção, desenvolvimento e utilização de sistemas de IA sejam seguros e respeitem os valores canadenses.	O framework proposto pela AIDA pretende ser o primeiro passo em direção a um novo regime regulatório criado para guiar a inovação em IA em uma direção positiva, a partir de coordenação com outras iniciativas internacionais. Por isso, o documento serve como um resumo do que vem sendo implementado mundialmente, já que o objetivo canadense era de trazer esse diálogo entre fontes estrangeiras. A proposta caminha em direção à abordagem baseada em riscos e traz a ferramenta de avaliação de impacto, além de expressamente prever a não rivalização entre regulação e incentivo à inovação.

Relatórios do Comitê *Ad Hoc* de Inteligência Artificial do Conselho da Europa	CAHAI	Conselho da Europa - Comitê Ad Hoc de Inteligência Artificial (CAHAI)	O Comitê foi instalado no âmbito do Conselho da Europa, com mandato de 2019 a 2021, para examinar a viabilidade e os elementos potenciais, com base em amplas consultas multilaterais, de um quadro jurídico para o desenvolvimento, conceção e aplicação da inteligência artificial, baseado nas normas do Conselho da Europa em matéria de direitos humanos, democracia e Estado de direito.	O CAHAI foi fonte de diferentes estudos de viabilidade regulatória, inclusive estudos relativos à avaliação de impacto algorítmico que considera os Direitos Humanos, Democracia e Estado de Direito como norte de análise, o que é essencial para a análise da possível estrutura de uma avaliação de impacto algorítmico.
Projeto de Convenção de Inteligência Artificial, Direitos Humanos, Democracia e Estado de Direito	Convenção	Conselho da Europa - Comitê de Inteligência Artificial (CAI)	O CAI tem o mandato de 2022 e 2024 no âmbito do Conselho da Europa e tem como principal entregável, até 15/11/2023, um instrumento jurídico adequado (Convenção) para o desenvolvimento, concepção e aplicação de sistemas de inteligência artificial baseados nas normas do Conselho da Europa em matéria de direitos humanos, democracia e Estado de direito, e conducentes à inovação, em conformidade com as decisões relevantes do Comité de Ministros.	A futura Convenção de IA, Direitos Humanos, Democracia e Estado de Direito do Conselho da Europa será o primeiro documento que cria regras vinculantes para regular a IA em âmbito internacional, fruto de anos de estudo e pesquisa de grupo de trabalho especializado na área. Um dos achados principais de análise é a escolha pela regulação baseada no risco e a imposição de obrigações de elaboração de avaliações de impacto em determinados casos.
Washington SB 5116 - 2021-22	-	Estados Unidos (Washington)	Estabelece diretrizes para uso e compra governamentais de sistemas de decisão automatizados, com objetivo de proteger os consumidores, melhorar a transparência e criar mais previsibilidade do mercado.	Projeto de lei estadual que prevê ferramentas de prestação de contas para agências públicas que buscam desenvolver, usar ou comprar sistemas de IA para decisão automatizada, trazendo como medida, por exemplo, a figura do "relatório de prestação de contas algorítmica" como obrigatório, inclusive seu envio para a autoridade competente, que publicará o documento para comentários públicos. É um exemplo de projeto específico para o poder público que demanda o desenvolvimento de ferramentas de prestação de contas com participação pública.

Blueprint for an AI Bill of Rights	Blueprint	Estados Unidos (federal)	Documento não vinculante publicado pela Casa Branca em outubro de 2022 para guiar o design, desenvolvimento e implantação de sistemas de IA. O documento se baseia em 5 princípios: (i) sistemas seguros e eficazes; (ii) proteção contra discriminação algoritmica; (iii) privacidade de dados; (iv) aviso e explicação; (v) alternativas humanas, consideração e recurso.	Mesmo não apresentando força vinculante, a princípio[20], é um importante documento de referência, já que foi criado pela Casa Branca com o intuito de direcionar padrões para o design, desenvolvimento e implementação de sistemas de IA nos Estados Unidos.
Artificial Intelligence Risk Management Framework (AI RMF 1.0) - NIST	AI RMF (NIST)	Estados Unidos (federal)	O AI RMF é uma estrutura voluntária que visa fornecer às organizações um processo para ajudar a enfrentar os riscos ao longo do ciclo de vida da IA, com o objetivo de promover sistemas de IA confiáveis e responsáveis. Destina-se a ajudar a gerir os riscos empresariais e sociais relacionados com a concepção, desenvolvimento, implantação, avaliação e utilização de sistemas de IA. Foi desenvolvido pelo Instituto Nacional de Padrões e Tecnologia do Departamento de Comércio dos Estados Unidos (NIST). Não tem força vinculante e se baseia em 4 princípios principais	O objetivo é oferecer um recurso para as organizações que projetam, desenvolvem, implantam ou usam sistemas de IA para ajudar a gerenciar os seus riscos e promover o desenvolvimento e uso confiável e responsável de sistemas de IA. O framework é voluntário, não específico de um setor e independente do tamanho da organização que pretende usá-lo. É um importante modelo prático para a gestão de riscos de IA, já sendo seguido por diferentes organizações, mesmo não tendo força vinculante, à princípio.[21] Traz metodologia bem definida, inclusive com critérios qualitativos e quantitativos.

20 O documento torna-se vinculante para algumas instituições em determinados casos, tais como para órgãos do governo federal após a publicação da Ordem Executiva de 30 de outubro de 2023 pelo presidente Joe Biden (Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence) – vide Seg. 10, 10.2 (b) (iv).

21 Vide nota de rodapé 14.

Directive on Automated Decision-Making + Algorithmic Impact Assessment tool	-	Canadá	Diretiva aplicável para sistemas de decisão automatizadas desenvolvidos ou aplicados após abril de 2020. Nesta diretiva, é prevista a Ferramenta de avaliação de risco criada para ajudar os departamentos e agências do Canadá a entender e gerenciar melhor os riscos associados aos sistemas de decisão automatizadas. A ferramenta é um questionário que determina o nível de impacto de um sistema de decisão automatizado (composta por 48 questões de riscos e 33 questões de mitigações). É uma ferramenta de avaliação de risco obrigatória e destinada a apoiar a diretiva do Conselho do Tesouro do Canadá sobre tomada de decisão automatizada.	A diretiva canadense é um exemplo já em vigor de legislação direcionada especificamente para o uso de IA em sistemas automatizados para tomada de decisões administrativas para concessão de benefícios sociais. Seu diferencial é o fornecimento de ferramenta própria para que os órgãos públicos façam o gerenciamento dos riscos de seus sistemas, inclusive com a definição dos níveis de impacto e respectivas medidas de mitigação.
Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems	-	Canadá	O código fornece temporariamente às empresas canadenses para o desenvolvimento e utilização de sistemas de IA generativos de forma responsável até que a regulamentação formal entre em vigor.	Apesar de voluntário, fornece boas práticas concretas para o desenvolvimento e uso de IA generativa, o que pode servir de exemplo para pensar na regulação desse uso de IA.
Organização para a Cooperação e Desenvolvimento Econômico	OCDE	Internacional	A OCDE apoia os governos através da medição e análise dos impactos econômicos e sociais das aplicações de IA para identificar boas práticas para as políticas públicas, tendo uma série de publicações sobre IA e sua governança. A organização publicou em 2019 Princípios para a IA e criou um Observatórios para políticas públicas de IA, além de possuir diferentes estudos sobre a temática, como o modelo para classificação de IA e um relatório sobre prestação de contas em IA por meio da governança e gerenciamento de seus riscos, além do recente guia de interoperabilidade entre sistemas de gerenciamento de risco de IA.	Os princípios da OCDE representam o primeiro modelo de política de IA, servindo como base para outros documentos, nacionais e internacionais, e para avaliação do status de governança de IA em cada país. Logo, os documentos da organização, como os que visam auxiliar na classificação de sistemas de IA e sua prestação de contas por meio do gerenciamento desses riscos, servem também como importante base para os modelos regulatórios que pretendem regular os usos desta tecnologia.

Interim Measures for the Management of Generative Artificial Intelligence Services	-	China	Regras adotadas pelo Ministério da Ciência e Tecnologia da China para a IA generativa, com aplicação a partir de 15 de agosto de 2023.	É o primeiro documento chinês a lidar com as ferramentas de IA Generativa, servindo como exemplo de como a regulação desse uso da tecnologia pode ser desenvolvido.
Recommendation on the Ethics of Artificial Intelligence (UNESCO)	-	UNESCO (internaciona)	Adotado pela UNESCO em novembro de 2021, foi o primeiro instrumento padrão global para a Ética da IA adotado pelos 193 Estados-Membros, tendo como norte a proteção dos direitos humanos e da dignidade. Ademais, as recomendações trazem também áreas determinadas de ação política, que auxiliam os decisores políticos a traduzir os valores e princípios fundamentais em ação. Por fim, a recomendação já apresenta duas metodologias práticas que auxiliam também na sua aplicação prática: (i) Metodologia de Avaliação de Prontidão (MAP); (ii) Avaliação de Impacto Ético (AIE).	A Recomendação da UNESCO, assim como suas ferramentas práticas (especialmente a AIE) são referência internacional para países e organizações que pretendem desenvolver, implementar e utilizar sistemas de IA.
Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence	Executive Order	Estados Unidos (federal)	A ordem executiva, assinada em 30 de outubro de 2023 pelo presidente Joe Biden, possui força vinculante para órgãos públicos dos EUA, que terão que seguir com uma série de obrigações relativas à proteção dos cidadãos norte-americanos contra os potenciais riscos trazidos por sistemas de IA.	A Ordem Executiva torna vinculantes o framework de avaliação de riscos do NIST e o Blueprint para o governo federal dos EUA, o que faz com que sua análise seja necessária, especialmente a Seg. 10, 10.2 (b) (iv), além das sugestões para a regulação de modelos fundacionais de IA (IA generativa).

4. Eixos temáticos de análise

EIXO 1 – Regulação baseada no risco

Para se aprofundar na regulação baseada em risco, é necessário retornar alguns passos para explorar a noção de risco, que é um elemento inerente à vida que se afasta de um significado dualista (de existir ou não). Gellert (2017) conceitua o risco como uma ferramenta que auxilia o processo de tomada de decisão, direcionando sua análise não para sua existência – presumida –, mas para o quanto de risco determinado agente é capaz ou está disposto a assumir e quanto é capaz de mitigar[22].

De forma similar, Hood et al (2001) define risco como uma probabilidade de consequências adversas, sendo a regulação do risco a interferência governamental nos processos de mercado ou sociais para controlar essas potenciais consequências adversas[23]. Citando Beck (1992), Hood et al destacam que a atividade humana e a tecnologia na modernidade têm como efeito colateral riscos que dependem de experts para avaliar e reconhecer, são coletivos, globais e irreversíveis em seu impacto, resultando em uma "sociedade do risco", distinta dos períodos históricos anteriores[24].

Logo, o risco pode ser entendido como "a capacidade de definir o que pode acontecer no futuro e escolher entre alternativas"[25], funcionando como uma ferramenta para a tomada de decisões, na medida em que torna certo o incerto. Seus elementos constitutivos são duas operações distintas, porém unidas: previsão do futuro (com a ajuda de números) e a tomada de decisões com base nisso. Assim, o risco, embora associado a algo mais quantificável, também pode ser entendido como um elemento qualitativo e valorativo que necessita ser avaliado considerando diferentes perspectivas.

a.1) Regulação assimétrica e risco: panorama geral

A regulação do risco, como demonstrou Hood et al, varia consideravelmente em relação a quais riscos são escolhidos para regulação e a maneira na qual a regulação fun-

22 GELLERT, Raphaël. Understanding the notion of risk in the General Data Protection Regulation. **Computer Law & Security Review**: The International Journal of Technology Law and Practice (2017). p. 02; GOMES, Maria Cecília O.Entre o método e a complexidade: compreendendo a noção de risco na LGPD. *In:* **Temas atuais de proteção de dados**. PALHARES, Felipe (Coord.). São Paulo: Thomson Reuters Brasil, 2020, pp 245-271.

23 HOOD, Christopher; ROTHSTEIN, Henry; BALDWIN, Robert.The Governance of Risk: Understanding Risk Regulation Regimes. Nova York: **Oxford University Press**, 2001. ISBN 0-19-924363-8.

24 Nas palavras dos autores: *"(...) we live today in a 'risk society'. By that Beck means that risk has a different significance for everyday life from that applying in previous historical eras. Human activity and technology in 'advanced modernity',he claims, produces a side-effect risks that need specialized expertise to assess and recognize, are collective, global, and irreversible in their impact, and thus potentially catastrophic on a scale never seen before.";* HOOD, *et al*, 2001, p. 3.

25 BERNSTEIN, Peter L. Against the Gods: The Remarkable Story of Risk. Wiley, 1996. p. 2.

ciona, não só entre diferentes ordenamentos jurídicos (diferentes países), como dentro de um mesmo ordenamento jurídico[26].

Em relação à governança dos sistemas de Inteligência Artificial, Kaminski (2022) constata a escolha de ferramentas de regulação de risco tanto nos Estados Unidos quanto na Europa para tanto. Essa escolha comum, entretanto, não significa que existe um único modelo de regulação de risco nas normativas analisadas. A autora destaca quatro modelos de regulação de risco. São eles:

(i) modelo de regulação de risco quantitativo, originado no direito administrativo dos EUA;

(ii) modelo de regulação de risco que estabelece supervisão democrática para problemas que afetam toda população, como a legislação ambiental dos EUA ("NEPA");

(iii) (iii) modelo de regulação de risco que distribui a aplicação da legislação e capacidade regulatória baseada nos riscos, ou seja, aloca recursos regulatórios contextualmente e de forma transversal, como ocorre no Reino Unido;

(iv) (iv) modelo de regulação de risco empresarial.

(KAMINSKI 2022)

O modelo de regulação de risco originado no direito administrativo dos EUA nos anos 1960-1980 para regular questões de saúde, segurança e ambientais tem como características a definição formal e quantitativa de risco e análises de custo-benefício, onde os potenciais danos deveriam ser conhecidos e medidos para que fossem ou regulados ou banidos. Nesse modelo, o risco é calculado como o produto da probabilidade e da severidade de sua consequência para definir o quanto de risco é "aceitável" na prática em troca dos benefícios potencialmente gerados[27]. Nesse contexto, segundo Boyd (2012), a noção de segurança no âmbito da proteção do consumidor foi explicitamente redefinida como a de "risco aceitável", a exemplo da definição do nível de 100 milhões como uma quantidade "aceitável" pela U.S. Food and Drug Administration (FDA) para um agente cancerígeno chamado dietilstibestol nos anos 1970.

Outra abordagem surgida nos EUA, o modelo de regulação de risco estabelecendo supervisão democrática para problemas que afetam toda a população, tem como principal exemplo a Política Nacional do Meio Ambiente (NEPA) que requer a condução de

26 HOOD *et al*, 2001.

27 KAMISNKI, 2022, p. 36; BOYD, Willian. Genealogies of Risk: Searching for Safety, 1930s–1970s. Ecology Law Quarterly, nº 895, 2012. Disponível em: https://scholar.law.colorado.edu/faculty-articles/143/.

uma avaliação de risco antes do início de um projeto, com sua posterior publicação para a população em geral para ser discutida[28]. De forma similar, o Brasil também apresenta uma estrutura muito rica de supervisão democrática de problemas que afetam toda a população, como no caso de questões ambientais. Nesse contexto, o ordenamento jurídico brasileiro apresenta diferentes formas de a sociedade civil atuar na proteção ambiental, a exemplo da participação, na formulação e na execução de políticas ambientais, seja mediante a atuação de representantes da sociedade civil em órgãos colegiados dotados de poderes normativos, seja na possibilidade de participação em audiências públicas no âmbito dos estudos de impacto ambiental, ou até mesmo pela participação nos conselhos municipais[29] do meio ambiente[30].

O terceiro modelo se iniciou no Reino Unido[31] e se espalhou internacionalmente nos anos 2000. O foco da regulação está em uma administração central que avalia riscos no nível macro e aloca recursos para tanto. Nesse modelo, os reguladores fazem a identificação do risco a ser gerenciado; selecionam o seu nível de tolerância; avaliam os danos e a probabilidade de sua ocorrência; estabelecem scores de risco para empresas e atividades (a exemplo de "alto", "médio" ou "baixo"); e conectam a alocação de recursos de compliance e inspeção a esses scores de risco[32]. Nesse sentido:

> "Os reguladores e o sistema regulatório como um todo devem usar uma avaliação de risco abrangente para concentrar recursos nas áreas que mais precisam deles"
>
> (...)
>
> "Os reguladores devem reconhecer que um elemento-chave de sua atividade será permitir, ou mesmo encorajar, o progresso econômico e apenas intervir quando houver um caso claro

28 KAMINSKI, 2022.

29 Conselhos de Meio Ambiente existem em âmbito federal (Conselho Nacional de Meio Ambiente (CONAMA)), nos Estados (Conselho Estadual de Meio Ambiente (COEMA)) e nos Municípios (Conselho Municipal de Meio Ambiente (CONDEMA) como integrantes do Sistema Nacional do Meio Ambiente (Sisnama – criado pela Lei 6.938/81). Tais conselhos são órgãos colegiados normativos – isto é, com poderes robustos para propor normas e diretrizes relativas à gestão ambiental – e compostos por representantes de órgãos públicos, setor empresarial e sociedade civil; FIGUEIRA, Paulo Sérgio Sampaio. O papel do conselho do meio ambiente nas políticas públicas ambientais. Publicado em 14 de abril de 2022. Disponível em: https://direitoambiental.com/o-papel-do-conselho-do-meio-ambiente-nas-politicas-publicas-ambientais/.

30 COLOMBO, Silvana. Os mecanismos de participação popular na gestão do meio ambiente à luz do texto constitucional: aspectos positivos e negativos. Editora Unijuí: Revista Direitos Humanos e Democracia, ano 9, nº 18, jul/dez 2021.

31 A escola inglesa de estudos sobre regulação pode ser entendida como o conjunto de teorias, estudos e pesquisas que foram desenvolvidas no Centre for Analysis of Risk and Regulation (CARR) da London School of Economics„ em que se destacam autores como Christopher Hood e Julia Black; ZANATTA, Rafael A. F. Proteção de Dados Pessoais como Regulação de Risco: uma nova moldura técnica?. I ENCONTRO DA REDE DE PESQUISA EM GOVERNANÇA DA INTERNET, NOVEMBRO DE 2017. Disponível em: https://www.redegovernanca.net.br/public/conferences/1/anais/ZANATTA,%20Rafael_2017.pdf. p. 182.

32 KAMINSKI, 2022, p. 37.

de proteção.

(HAMPTON, 2005, p. 7, tradução própria)

Por fim, no gerenciamento de risco empresarial, as empresas se auto-organizam para reduzir seus riscos, como responsabilidades ou outras penalidades, sejam de mercado ou regulatórias. Para isso, realizam avaliações cíclicas e contínuas de risco, baseadas em uma cultura organizacional da empresa de redução de riscos desde o design de seus produtos até o momento posterior à implementação[33]. Esse modelo aproxima-se mais da noção de autorregulação, já que, diferentemente dos outros modelos mencionados, há menor presença do Estado no gerenciamento do risco, pois a gestão do risco pode ocorrer por iniciativa própria das instituições na ausência de regulação ou por meio de incentivo estatal no caso de publicação de recomendações, fiscalização ou de ameaça de aplicação regulamentar[34].

O que se observa é que a análise de risco e mitigação pode ser realizada em um nível micro (da empresa e de um setor, por exemplo) ou no nível macro (como nos mercados em geral de forma mais transversal, com participação e intervenção do estado). No nível micro, o processo costuma ser iniciado por uma análise do sistema para identificação de riscos, seguida de sua mitigação e, ao final, teste para garantir que esta mitigação foi efetiva. Esse processo pode ser contínuo, incluindo a análise do comportamento da tecnologia após inserida no mercado. No nível do mercado, a análise de risco e mitigação se torna uma abordagem regulatória. Os reguladores, entidades públicas e privadas[35], identificam riscos e catalogam com diferentes níveis de risco determinadas empresas ou atividades. A partir dessa avaliação, são distribuídos os recursos de inspeção e investigação do regulador para monitorar a atividade empresarial[36].

Para compreender o panorama internacional da regulação de riscos e transportá-lo para o contexto de IA, também é necessário resgatar a abordagem baseada em riscos adotada por outros campos. Este é o caso do de saúde, alimentos, ambiental, de seguros, consumidor e, mais recentemente, da proteção de dados pessoais, já que os "regimes" de regulação de risco podem variar de um domínio para outro, inclusive se alterando ao

33 Ibid.

34 Ibid, p. 36.

35 De acordo com Julia Black, a regulação pode ser vista a partir de uma perspectiva descentralizada, no que conceitua como regulação policêntrica ou multimodal, isto é, não dependente apenas das forças do Estado, mas que decorre de muitos fóruns (nacionais, subnacionais e internacionais), inclusive atores não estatais como empresas, a sociedade civil organizada, pessoas que controlam os principais recursos de que as empresas necessitam (por exemplo, agências de notação de crédito, seguradoras, auditores, fornecedores de serviços de Internet, etc), "empresários políticos", entre outros; BLACK, Julia. *Proceduralisation and polycentric regulation.* Revista Direito GV, Especial 1, pp. 099–130, 2005. p. 105–110.

36 KAMINSKI, 2022.

longo do tempo[37]. Nesse contexto, o risco pode ser entendido como elemento central e que foca em processos, tais como a coleta de informação e cognição dos riscos; desenvolvimento de regras e padrões de conduta; e enforcement e monitoramento da modificação dos comportamentos de acordo com os padrões criados[38].

No caso específico da proteção de dados pessoais, o Regulamento Geral de Proteção de Dados da União Europeia (General Data Protection Regulation – GDPR) é um exemplo de regulação baseada no risco, que prevê uma versão flexível da regulamentação de baixo para cima (ou flexível de cima para baixo), responsiva às entidades reguladas e alocando os recursos de fiscalização dos reguladores por risco. Nesse sentido, Quelle[39], definindo a abordagem baseada em riscos da GDPR ("risk-based approach"), afirma que ela "introduz a noção de risco como uma referência obrigatória para a calibragem das obrigações legais dos controladores"[40]. Em outras palavras, a abordagem baseada em riscos afeta quais são as obrigações dos controladores em cada caso concreto, o que faz com que a lei de proteção de dados se aplique de forma diferente a depender do nível de risco de determinada atividade[41].

Nesse caso, não há uma substituição dos princípios e regras da proteção de dados por uma mera análise de riscos nessa abordagem. A partir do grau de risco, considerando contextualmente a severidade e probabilidade, as obrigações de cada um dos controladores é parametrizada com mais ou menos obrigações, direitos e deveres – quanto maior o risco, maior a carga obrigacional[42].

Assim, na abordagem baseada em riscos da GDPR, são destacados os possíveis resultados de determinado tratamento de dados, para então avaliar se os direitos e liberdades dos indivíduos estão sendo respeitados nos termos da lei[43]. Essa avaliação, em situações de alto risco, é, por exemplo, consubstanciada na obrigação de elaboração de um DPIA ("Data Protection Impact Assessment" ou "avaliação de impacto em proteção de dados"), uma das formas de avaliação de impacto, considerada ferramenta chave na abordagem baseada em riscos – que será melhor explorada no eixo 2 deste estudo.

37 Hood *et al*, 2001, p. 8.

38 Ibid.

39 QUELLE, Claudia. 'The 'risk revolution' in EU data protection law: We can't have our cake and eat it, too' in R Leenes, R van Brakel, S Gutwirth and P De Hert (eds), Data Protection and Privacy: The Age of Intelligent Machines (Hart Publishing, forthcoming). 2017.

40 "The relationship between the risk-based approach and adherence to the legal requirements of data protection is addressed in particular by articles 24, 25(1) and 35 of the GDPR. These provisions determine how controllers should give hands and feet to data protection law in practice." ("Data Protection and Privacy : The Age of Intelligent Machines", 2017, p. 8); QUELLE, 2017, p. 1.

41 QUELLE, 2017.

42 QUELLE, 2015; ZANATTA, 2017.

43 QUELLE, 2017.

Dito isso, utilizando o contexto de proteção de dados como ilustrativo, a ideia em torno da regulação de risco assimétrica é calibrar o peso da regulação – a intensidade de obrigações, direitos e deveres de determinado agente regulado – de acordo com o nível do risco calculado.

Regulação assimétrica e risco nas propostas regulatórias brasileiras em IA

No contexto brasileiro, a versão final aprovada em setembro de 2021 pela Câmara dos Deputados do PL 21/20 traz uma regulação fundamentalmente principiológica e sucinta, com apenas 16 artigos. O termo "risco" aparece apenas em três deles: (a) artigo 2º, VI – apresenta uma definição do relatório de impacto de inteligência artificial; (b) artigo 6,V – na definição do princípio de segurança; e (c) artigo 9, IV – coloca como dever dos agentes de IA a implantação de sistema de inteligência artificial após avaliação dos seus objetivos, benefícios e riscos relacionados a cada fase do sistema.

Nesse caso, o risco não é utilizado de maneira sistemática para organizar a abordagem regulatória, figurando apenas como mais um de vários outros elementos conceituais do projeto. Por isso, nesse caso, não há nem que se falar em regulação baseada no risco e muito menos em diferentes graus de risco de cada um dos sistemas de IA, já que o PL não a procedimentaliza ao não trazer uma definição mínima dos possíveis graus de risco e dos instrumentos de governança. Tal conclusão é também extraída da análise do PL 872/21 que menciona "riscos" apenas uma vez ao definir, no inciso VII do art. 4º, que as soluções de IA devem "seguir padrões de governança que garantam o contínuo gerenciamento e a mitigação dos riscos potenciais da tecnologia", porém, sem trazer maiores explicações do que consistiria esses processos. Já o PL 759/23 sequer menciona o termo risco. A regulamentação é ainda mais sucinta, com 7 artigos, dividida entre: princípios (artigo 2º), diretrizes (art. 3º), critérios a serem obedecidos pelos sistemas de Inteligência artificial (art. 4º), obrigação do Poder Executivo de criar uma Política Nacional de Inteligência artificial (art. 5º) e faculdade dos entes públicos para celebração de convênios com entidades privadas ou públicas, nacionais ou internacionais para apoio e fortalecimento da Política Nacional de Inteligência Artificial (art. 6º). Nenhuma menção ao termo "risco" aparece no PL 5051/19, que aparece como outro exemplo de projeto com abordagem regulatória de cunho principiológico.

O PL 2338/23 alia duas abordagens diferentes: baseada em direitos e em risco. A abordagem baseada em direitos permite a proteção da pessoa natural impactada por sistemas de inteligência artificial, ao mesmo tempo em que a abordagem baseada em risco, ao regulamentar a governança dos sistemas de inteligência artificial, garante previsibilidade e segurança jurídica para inovação e o desenvolvimento tecnológico. Essa

aliança busca harmonizar a proteção de direitos e liberdades fundamentais, valorização do trabalho e dignidade da pessoa humana à criação de novas cadeias de valor e desenvolvimento da ordem econômica:

> Estruturalmente, a proposição estabelece uma regulação baseada em riscos e uma modelagem regulatória fundada em direitos. Apresenta ainda instrumentos de governança para uma adequada prestação de contas dos agentes econômicos desenvolvedores e utilizadores da inteligência artificial, incentivando uma atuação de boa-fé e um eficaz gerenciamento de riscos.
>
> (PL 2338/23, p.29)

O modelo adotado pelo PL 2338/23 utiliza o risco, à semelhança das mais recentes iniciativas de regulação de IA (vide: tabela abaixo), para calibrar as obrigações dos agentes dos sistemas de IA. Existem direitos básicos que se aplicam a qualquer interação entre o sistema de IA e um ser humano (conforme art.5, I, II, IV, V e VI, art.7, art.8, art. 12) na toada de uma regulação baseada em direitos, como informação e transparência. Entretanto, há mais obrigações quando há um maior risco a direitos (art. 5, III, art. 9, art. 10, art. 11). Da mesma maneira, as medidas de governança dos sistemas de inteligência artificial também são divididas de acordo com o risco:

> Além de fixar direitos básicos e transversais para todo e qualquer contexto em que há interação entre máquina e ser humano, como informação e transparência, intensifica-se tal obrigação quando o sistema de IA produz efeitos jurídicos relevantes ou impactem os sujeitos de maneira significativa (ex: direito de contestação e intervenção humana). Assim, o peso da regulação é calibrado de acordo com os potenciais riscos do contexto de aplicação da tecnologia. Foram estabelecidas, de forma simétrica aos direitos, determinadas medidas gerais e específicas de governança para, respectivamente, sistemas de inteligência artificial com qualquer grau de risco e para os categorizados como de alto risco.
>
> (PL 2338/23, p.30-31)

A separação de medidas de governança de acordo com o risco de determinado sistema de IA em um contexto específico se assemelha à regulamentação baseada em risco originada no Reino Unido pela escola inglesa de regulação[44], descrita por Kaminski, e à abordagem do EU AI Act. Nessa versão de regulação de risco, há, historicamente, previsão de supervisão estatal que co-avalia e co-atribui riscos a determinadas empresas e atividades, dividindo-os, por exemplo, entre alto, médio e baixo, e aloca os recursos de aplicação da lei e investigação de acordo com esses riscos[45]. É o que faz, por exemplo, o PL 2338/23 ao prever que, além de haver a necessária designação de uma autoridade competente (art. 32, caput) que este órgão regulador deverá cooperar com outros de competências correlatas para cognição e gerenciamento de riscos (artigo 32, incisos V, VII e VIII). Em especial, quando se tratar de setores econômicos específicos em que haverá variação contextual dos riscos no desenvolvimento e implementação de IA (art. 34, caput e §1º).

Ainda, o PL 2338/23 diferencia-se qualitativamente das demais propostas regulatórias nacionais ao prever um capítulo para governança e boas práticas de modo a incentivar que os próprios agentes econômicos gerenciem os riscos das suas próprias atividades econômicas. Inclusive, com expressa previsão de que cabe às autoridades regulatórias promover "estudos sobre boas práticas no desenvolvimento e utilização de sistemas de inteligência artificial" e "ambientes regulatórios experimentais" (sandboxes). Esse arranjo de parceria pública- privada na cognição dos riscos associadas à IA é ainda complementada pelo dever do aparato estatal prestar contas sobre suas escolhas regulatórias (e.g., avaliações de impacto regulatório e consultas públicas) e, como será visto mais à frente, por um modelo o mais participativo possível para a elaboração de avaliações de impacto algorítmico que será uma das documentações a comporem uma base de dados pública e aberta sobre IAs de alto risco.

Em resumo, como sintetiza a tabela abaixo, o PL 2338/23 não apenas quantitativamente enuncia risco como elemento organizador da regulação. Também, de forma qualitativa, procedimentaliza a maneira pela qual devem ser alocados recursos institucionais e ferramentas para, democraticamente, se decidir quais riscos são (in)aceitáveis e como gerenciá-los.

44 A escola inglesa de estudos sobre regulação pode ser entendida como o conjunto de teorias, estudos e pesquisas que foram desenvolvidas no Centre for Analysis of Risk and Regulation (CARR) da London School of Economics„ em que se destacam autores como Christopher Hood e Julia Black; ZANATTA, Rafael A. F. Proteção de Dados Pessoais como Regulação de Risco: uma nova moldura técnica?. I ENCONTRO DA REDE DE PESQUISA EM GOVERNANÇA DA INTERNET, NOVEMBRO DE 2017. Disponível em: https://www.redegovernanca.net.br/public/conferences/1/anais/ZANATTA,%20Rafael_2017.pdf. p. 182.

45 KAMINSKI, 2022.

MODELOS DE REGULAÇÃO DO RISCO CONFORME A PROPOSTA DE REGULAÇÃO DE IA OU DOCUMENTO INTERNACIONAL[46]

Exposição de motivos da proposta / Explicação do documento de *soft law*		Modelo quantitativo	Modelo Quantitativo (Potencialmente Híbrido)		
			Supervisão democrática	Previsão de alocação de recursos (e.g., criação ou designação órgão regulador)	Regulação de risco empresarial
PL 21/20	Justificativa do texto do PL 21/20 do deputado Eduardo Bismarck menciona o termo risco no contexto de deveres relacionados ao gerenciamento dos riscos gerados por sistema de IA, em avaliação conjunta com seus potenciais benefícios. A regulação que busca "tornar obrigatórios princípios consagrados internacionalmente e disciplinar direitos e deveres", fomentando a adoção da IA para "promover a pesquisa e inovação, aumento da produtividade, desenvolvimento de atividade econômica sustentável, melhoria do bem estar das pessoas e auxílio na resposta aos principais desafios globais". Porém, o texto aprovado na Câmara dos Deputados é oriundo do substitutivo da deputada Luísa Canziani que, apesar de alterar o projeto, não possui exposição de motivos.	Baixo-Inexistente	Baixo-Inexistente	Baixo-Inexistente (Propõe apenas Diretrizes para atuação do Poder Público e de coordenação de Autoridades Setoriais já existentes - e.g, Art. 6º)	Médio (e.g, Arts. 3º, VIII, e 8º, III)

46 Esta tabela não apresenta uma comparação conclusiva opinativa de todas as propostas regulatórias, mas auxilia no mapeamento inicial de seus pontos relevantes para avançar nesse tipo de comparação, onde uma proposta poder ser mais ou menos híbrida de acordo com as modalidades de regulação de risco classificadas por Kaminski (2022).

PL 2338/23	A exposição de motivos da CJSUBIA no relatório apresentado em dezembro de 2022 e a justificativa do PL 2338/23 destacam que "a proposição estabelece uma regulação baseada em riscos e uma modelagem regulatória fundada em direitos. A proposta busca "conciliar, na disciplina legal, a proteção de direitos e liberdades fundamentais, a valorização do trabalho e da dignidade da pessoa humana e a inovação tecnológica representada pela inteligência artificial. A ideia é que o peso da regulação seja dinamicamente "calibrado de acordo com os potenciais riscos do contexto de aplicação da IA". Para isso, "foram estabelecidas, de forma simétrica aos direitos, determinadas medidas gerais e específicas de governança para, respectivamente, sistemas de inteligência artificial com qualquer grau de risco e para os categorizados como de alto risco". Em outras palavras, o peso regulatório (maior número de obrigações legais) aumenta conforme aumenta também o nível de risco do sistema de IA.	Alto (Prevê ampla taxonomia de riscos: vide tabela do tópico a.2 infra - Graus de Risco, Riscos Excessivos e Alto Risco - e.g, Arts. 14 e 17)	Alto (Prevê uma série de obrigações de controle social e participação pública na produção regulatória e gerenciamento de risco - e.g, Capítulo IV sobre governança dos sistemas de IA)	Médio (Além de propor Diretrizes para atuação do Poder Público e de coordenação de Autoridades Setoriais já existentes, há previsão de uma nova autoridade para coordenação de tais esforços por parte do Poder Executivo - e.g, Art. 21 e Seção I do Capítulo VIII)	Alto (e.g, Capítulo VI - Códigos de Boas Práticas e Seção Dedicada a Fomentar Inovação)
EU AI Act	Nas razões e objetivos da Proposta, destaca-se que a União "está empenhada em alcançar uma abordagem equilibrada". Nesse sentido, destacam que "é do interesse da União preservar a liderança tecnológica da UE e assegurar que novas tecnologias, desenvolvidas e exploradas respeitando os valores, os direitos fundamentais e os princípios da União, estejam a serviço dos cidadãos europeus".	Alto (Prevê ampla taxonomia de riscos: vide tabela do tópico a.2 infra - Inaceitável e Alto - e.g, Arts. 5º e 6º + Anexo III)	Alto (e.g, Considerando 81 e Art. 29a (4) - versão do texto do PE)	Alto (e.g, diferentes medidas de governança e implementação nos títulos VI, VII e VIII, respectivamente, com a criação do Conselho Europeu de Inteligência Artificial, coordenação entre esse Conselho e as demais autoridades nacio-	Alto (e.g, Título IX sobre Códigos de Conduta; Título V de medidas de apoio à inovação)

EU AI Act	Menciona-se a adoção pela "abordagem regulamentar baseada no risco bem definida que não cria restrições desnecessárias, uma vez que a intervenção legal é adaptada às situações concretas em que exista um motivo justificado de preocupação ou quando tal preocupação puder ser razoavelmente antecipada em um futuro próximo", além de trazer "mecanismos flexíveis" de adaptação dinâmica da regulação de acordo com os avanços tecnológicos e surgem novas situações preocupantes.			nais e medidas de monitoramento pós-comercialização))	
Directive on Automated Decision-Making + Algorithmic Impact Assessment tool (Canadá)	Na explicação da Diretiva há a definição de que seu objetivo é garantir que sistemas de decisões automatizadas sejam implementados de forma a reduzir os riscos para a sociedade canadense. Para isso, prevê a obrigação de condução de uma avaliação de impacto algorítmico, inclusive fornecendo ferramenta prática (tool).	Alto (Prevê ampla taxonomia de riscos: vide tabela do tópico a.2 infra - e.g. Anexo B)	Alto (menciona a necessidade de consulta às partes interessadas internas e externas na página da ferramenta)	Médio (designa autoridade já existente - e.g. Art.2º)	Baixo-Inexistente (a princípio, aplica-se apenas para decisões automatizadas utilizadas para tomar decisão administrativa - e.g. Art. 5º)
AIDA	O documento explicativo da AIDA menciona que a regulação terá uma abordagem "baseada no risco", de forma a alinhar-se com outras regulações em desenvolvimento no plano internacional. O objetivo é construir um "quadro destinado a garantir a identificação proativa e a mitigação de riscos, a fim de prevenir danos e resultados discriminatórios, reconhecendo ao mesmo tempo a natureza única do ecossistema de IA e garantindo que a investigação e a inovação responsável sejam apoiadas". Nesse sentido, o documento é expresso em definir que "na medida em que a	Alto (menção expressa a sistemas de IA de alto risco e práticas proibidas)	Alto (o texto da AIDA ainda não foi divulgado, mas em seu documento de estudo há clara menção à "extensa consulta a uma gama de agentes interessados" para a construção da regulação)	Alto (mencionam que já há autoridades, mas que os riscos da IA criam a necessidade de novas ações + designação do Ministério de Inovação, Ciência e Indústria como a autoridade competente para implementar e fiscalizar a AIDA)	Alto (clara preocupação de criar uma regulação proporcional que não impeça a inovação)

AIDA	tecnologia evolui, novas capacidades e usos de IA irão surgir e o Canadá precisa de uma abordagem que se adapte a esse cenário de constante mudança". Ademais, o documento explicativo da proposta da AIDA salienta que as medidas obrigacionais dos atores de IA serão determinadas de acordo com o "contexto e os riscos associados com atividades reguladas específicas dentro do ciclo de vida de um sistema de IA de alto risco". Nesse sentido, "as atividades reguladas definidas na AIDA seriam associadas com obrigações distintas que são proporcionais ao risco", evitando "impactos indevidos na inovação".				
Algorithmic Accountability Act EUA	O projeto visa instruir a Federal Trade Commission a exigir avaliações de impacto de sistemas de decisão automatizados e processos de decisão críticos. Assim, pretende exigir que as empresas avaliem os impactos dos sistemas automatizados que utilizam e vendem, além de criar uma nova transparência sobre quando e como os sistemas automatizados são utilizados e capacitar os consumidores a fazerem escolhas informadas sobre a automatização de decisões críticas.	Médio (não traz uma taxonomia de riscos bem definida, mas direciona a regulação para riscos mais elevados)	Alto (na condução de avaliações de impacto, deve haver consulta a importantes stakeholders - e.g, Section 3, (b) 1. (G); Section 4, (a) (2))	Alto (direciona ao Federal Trade Commission (FCT) a obrigação de exigir avaliações de impacto dos agentes que utilizam sistemas de decisões automatizadas, além de criar obrigações para outras autoridades - e.g. Sections 8 e 9)	Médio (possibilidade de assistência técnica e orientação dos atores regulados pela FCT - e.g. Section 7)
Proyecto de Ley 15869/19 (Chile)	O projeto tem como inspiração o EU AI Act e, por isso, também apresenta uma abordagem baseada no risco, trazendo como fundamento a necessidade de enfrentar o rápido avanço das tecnologias, tanto nos seus aspectos positivos como nos riscos associados ao seu uso.	Alto (Prevê taxonomia de riscos: vide tabela do tópico a.2 infra - Graus de Risco, Riscos Ina-	Alto (Prevê uma série de obrigações de controle social e participação pública na produção regulatória e gerencia-	Médio (Além de propor Diretrizes para atuação do Poder Público e de coordenação de Autoridades Setoriais já existentes, há previsão de uma nova	Alto (e.g, Capítulo VI - Códigos de Boas Práticas e Seção Dedicada a Fomentar Inovação)

Proyecto de Ley 15869/19 (Chile)	O objetivo é "estabelecer uma área de soberania digital para sistemas de inteligência artificial, na qual o Estado do Chile é quem discute as considerações éticas e legais, além de regular os riscos decorrentes do desenvolvimento, distribuição, comercialização e utilização desta tecnologia" e estabelecer limites, formalidades e requisitos de implementação e aplicação para qualquer pessoa que exerça as suas ações com a tecnologia. Para cumprir este objetivo, o projeto de lei estabelece a criação da Comissão Nacional de Inteligência Artificial, que terá, entre suas competências, propor a ampliação ou atualização da regulamentação sobre IA; avaliar e autorizar (ou proibir) sistemas de IA; e manter um registro dos sistemas autorizados.	ceitáveis e Alto Risco - e.g, Arts. 3º e 4º)	mento de risco - e.g, Capítulo IV sobre governança dos sistemas de IA)	autoridade para coordenação de tais esforços por parte do Poder Executivo - e.g, Art. 21 e Seção I do Capítulo VIII)	
Projeto de Convenção de Inteligência Artificial, Direitos Humanos, Democracia e Estado de Direito (rascunho) – CAI	O projeto (draft) de texto da Convenção desenvolvida pelo Comitê de Inteligência Artificial (CAI) do Conselho da Europa traz um artigo específico sobre a "abordagem baseada no risco", em que define que cada Estado-Membro "manterá e tomará medidas graduadas e diferenciadas em seu âmbito jurídico interno, conforme necessário e apropriado, tendo em vista a gravidade e a probabilidade de ocorrência de impactos adversos nos direitos humanos e nas liberdades fundamentais, na democracia e o Estado de direito durante a concepção, desenvolvimento, utilização e descontinuidade de sistemas de inteligência artificial".	Médio (Orienta a adoção de uma abordagem baseada no risco, apesar de não definir esses níveis - e.g, Art. XX)	Alto (previsão de discussão e consulta pública diversa e adequada - e.g. Art. 19)	Médio (e.g. Capítulo VII sobre mecanismos de follow-up e cooperação da aplicação da Convenção)	Médio (e.g. art. 12 sobre inovação segura)

	Somado a isso, o projeto determina que os Estados-Membros "devem tomar medidas para identificação, avaliação, prevenção e mitigação de riscos e impactos aos direitos humanos, democracia e o Estado de Direito decorrentes do projeto, desenvolvimento, uso ou descontinuação de sistemas de IA", levando em conta a abordagem baseada no risco.				
CAHAI	No estudo de viabilidade regulatória, o CAHAI explicita que os riscos vindos de sistemas de IA dependem do contexto da aplicação, da tecnologia e das partes interessadas envolvidas. Por isso, para combater qualquer sufocamento da inovação de IA e para garantir que os benefícios dessa tecnologia possam ser colhidos ao mesmo tempo em que enfrenta adequadamente seus riscos, o CAHAI recomenda que um futuro quadro jurídico criado pelo Conselho da Europa sobre IA siga uma abordagem baseada em risco. Além dela, o Comitê também salienta que, quando relevante, deve ser considerada uma abordagem preventiva, incluindo possíveis proibições. Assim, de acordo com o estudo uma estrutura legal abrangente para sistemas de IA, guiada por uma abordagem baseada em risco, pode ajudar a fornecer os contornos nos quais a inovação benéfica pode ser estimulada e aprimorada , e os benefícios da IA podem ser otimizados, garantindo – e maximizando – a proteção dos direitos humanos, da democracia e do estado de direito por meio de recursos legais eficazes.	Alto (estudo de viabilidade menciona a necessidade de uma abordagem baseada em risco com a definição de graus de riscos e possíveis proibições)	Alto (envolvimento das partes interessadas nas elaborações de avaliações de impacto em IA)	Alto (necessidade de autoridades nacionais de IA)	Médio (e.g. menciona medidas de compliance, como sandboxes, porém, em conjunto com avaliações de impacto)

Blueprint for an AI Bill of Rights	No tópico explicativo sobre a aplicação do Blueprint, ressalta-se que as medidas tomadas para concretizar a visão apresentada no documento devem ser proporcionais à extensão e a natureza do dano, ou risco de dano, aos direitos, oportunidades e acesso fruto de sistemas de IA.	Médio (não há taxonomia bem definida, mas prevê diferentes recomendações com base no risco do sistema de IA)	Alto (e.g. Princípio de Segurança e Efetividade dos Sistemas menciona a necessidade de consulta a diferentes atores)	Médio (documento é, a princípio, voluntário – mas torna-se obrigatório para órgãos do governo federal após a Ordem Executiva do presidente Biden de 30 de outubro de 2023)	Alto (documento se pretende ser voluntário para as organizações, além de prever medidas de suporte à inovação)
NIST – AI RMF 1.0	O Sumário Executivo do framework menciona que o gerenciamento dos riscos da IA é um componente essencial para o desenvolvimento e uso responsável da tecnologia. O objetivo central do AI RMF do NIST é ser uma fonte auxiliar de gerenciamento dos riscos para as organizações que desenvolvem, implementam e usam a IA. A ferramenta é, a princípio, voluntária, preocupada com a preservação de direitos, não específica de nenhum setor e "agnóstica" em termos de casos de uso, fornecendo flexibilidade para organizações de todos os tamanhos e de todos os setores, inclusive permitindo adaptação ao longo do desenvolvimento da tecnologia.	Médio (define que o AI RMF pode ser usado para priorizar o risco, mas não define a tolerância ao risco, que deve ser definida por cada uma organização de acordo com o quanto de risco está disposta a assumir, mas menciona a possibilidade de riscos: inaceitáveis, altos e baixos. - tópico 1.2.2)	Alto (e.g.tópico 5.2, página 25 - menciona a necessidade de incorporação de times internos diversos e o envolvimento de diferentes agentes externos, inclusive indivíduos e grupos potencialmente impactados pela tecnologia.	Baixo-Inexistente (trata-se de documento voluntário, a princípio)	Alto (o framework pretende ser aplicado voluntariamente em diferentes organizações de variados tamanhos e setores)
OCDE	A OCDE acompanha o desenvolvimento da governança de IA desde 2019, a partir da publicação de seus **Princípios** e a criação do **Observatório de IA**, com forte preocupação nos impactos econômicos e sociais dessa tecnologia. Possui diferentes relatórios e estudos sobre a regulação de IA, com destaque para o	Alto (já se manifestou a favor da abordagem baseada no risco para regular a IA, a fim de concentrar a supervi-	Alto (necessidade de monitoramento e participação dos atores interessados nos processos de prestação de contas de IA))	Alto (necessidade de autoridades de supervisão para monitoramento das políticas de IA)	Alto (diferentes princípios e boas práticas a serem implementados pelas organizações para uma IA de confiança)

OCDE	*framework* de classificação de sistemas de IA e o relatório sobre medidas de prestação de contas, além das recentes orientações para interoperabilidade na gestão de riscos de IA, que demonstram a tendência internacional de se regular sistemas de IA por meio da abordagem baseada em risco com a necessidade de ferramentas de *accountability*.	são e a intervenção onde são mais necessárias, evitando ao mesmo tempo obstáculos desnecessários à inovação))			
UNESCO	A UNESCO produziu o primeiro padrão global sobre ética na IA – a "Recomendação sobre a Ética da Inteligência Artificial", publicado em novembro de 2021, a partir de uma abordagem de direitos humanos. Este quadro foi adotado por todos os 193 Estados-Membros. A proteção dos direitos humanos e da dignidade é a pedra angular da Recomendação, baseada em princípios fundamentais como a transparência, equidade e supervisão humana dos sistemas de IA. O documento se baseia em quatro valores centrais: direitos humanos e dignidade; vida em uma sociedade justa, pacífica e interconectada; garantia de diversidade e inclusão; e proteção do meio-ambiente e ecossistema. Além disso, ele segue os movimentos recentes de enfatizar a necessidade de irmos além dos princípios éticos para a efetiva estratégia prática. Para isso, a Recomendação criar 11 áreas-chave para ações políticas ("actionable policies") e e fornecer duas metodologias práticas de (i) avaliação de impacto ético (EIA); (ii) avaliação de prontidão	Alto (e.g, Arts. 25 e 50-53 da Recomendação + Ferramenta de Avaliação de Impacto Ético traz 4 níveis de risco: muito alto, alto, médio e moderado/baixo)	Alto (e.g, Arts. 50-53 da Recomendação)	Alto (e.g, Capítulo V da Recomendação sobre monitoramento e avaliação)	Médio (e.g, Art. 69)

A partir dos documentos comparados, destaca-se o alinhamento do PL 2338/23 do Brasil às discussões internacionais, não apenas vindas do contexto europeu, mas de padrões globais, a exemplo da OCDE e UNESCO. Tal posição dá destaque ao modelo de regulação do risco definido por Kaminski (2022) na modalidade de supervisão democrática e alocação prévia de recursos, já que grande parte das propostas de governança de IA dão enfoque à necessidade de participação pública nos processos regulatórios, especialmente quando envolve a realização de avaliações de impacto, além de dar grande atenção à divisão dos esforços regulatórios de acordo com os riscos dos sistemas. Ainda, há a previsão de ferramentas regulatórias para destravar uma espécie de parceria público-privada no gerenciamento de risco, um meio do caminho entre regulação monopolizada pelo estado (comando e controle) ou somente pelo próprio agente econômico (autorregulação)[47].

a.2) Taxonomia de riscos

O modelo de regulação de riscos, de regulação assimétrica, intensifica tanto os recursos usados pelo regulador para fiscalização quanto às obrigações que as empresas devem cumprir em relação aos produtos ou serviços que apresentam um maior risco. Há, nesse modelo, uma alocação de riscos de maneira macro para determinadas empresas ou atividades, construindo-se faixas de risco, que podem variar de acordo com a metodologia escolhida, ser divididos entre "alto", "médio" ou "baixo", por exemplo.

Essa estratégia de regulação e designação macro de faixas de risco foi adotada por diversas normativas regulando sistemas de inteligência artificial: o PL 2338/23 no Brasil, o EU AI Act, o Proyecto de ley 15869/19 do Chile[48] e a ferramenta de avaliação de impacto do Canadá (Canada's Algorithmic Impact Assessment tool no âmbito da Diretiva sobre Tomada de Decisão Automatizada do Canadá).

No Brasil, especificamente no PL 2338/23, o risco é dividido em três faixas: excessivo, alto e moderado/baixo (esta é uma categoria residual, não explicitada na legislação). Não há uma definição de cada categoria do risco, uma vez que o risco é classificado a partir de um rol exemplificativo, com previsão de elementos quantitativos e qualitativos para atualização do rol de sistemas de risco inaceitável e alto pela autoridade competente, conforme artigo 18.

47 Para fins de análise da relação entre supervisão democrática e modelos de regulação, veja-se, entre outros: a) como o princípio da precaução é organizativo para fins de deliberação pública sobre quais são os riscos aceitáveis de uma atividade econômica ou tecnologia (BIONI, Bruno; LUCIANO, Maria. O Princípio da Precaução para a Regulação da Inteligência Artificial: Seriam as Leis de Proteção de Dados seu Portal de Entrada. *In*: Frazão, Ana. Mullholand, Caitlin. *Inteligência Artificial e Direito: ética, regulação e responsabilidade*. São Paulo: Revista dos Tribunais, 2019); b) a noção de co-deliberação informacional em complemento a de autodeterminação informacional (BIONI, Bruno Ricardo. *Regulação e Proteção de Dados Pessoais – O Princípio da Accountability*. São Paulo: Editora Forense, 2022. 320p).

48 https://www.camara.cl/verDOC.aspx?prmID=72777&prmTipo=FICHAPARLAMENTARIA&prmFICHATIPO=-DIP&prmLOCAL=0.

Na União Europeia, a proposta de regulamentação ("EU AI Act") divide o risco da mesma maneira que o PL 2338/23, em três faixas, entretanto, o nível mais alto de risco é denominado inaceitável. A faixa de risco mais baixa, (moderado ou limitado), assim como no PL, não é explicitada pela legislação, sendo uma categoria residual. Tal classificação também é seguida pelo projeto chileno 15869-19, que divide o risco em inaceitável e alto, além da categoria residual dos sistemas não classificados pelos dois níveis de risco.

Já a ferramenta de Avaliação de Impacto Algorítmico do Canadá, no âmbito da Diretiva canadense sobre Tomada de Decisão Automatizada, divide os impactos de automatizar uma decisão administrativa em 4 níveis, sendo que cada nível tem uma faixa percentual de impacto, a depender da reversibilidade das decisões automatizadas e da duração esperada da decisão tomada. Nesse sentido, decisões reversíveis e breves são de pouco impacto (nível I) e decisões irreversíveis e perpétuas são de muito alto impacto (nível IV).

Normativa	Níveis de risco	Nomenclatura	Obrigações relacionadas
PL 2338/23	3	Excessivo	Proibição (art. 14)
		Alto	Documentação, realização de testes de confiabilidade, adoção de medidas técnicas para viabilizar a explicabilidade dos resultados, entre outras (art. 20) + obrigações gerais (art. 19) + obrigação de elaboração de uma avaliação de impacto algorítmico (art. 22)
		Moderado/baixo (categoria residual)	Medidas de transparência, medidas de gestão de dados adequadas para mitigar e prevenir vieses discriminatórios, medidas de segurança da informação desde a concepção, dentre outras (artigo 19).
EU AI Act	3[49]	Inaceitável	Proibição (art. 5º)
		Alto	Obrigação de elaborar uma avaliação de impacto em direitos fundamentais (art. 29a), sistema de gestão de qualidade, elaboração de documentação técnica, manutenção de registros, sujeição ao procedimento de
EU AI Act	3		avaliação da conformidade, adoção de medidas corretivas, entre outros (capítulo 2 e 3 - artigos 8º em diante), dentre outros.
		Moderado	Obrigações de transparência bastante limitadas, por exemplo, no que diz respeito à prestação de informações para sinalizar a utilização de um sistema de IA quando este interage com seres humanos (título IV).

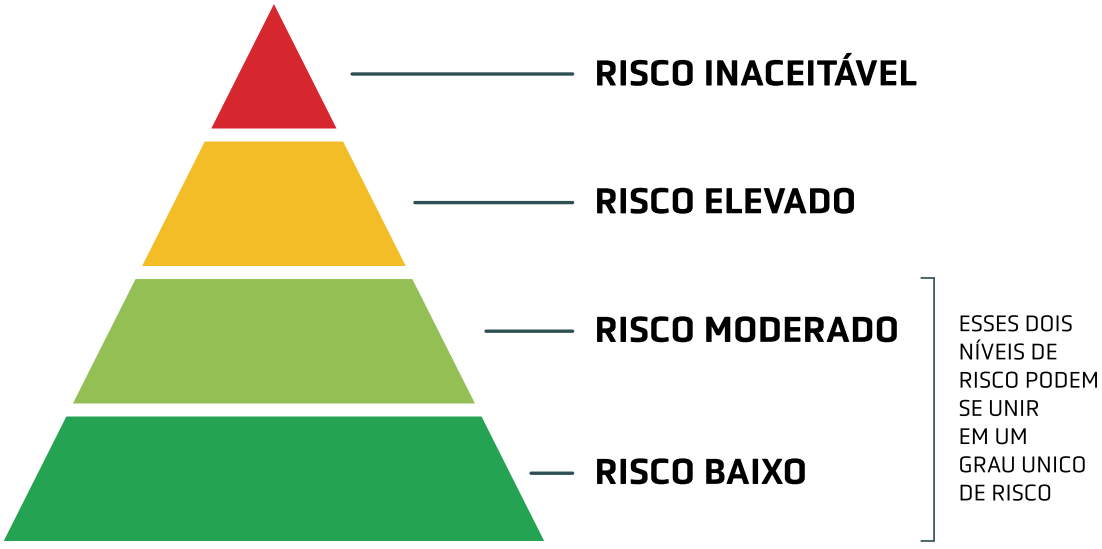
49 A última versão do texto, vinda do Parlamento Europeu em junho de 2023, criou obrigações específicas para os fornecedores de modelos de IA fundacionais no artigo 28b, para além da questão do risco associado.

Proyecto de ley 15869/19 (Chile)	3	Inaceitável	Proibição por meio de não autorização pela autoridade competente (art. 8).
		Alto	Obrigações prévias, como a implementação de um plano de gestão de riscos, plano de gestão de dados de entrada e plano de gestão de qualidade, manutenção de registros, fornecer informações, intervenção humana, entre outros (art. 9).
		Residual	Informar as pessoas que estão interagindo com um sistema de IA (art. 10) e informar à autoridade no caso de incidente grave ou defeito de funcionamento (art. 11).
Directive on Automated Decision-Making + Algorithmic Impact Assessment tool	4	Pouco a nenhum impacto (nível I)	Explicação significativa para resultados de decisões automatizadas.
		Impacto moderado (nível II)	*Peer review* (consulta a pelo menos dois especialistas e publicação do resumo dos resultados no site do Governo do Canadá); análise de gênero, aviso em linguagem simples publicado em todos os canais de prestação de serviços em uso; explicação significativa fornecida ao cliente em qualquer decisão que resulte na negação de um benefício ou serviço; documentação do design e funcionamento do sistema, entre outros.
		Impacto Alto (nível III)	*Peer review* (consulta a pelo menos dois especialistas e publicação dos resultados completos no site do Governo do Canadá); decisões não podem ser feitas sem intervenção humana; análise de gênero, aviso em linguagem simples publicado em todos os canais de prestação de serviços em uso; explicação significativa fornecida ao cliente em qualquer decisão que resulte na negação de um benefício ou serviço; documentação do design e funcionamento do sistema, operação depende de aprovação do *Deputy Head*, entre outros.
		Impacto muito alto (nível IV)	*Peer review* (consulta a pelo menos dois especialistas e publicação dos resultados completos no site do Governo do Canadá); decisões não podem ser feitas sem intervenção humana; análise de gênero, aviso em linguagem simples publicado em todos os canais de prestação de serviços em uso; explicação significativa fornecida ao cliente em qualquer decisão que resulte na negação de um benefício ou serviço; documentação do design e funcionamento do sistema, cursos de formação recorrentes (e um meio para verificar se o treinamento foi concluído); operação depende de aprovação do *Treasury Board*, entre outros.

Assim, constata-se que diferentes iniciativas regulatórias no âmbito internacional vindas da Europa, Canadá e América Latina aplicam a abordagem com base no risco que o divide em graus, níveis ou faixas para que os recursos regulatórios, assim como a carga obrigacional para os agentes regulados, sejam adequadamente distribuídos entre eles. No contexto brasileiro, isso é definido apenas no PL 2338/23, que faz a divisão dos graus de risco para alocação das obrigações legais, assim como onde a regulação em rede deve ser mais intensa, o que não é abordado nos demais projetos brasileiros atualmente em disputa no país.

a.3) Graus de risco

Como mencionado, por meio da taxonomia de riscos em uma regulação de risco assimétrica, há gradações de risco (níveis diferentes de mensuração de risco), que podem variar de acordo com o referencial utilizado. Os diferentes graus de risco darão ensejo a obrigações regulatórias mais ou menos fortes. A ideia é que não se crie restrições desnecessárias ao comércio, prestação de serviços ou inovação por meio de intervenção jurídica adaptada às situações concretas em que existe um motivo de preocupação justificado ou em que tal preocupação pode ser razoavelmente antecipada num futuro próximo[50].



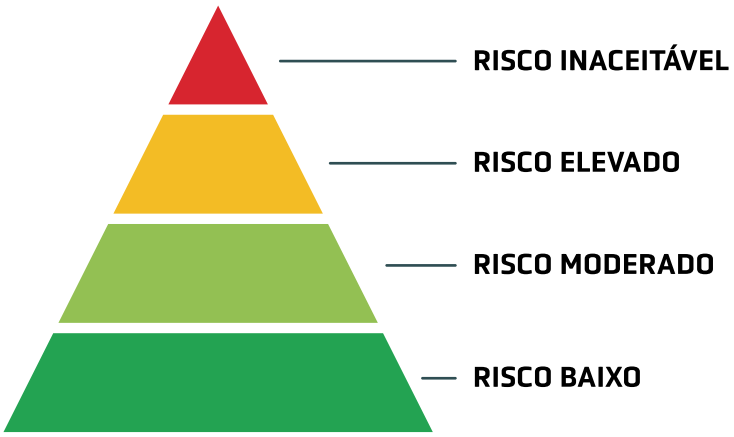
As gradações de risco (níveis diferentes de mensuração de risco) podem variar de acordo com o referencial utilizado

Assim, a gradação de riscos vai destravar pesos regulatórios distintos de acordo com o risco identificado para quem desenvolve a tecnologia e vai lançá-la no mercado ou

50 EU AI Act, exposição de motivos.

colocá-la à serviço. A intensidade da regulação será maior quanto maior o risco associado ao seu lançamento. Nesse ponto, é possível relacionarmos a gradação de riscos à ideia de uma pirâmide, em que a base é o risco mínimo/médio (sem grandes deveres), o meio é o alto risco (há muitos deveres para que a tecnologia seja permitida de ser implementada no mercado) e, por fim, o topo em que há riscos inaceitáveis, ou seja, em que a regulação vai impedir o uso da tecnologia porque ela traz mais riscos do que benefícios para direitos, democracia e sociedade.

Porém, além da definição dos diferentes níveis de risco, é essencial que sejam previstos os elementos qualitativos para definição de cada um desses riscos. Em outras palavras, ao invés de apenas definir os graus de risco (por exemplo, baixo/médio/alto risco) de forma generalista, é indispensável ter critérios mínimos para identificação dos sistemas em cada um desses níveis. A título exemplificativo:



As gradações de risco (níveis diferentes de mensuração de risco) podem variar de acordo com o referencial utilizado

RISCO INACEITÁVEL
Quando há riscos "significativos" ou desconhecidos para os direitos humanos, democracia ou estado de direito + não existem medidas de mitigação apropriadas dentro das estruturas legais existentes para mitigar adequadamente

RISCO ELEVADO
Quando há riscos "significativos" ou desconhecidos para os direitos humanos, democracia ou estado de direito + existem medidas de mitigação apropriadas dentro das estruturas legais existentes que podem auxiliar na mitigação adequada

RISCO MODERADO
Quando, apesar de existir riscos, eles não são significativos para os direitos humanos, democracia ou estado de direito + existem medidas de mitigação apropriadas dentro das estruturas legais existentes que auxiliam na mitigação

RISCO BAIXO
Não é possível a presença de zero risco, mas os riscos existentes não são significativos para os direitos humanos, democracia ou estado de direito, sendo apenas residuais, não demandando requisitos adicionais para seu uso

Por exemplo, pensar no contexto de aplicação da tecnologia de IA em específico é um elemento essencial, pois é a partir do contexto de aplicação que é possível uma análise mais granular para parametrização de risco. Além do contexto, outros elementos podem ser utilizados como critério para definição do grau de risco, a exemplo do escopo, explicabilidade, quantidade de dados processados, nível de automação, dentre outros, o que deve ser minimamente explicitado na regulação.

a.3.1) *Risco Inaceitável/Excessivo*

De acordo com Mantelero (2022), sempre que uma nova aplicação de tecnologia possa produzir riscos potenciais graves para os indivíduos e a sociedade, que não podem

ser calculados ou quantificados com precisão e antecedência, uma abordagem de precaução deve ser adotada. Nesses casos, a adoção de mecanismos de governança, como a elaboração de uma avaliação de impacto adequada, por exemplo, é impossível, mas o impacto potencialmente alto na sociedade justifica medidas de precaução específicas (por exemplo, proibição ou restrição do uso da tecnologia). É nesse grau de risco que se encontra o nível mais interventivo e regulatório, já que se define a proibição de utilização e desenvolvimento dos sistemas de forma *ex-ante*.

Sistemas de inteligência artificial de risco inaceitável e excessivo motivam a intervenção regulatória mais forte do EU AI Act, do Projeto de lei chileno 15869/19 e do PL 2338/23, respectivamente, resultando na proibição desses sistemas de maneira prévia (ex ante). Tal nível pode ser entendido como parte da conciliação entre uma abordagem baseada em direitos e riscos. Ou seja, existem determinados direitos que são inegociáveis e certas aplicações da inteligência artificial gerariam riscos intoleráveis.

Um exemplo dessa proibição são os sistemas de crédito social (social scoring), que condicionam o acesso a bens,serviços e políticas públicas a uma avaliação do indivíduo com base em seu comportamento social ou características da sua personalidade.

A ideia de sistemas de crédito social foi divulgada mundialmente a partir de experiências na China. Em 2014, o Governo Central Chinês anunciou um plano de seis anos para estabelecer um sistema de crédito social ("social credit system"), em que ações que criam confiança na sociedade seriam recompensadas, e punidas as que vão em sentido contrário[51]. O termo crédito social abrange não apenas o que é visto tradicionalmente como score de crédito, ou seja, o histórico financeiro de indivíduos e empresas e uma previsão de se pagarão futuros empréstimos, como também o score social ("social creditworthiness"), relacionado a confiabilidade de um indivíduo a partir de atividades não financeiras[52]. Em escala nacional, o que existe por enquanto é um sistema focado em empresas, que agrega dados sobre cumprimento de regulações de diferentes agências governamentais, disponibilizado em um site chamado "Credit China"[53]. Apesar do foco em empresas, no site existem informações de indivíduos e outras organizações, reunindo bases de dados variadas, mas não sistematizadas, com informações como, por exemplo, quais indivíduos descumpriram medidas judiciais, quais universidades chinesas são legítimas, entre outros[54].

Os exemplos de sistemas mais desenvolvidos de crédito social vêm de governos

51 YANG, Zeyi. China just announced a new social credit law. Here's what it means. MIT Technology Review, publicado em 22 nov. 2022. Disponível em: https://www.technologyreview.com/2022/11/22/1063605/china-announced-a--new-social-credit-law-what-does-it-mean/.

52 Ibid.

53 Ibid.

54 Ibid.

locais que implementaram programas pilotos[55]. Na cidade de Rongcheng, com meio milhão de habitantes, foi implementado em 2013 um sistema que dava como base de crédito social para cada cidadão 1000 pontos, sendo o número de pontos influenciado por ações individuais, como espalhar informação maliciosa em redes de comunicação social, que reduzia 50 pontos, ou ganhar uma competição de esportes ou cultura de alcance nacional, que adicionaria 40 pontos[56]. Esses programas se mantiveram restritos a cidades, não atingindo províncias inteiras ou o país[57]. Inclusive, em Dezembro de 2020, em um guia publicado pelo Conselho de Estado Chinês, foi recomendado que governos locais só punissem comportamentos que já são ilegais na legislação Chinesa. Voltando ao exemplo da Cidade de Rongcheng, a regulação de crédito social foi atualizada para permitir que os cidadãos saiam do programa se desejarem e houve modificação de alguns critérios[58].

Outro exemplo relevante é o caso do uso de sistemas de identificação biométrica à distância em espaços públicos, como é o caso do reconhecimento facial no âmbito da segurança pública. Diferentes estudos[59] já comprovaram que, no atual estágio de desenvolvimento, esses sistemas apresentam imprecisões de falsos positivos e negativos, principalmente contra grupos já marginalizados e vulneráveis, especialmente quando analisado sob lentes de interseccionalidade. Logo, a implementação e uso desta tecnologia para fins de segurança pública pelo Estado, principalmente quando aplicado de forma massiva em tempo real para identificação e rastreamento pode interferir negativamente em diferentes direitos fundamentais, inclusive reforçando discriminações estruturais.

Nesse contexto, fala-se da necessidade de banimento ou de uma moratória para o desenvolvimento e utilização de sistemas de reconhecimento facial na segurança pública pelo Estado. No caso do banimento, defende-se uma total proibição de utilização desses sistemas por entender que seus benefícios não superam os malefícios trazidos pela violação de direitos e valores inegociáveis, como a não discriminação. Já no caso da moratória há uma proibição por determinado período ou diante de certas circunstâncias, até que a tecnologia evolua ou mecanismos de governança eficientes sejam desenvolvidos para que direitos inegociáveis não sejam violados por tais sistemas.

55 Ibid.

56 Ibid.

57 Ibid.

58 Ibid.

59 Buolamwini, Joy; Gebru, Timnit. (2018) Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Conference on Fairness, Accountability and Transparency. Proceedings of Machine Learning Research 81:1–15, 2018. Disponível em: <http://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf>; COSTANZA-CHOCK, Sasha. Design Justice, A.I., and escape from the matrix of domination. Cambridge: Journal of Design and Science, Jul. 2018. DOI:10.21428/96c8d426. Disponível em: https://jods.mitpress.mit. edu/pub/costanza-chock/release/4; VARON, Joana; SILVA, Mariah Rafaela. Reconhecimento facial no setor público e identidades trans: tecnopolíticas de controle e ameaça à diversidade de gênero em suas interseccionalidades de raça, classe e território. Disponível em: <https://codingrights.org/docs/rec-facial-id-trans.pdf>.

O PL 2338/23, na seção de risco excessivo, traz como uma das suas hipóteses o uso de sistemas de identificação biométrica à distância em atividades de segurança pública.

Na prática, o que o **artigo cria é uma moratória**, condicionando o uso de tais sistemas à dois fatores: (i) promulgação de lei federal específica, (ii) autorização judicial de utilização, que deve estar conectada à atividade de persecução penal individualizada, para crimes passíveis de pena máxima de reclusão superior a dois anos, busca de vítimas de crimes ou pessoas desaparecidas ou crime em flagrante[60]. A lei federal deve prever medidas proporcionais e estritamente necessárias ao atendimento do interesse público, além da necessidade de revisão por agente público responsável pela inferência algorítmica antes da tomada de ação relativa à pessoa identificada.

O projeto de lei chileno também traz esta hipótese de moratória ao uso de sistemas de identificação biométrica à distância em espaços de acesso público nos casos que esta utilização seja considerada estritamente necessária para: (i) busca de possíveis vítimas específicas de um crime, incluindo menores desaparecidos; (ii) prevenção de ameaça específica, significativa e iminente à vida ou à segurança física das pessoas ou de um ataque terrorista; (iii) detecção, localização, identificação ou acusação da pessoa que cometeu, ou é suspeita de ter cometido, algum dos crimes previstos no Código Penal. Nestes três casos excepcionais, o projeto também determina que sejam sempre sujeitas à decisão prévia proferida por um Tribunal de Justiça e apenas aplicadas pelos Carabineros do Chile (uma espécie de polícia ostensiva) e pela Polícia Investigativa.

Na regulamentação europeia, existem duas posições diferentes sobre a regulamentação da identificação biométrica. A posição do Parlamento Europeu proíbe o uso de sistemas de identificação biométrica remota em tempo real em espaços públicos (por entes públicos ou privados), assim como o uso de sistemas para análise de gravações de espaços públicos com identificação biométrica remota. Há exceção para o uso retroativo, ou seja, de gravações para identificação biométrica remota se houver autorização judicial prévia para o uso, que deve se dar no contexto da persecução penal, quando estritamente necessário, e ser relativo a um crime sério que já ocorreu[61].

Já o Conselho da União Europeia, composto pelos Estados Membros, enfraqueceu a proibição do uso em tempo real dos sistemas de identificação biométrica remota, já que

60 A partir desta redação da proposta de lei, o Brasil proibiria a prática do Predictive Policing, isto é, a utilização de algoritmos para análise de grandes bancos de dados para prever informações relacionadas a crimes, como quando e onde ele acontecerá no futuro ou quem é mais provável de cometê-lo, e, a partir desta informação, tomar a decisão de onde alocar maior contingente policial. Isso porque o uso de sistemas de identificação biométrica à distância em atividades de segurança pública só poderia existir nos casos listados nos incisos do art. 15, não mais analisando o público geral como suspeitos, uma vez que as análises seriam apenas possíveis no contexto de uma persecução penal individualizada. ACLU of Washington. How Automated Decision Systems are used in Policing. Publicado em 26 dez. 2022. Disponível em: https://www.aclu-wa.org/story/how-automated-decision-systems-are-used-policing.

61 Emenda 41 (Amendment 41) – https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf.

defende a permissão do uso em situações excepcionais, listada de maneira exaustiva, nas quais o interesse público prevalecerá sobre os riscos. Exemplos de tais situações são a busca de vítimas potenciais de um crime, incluindo crianças desaparecidas; algumas ameaças à vida ou segurança física de pessoas naturais ou de um ataque terrorista; e a detecção, localização, identificação ou persecução criminal de suspeitos dos 32 crimes listados na decisão do Conselho (Council Framework Decision 2002/584/JHA) se os crimes forem puníveis nos Estados Membro por uma sentença de custódia ou detenção por um período máximo de ao menos três anos. Além disso, na proposta do Conselho, também é permitida a utilização de tais sistemas por forças policiais de fronteira, imigração ou asilo para identificar uma pessoa que se recusa a ser identificada ou não consegue provar sua identidade[62].

Alguns documentos internacionais de referência não especificam claramente quais os casos de IA de risco excessivo, mas preveem que, mediante certas condições, alguns sistemas de IA devem ser submetidos a moratória ou proibição prévias. No projeto de Convenção sobre Inteligência Artificial, Direitos Humanos, Democracia e Estado de Direito do Comitê de Inteligência Artificial (CAI) do Conselho da Europa, cria-se uma obrigação para que os Estados-parte tomem medidas legislativas necessárias para criar moratória ou proibição de certos sistemas de IA sempre que considerados incompatíveis com o respeito dos direitos humanos, o funcionamento da democracia e do Estado de direito (art. 15(3)). Já a Recomendação sobre a Ética da Inteligência Artificial da UNESCO menciona a proibição de sistemas de IA que tenham efeitos negativos desproporcionais em impactos ambientais (art. 86), assim como daqueles que tenham o poder de tomar decisões de vida ou morte (art. 36), além da clara menção à recomendação de não utilização de IA para fins de vigilância em massa e crédito social (art. 26).

62 Parágrafo 19 e seguintes – https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf.

Normativa	Previsão de riscos proibidos de forma *ex-ante*	Quantidade de situações proibidas *ex-post*	Uso de dados biométricos em IA para fins de persecução penal
PL 2338/23	Sim	4	Moratória
EU AI Act (versão do PE)	Sim	8	Banimento
EU AI Act (versão do Conselho da União Europeia)	Sim	4	Moratória
Proyecto de ley 15869/19 (Chile)	Sim	4	Moratória
CAI	Sim	Não há definição de quantidade	Não menciona especificamente
UNESCO	Sim	Não há definição de quantidade	Não menciona especificamente

NORMATIVAS ANALISADAS Hipóteses de risco excessivo/inaceitável				
	PL 2338/2023	EU AI ACT (Conselho)	EU AI ACT (PE)	PL 15869-19 (Chile)
Definição do risco/ consequência (proibição)	Art. 14. São vedadas a implementação e o uso de sistemas de inteligência artificial:	Art. 5º. 1. São proibidas as seguintes práticas de inteligência artificial:	Art. 5º. 1. São proibidas as seguintes práticas de inteligência artificial:	Artigo 3º. Serão classificados como sistemas de IA de risco inaceitável: Artigo 8. A Comissão não autorizará o desenvolvimento, distribuição, comercialização ou utilização de sistemas de IA cujo risco seja inaceitável
Técnicas de indução de comportamentos que possa causar danos (físicos e psicológicos)	I – que empreguem técnicas subliminares que tenham por objetivo ou por efeito induzir a pessoa natural a se comportar de forma prejudicial ou perigosa à sua saúde ou segurança ou contra os fundamentos desta Lei;	a) A colocação no mercado, a entrada em serviço ou a utilização de um sistema de IA que utilize técnicas subliminares para além da consciência de uma pessoa com o objetivo ou o efeito de distorcer materialmente seu comportamento de forma a causar ou que tenha probabilidade razoável de causar danos físicos ou psicológicos a essa pessoa ou a outra pessoa;	a) A colocação no mercado, a entrada em serviço ou a utilização de um sistema de IA que utilize técnicas subliminares que vão além da consciência de uma pessoa ou técnicas intencionalmente manipuladoras ou enganosas, com o objetivo ou o efeito de distorcer materialmente a situação de uma pessoa ou de um grupo de pessoas, prejudicando sensivelmente a sua capacidade de tomar uma decisão informada, fazendo com que ela tome uma decisão que de outra forma não teria tomado, causando ou sendo suscetível de causar a essa pessoa, outra pessoa ou grupo de pessoas danos significativos. A proibição de sistemas de IA que utilizem técnicas subliminares a	1. Aqueles que utilizam técnicas subliminares que transcendem a consciência de uma pessoa para alterar substancialmente seu comportamento de forma que cause ou possa causar danos físicos ou mentais àquela pessoa ou outra.

			que se refere o primeiro parágrafo não se aplica a sistemas de IA destinados a serem utilizados para fins terapêuticos aprovados com base no consentimento informado específico das pessoas a eles expostas ou, se for caso disso, de seu responsável legal;	
Técnicas que exploram vulnerabilidades	II – que explorem quaisquer vulnerabilidades de grupos específicos de pessoas naturais, tais como as associadas a sua idade ou deficiência física ou mental, de modo a induzi-las a se comportar de forma prejudicial a sua saúde ou segurança ou contra os fundamentos desta Lei;	b) A colocação no mercado, a entrada em serviço ou a utilização de um sistema de IA que explore qualquer uma das vulnerabilidades de um grupo específico de pessoas devido à sua idade, deficiência ou a uma situação social ou econômica específica, com o objetivo ou a efeito de distorcer materialmente o comportamento de uma pessoa pertencente a esse grupo de maneira a causar ou tenha probabilidade razoável de causar danos físicos ou psicológicos a essa pessoa ou a outra pessoa;	b) A colocação no mercado, a entrada em serviço ou a utilização de um sistema de IA que explore qualquer uma das vulnerabilidades de uma pessoa ou de um grupo específico de pessoas, incluindo características dos traços de personalidade conhecidos ou previstos dessa pessoa ou de um grupo ou características sociais ou situação econômica, idade, capacidade física ou mental com o objetivo ou o efeito de distorcer materialmente o comportamento dessa pessoa ou de uma pessoa pertencente a esse grupo de uma maneira que cause ou possa causar a essa pessoa ou a outra pessoa danos significativos;	2. Qualquer sistema de IA que se aproveite de alguma das vulnerabilidades de uma pessoa ou de um determinado grupo de pessoas devido à sua idade ou deficiência física ou mental para alterar substancialmente o comportamento de uma pessoa pertencente a esse grupo de uma forma que cause ou possa causar dano físico ou psicológico a essa pessoa ou a outra.
Categorização biométrica para classificar pessoas de acordo com características sensíveis ou protegidas	Sem correspondência.	Sem correspondência.	b) A colocação no mercado, a entrada em serviço ou a utilização de sistemas de categorização biométrica que classificam pessoas de acordo com atributos ou características sensíveis ou protegidos ou com base na inferência desses atri-	Sem correspondência.

			butos ou características. Esta proibição não se aplica aos sistemas de IA destinados a serem utilizados para fins terapêuticos aprovados com base no consentimento informado específico das pessoas a eles expostas ou, se for caso disso, do seu tutor legal.	
Score social	III – pelo poder público, para avaliar, classificar ou ranquear as pessoas naturais, com base no seu comportamento social ou em atributos da sua personalidade, por meio de pontuação universal, para o acesso a bens e serviços e políticas públicas, de forma ilegítima ou desproporcional.	c) A colocação no mercado, a entrada em serviço ou a utilização de sistemas de IA para a avaliação ou classificação de pessoas durante um determinado período de tempo com base no seu comportamento social ou em características pessoais ou de personalidade conhecidas ou previstas, com a pontuação social a liderar a um ou ambos os seguintes: (i) tratamento prejudicial ou desfavorável de certas pessoas ou grupos em contextos sociais que não estão relacionados com os contextos em que os dados foram originalmente gerados ou recolhidos; (ii) tratamento prejudicial ou desfavorável de certas pessoas ou grupos que seja injustificado ou desproporcional ao seu comportamento social ou à sua gravidade;	c) A colocação no mercado, a entrada em serviço ou a utilização de sistemas de IA para a avaliação ou classificação de pontuação social de pessoas ou grupos de pessoas singulares durante um determinado período de tempo com base no seu comportamento social ou características pessoais ou de personalidade conhecidas, inferidas ou previstas, com a pontuação social levando a uma ou ambas as consequências: (i) tratamento prejudicial ou desfavorável de certas pessoas ou grupos inteiros em contextos sociais que não estão relacionados aos contextos em que os dados foram originalmente gerados ou coletados; (ii) tratamento prejudicial ou desfavorável de certas pessoas ou de grupos inteiros delas, que seja injustificado ou desproporcional ao seu comportamento social ou à sua gravidade;	3. Aquela utilizada pelas autoridades públicas ou em seu nome para avaliar ou classificar a confiabilidade das pessoas físicas durante um determinado período de tempo com base em seu comportamento social ou em características pessoais ou de personalidade conhecidas ou previstas, de modo que a classificação social resultante em uma ou mais das seguintes situações: a. Tratamento prejudicial ou desfavorável a determinadas pessoas ou grupos em contextos sociais que não estão relacionados com os contextos onde os dados foram originalmente gerados ou coletados. b. Tratamento prejudicial ou desfavorável a determinadas pessoas ou grupos, de forma injustificada ou desproporcional ao seu comportamento social ou à sua gravidade.

Identificação biométrica à distância em tempo real (contínua)	Art. 15. No âmbito de atividades de segurança pública, somente é permitido o uso de sistemas de identificação biométrica à distância, de forma contínua em espaços acessíveis ao público, quando houver previsão em lei federal específica e autorização judicial em conexão com a atividade de persecução penal individualizada, nos seguintes casos: I – persecução de crimes passíveis de pena máxima de reclusão superior a dois anos; II – busca de vítimas de crimes ou pessoas desaparecidas; ou III – crime em flagrante.	d) A utilização de sistemas de identificação biométrica remota em tempo real em espaços acessíveis ao público pelas autoridades responsáveis pela aplicação da lei ou em seu nome para efeitos de aplicação da lei, a menos que e na medida em que tal utilização seja estritamente necessária para um dos seguintes objetivos: (i) a procura direcionada de potenciais vítimas específicas de crimes; (ii) a prevenção de uma ameaça específica e substancial às infraestruturas críticas, à vida, à saúde ou à segurança física de pessoas ou à prevenção de ataques terroristas; (iii) a localização ou identificação de uma pessoa singular para efeitos de investigação criminal, ação penal ou execução de sanções penais (...) e puníveis no Estado-Membro abrangido por uma pena ou medida de segurança privativas da liberdade por um período máximo de pelo menos três anos, ou outras infrações específicas puníveis no Estado-Membro em questão com uma pena ou medida de segurança privativas da liberdade por um período máximo de pelo menos cinco anos, como determinado pela legislação desse Estado-Membro.	d) A utilização de sistemas de identificação biométrica remota em tempo real em espaços acessíveis ao público;	4. A de identificação biométrica remota, em tempo real ou diferida em espaços de acesso público, salvo e na medida em que essa utilização seja estritamente necessária para atingir um ou mais dos seguintes objetivos: a. A busca seletiva de possíveis vítimas específicas de um crime, incluindo menores desaparecidos. b. A prevenção de uma ameaça específica, significativa e iminente à vida ou à segurança física das pessoas ou de um ataque terrorista. c. A detecção, localização, identificação ou acusação da pessoa que cometeu, ou é suspeita de ter cometido, algum dos crimes previstos no Código Penal. As exceções consideradas no parágrafo 4 deste artigo estarão sujeitas a ordem emitida por um Tribunal de Justiça e só poderão ser aplicadas pelos Carabineros do Chile e pela Polícia Investigativa.

Avaliação de risco de cometimento de crimes ou reincidência	Sem correspondência.	Sem correspondência.	d-A) A colocação no mercado, a entrada em serviço ou a utilização de um sistema de IA para efetuar avaliações de risco de pessoas ou grupos de pessoas, a fim de avaliar o seu risco de cometer uma infração ou reincidência ou para prever a ocorrência ou reincidência de uma infração penal ou administrativa real ou potencial baseada na definição de perfil de uma pessoa ou na avaliação de traços e características de personalidade, incluindo a localização da pessoa, ou comportamento criminoso passado de pessoas ou grupos de pessoas;	Sem correspondência.
Criação ou expansão de bases de dados de reconhecimento facial	Sem correspondência.	Sem correspondência.	dB) A colocação no mercado, a entrada em serviço ou a utilização de sistemas de IA que criam ou expandem bases de dados de reconhecimento facial através da coleta não direcionada de imagens faciais da Internet ou de imagens CCTV;	Sem correspondência.
Inferência de emoções para determinados contextos	Sem correspondência.	Sem correspondência.	d-C) A colocação no mercado, a entrada em serviço ou a utilização de sistemas de IA para inferir emoções de uma pessoa nos domínios da aplicação da lei, da gestão das fronteiras, no local de trabalho e nas instituições de ensino.	Sem correspondência.

Sistemas de identificação biométrica remota posterior para análise de imagens gravadas de locais públicos	Sem correspondência.	Sem correspondência.	dd) A entrada em serviço ou a utilização de sistemas de IA para a análise de imagens gravadas de espaços acessíveis ao público através de sistemas de identificação biométrica remota posterior, a menos que estejam sujeitos a uma autorização pré-judicial nos termos do direito da União e sejam estritamente necessários para a pesquisa direcionada relacionada com uma infração penal grave específica, tal como definida no artigo 83.º, n.º 1, do TFUE, que já tenha sido realizada para efeitos de aplicação da lei.	Sem correspondência.

Diante do exposto, percebe-se a convergência quanto à proibição *ex-ante* de certos usos de IA que apresentam riscos potenciais tão graves que deflagram uma abordagem de precaução e, consequentemente, a intervenção regulatória mais intensa. Porém, há nuances quanto a alguns dos usos inaceitáveis, como no caso de sistemas biométricos de identificação, especialmente o reconhecimento facial, utilizados para fins de persecução penal. Nessa hipótese, ainda não há consenso a respeito da proibição prévia total, como propõe o Parlamento Europeu, ou se aplicar-se-á uma moratória que permita a sua utilização apenas em casos excepcionais, definidos pela lei.

a.3.2) *Risco Alto/Elevado*

O grau de risco elevado/alto dispara nível interventivo regulatório mais significativo, mas não proibitivo, já que autoriza a utilização da tecnologia mediante o cumprimento de algumas obrigações. Em regra, a técnica de classificação de risco ocorre de maneira bifurcada, por meio da criação de uma lista exemplificativa com rotulação de exemplos e/ou pelo estabelecimento de critérios quantitativos e qualitativos para que outras atividades sejam enquadradas como tal.

No caso do PL 2338/23, assim como na proposta de regulamentação europeia, na proposta de Lei de Inteligência Artificial e Dados canadense (AIDA) e no projeto de lei 15869/19 chileno, o nível de risco alto indica sistemas de inteligência artificial nos quais haverá intervenção regulatória significativa, mas que não são proibidos, conforme ilustrado pela pirâmide dos riscos.

Porém, não há em nenhuma destas normativas uma definição do que é um sistema de inteligência artificial de alto risco, entretanto, as propostas trazem exemplos de tais sistemas. No caso do Marco Legal de IA estipulado pelo PL 2338/23, a categorização dos riscos dos sistemas de IA se dá pela avaliação preliminar, conforme definido no artigo 13. O art. 17 deste projeto lista alguns sistemas de IA considerados de alto risco, a partir de suas finalidades, conforme tabela abaixo. Essa lista tem natureza exemplificativa e não exaustiva das hipóteses de sistemas de alto risco. Para classificar um sistema como de alto risco, a normativa se utiliza de dois caminhos: (i) se uma das finalidades do sistema está listada no artigo 17; e (ii) por meio de análise a partir de critérios quantitativos e qualitativos expostos no artigo 18, que trata da atualização da lista dos sistemas de IA de alto risco pela autoridade competente.

Já as hipóteses de sistemas de alto risco na proposta de regulamento europeia (versão do Parlamento Europeu de junho de 2023) estão no artigo 6º, complementado pelo Anexo III, identificando duas categorias principais de sistemas de IA de alto risco: (i) sistemas de IA destinados a serem utilizados como componentes de segurança de produtos sujeitos a avaliação de conformidade ex ante por terceiros; (ii) outros sistemas de IA autônomos com implicações principalmente nos direitos fundamentais, explicitamente enumerados pela sua área de atuação no anexo III. Diferentemente da proposta original da Comissão Europeia, tais opções são consideradas de alto risco, pela versão do Parlamento Europeu, desde que haja o cumprimento de requisito adicional: a existência de risco significativo de danos para a saúde, segurança ou direitos fundamentais das pessoas, o que seria definido pela Comissão Europeia pelo menos seis meses antes da entrada em vigor do regulamento, após consulta pública à Autoridade de IA e demais partes interessadas.

Assim como no PL 2338/23, a proposta europeia também prevê a possibilidade de

atualização da lista de casos de uso de IA de alto risco, já que apresenta número limitado de sistemas de IA cujos riscos já se concretizaram ou são suscetíveis de se materializar em um futuro próximo. Nesse sentido, para garantir a constante atualização do regulamento, de acordo com o art. 7º, a Comissão Europeia pode aumentar, modificar ou remover hipóteses, a partir de alguns critérios qualitativos e quantitativos. Já a proposta canadense (AIDA), que ainda está em processo de elaboração, os critérios para designação de sistemas de alto risco serão definidos na regulação, que deve estar alinhada à ideia de interoperabilidade com outros regramentos internacionais sobre IA em evolução, a exemplo do EU AI Act, dos Princípios de IA da Organização de Cooperação e Desenvolvimento Económico (OCDE) e da Estrutura de Gestão de Risco (RMF) do Instituto Nacional de Padrões e Tecnologia dos EUA (NIST). O documento complementar à proposta traz os principais fatores de análise para determinação se um sistema de IA é de alto risco, além de salientar a importância de se atentar às capacidades e aos contextos de utilização de sistemas de IA para designação do grau de risco e, por isso, trazem uma lista com exemplos de sistemas que são de interesse do governo canadense em termos dos seus potenciais altos impactos.

Ainda no contexto canadense, a Diretiva sobre Decisões Automatizadas, apesar de não prever expressamente a categoria de alto risco, estabelece o nível de impacto III que pode ser associado à essa categoria, já que está relacionado a decisões automatizadas que frequentemente conduzirão a impactos que podem ser difíceis de reverter e são contínuos. Nesse âmbito, a diretiva, em seu anexo B, prevê que este nível é vinculado a decisão que provavelmente terá grandes impactos em direitos dos indivíduos ou comunidades; a igualdade, dignidade, privacidade e autonomia dos indivíduos; a saúde ou o bem-estar de indivíduos ou comunidades; os interesses econômicos de indivíduos, entidades ou comunidades; e a sustentabilidade contínua de um ecossistema.

O projeto de lei 15869/19 do Chile também traz um rol exemplificativo de sistemas de IA de alto risco, similar ao EU AI Act, pois associa o nível de risco ao contexto de utilização e prevê a inclusão de mais hipóteses a esse rol nos casos que envolvem riscos de prejuízo à saúde, segurança ou repercussões negativas em direitos fundamentais. Um ponto de diferenciação da iniciativa do Chile, porém, é a obrigação de que todos os desenvolvedores, fornecedores e usuários de sistemas de IA solicitem autorização da Comissão Nacional de IA antes do início do desenvolvimento, comercialização, distribuição e utilização destes em território chileno, o que faz com que o nível de risco e o cumprimento das obrigações seja avaliado por essa Comissão em momento prévio e, mesmo após aprovado, caso o sistema passe por modificações substanciais. Nas demais propostas, a avaliação e supervisão dos riscos pelas autoridades ocorre em momento posterior.

Normativa	Lista de exemplos de atividades de alto risco
PL 2338/23	Sistemas de IA para as finalidades de: I – aplicação como dispositivos de segurança na gestão e no funcionamento de infraestruturas críticas; II – educação e formação profissional; III – recrutamento, triagem, filtragem, avaliação de candidatos, tomada de decisões sobre promoções ou cessações de relações contratuais de trabalho; IV – avaliação de critérios de acesso, elegibilidade, concessão, revisão, redução ou revogação de serviços privados e públicos que sejam considerados essenciais; V – avaliação da capacidade de endividamento das pessoas naturais ou estabelecimento de sua classificação de crédito; VI – envio ou estabelecimento de prioridades para serviços de resposta a emergências, incluindo bombeiros e assistência médica; VII – administração da justiça, incluindo sistemas que auxiliem autoridades judiciárias na investigação dos fatos e na aplicação da lei; VIII – veículos autônomos; IX – aplicações na área da saúde, inclusive as destinadas a auxiliar diagnósticos e procedimentos médicos; X – sistemas biométricos de identificação; XI – investigação criminal e segurança pública; XII – estudo analítico de crimes relativos a pessoas naturais; XIII – investigação por autoridades administrativas para avaliar a credibilidade dos elementos de prova no decurso da investigação ou repressão de infrações, para prever a ocorrência ou a recorrência de uma infração real ou potencial com base na definição de perfis de pessoas singulares; ou XIV – gestão da migração e controle de fronteiras.
EU AI Act (Comissão Europeia)	Sistema de IA destinado à utilização como componente de segurança de um produto ou de uma outra IA, além dos sistemas utilizados em um dos seguintes domínios (anexo III): I - Identificação biométrica e categorização de pessoas naturais; II - gestão e funcionamento de infraestruturas críticas; III- educação e formação profissional; IV- emprego, gestão de trabalhadores e acesso ao emprego por conta própria; V- Acesso a serviços privados e a serviços e prestações públicas essenciais, bem como o usufruto dos mesmos; VI- Manutenção da ordem pública; VII- Gestão da migração, do asilo e do controle das fronteiras; VIII- Administração da justiça e processos democráticos.
EU AI Act (Parlamento Europeu)	Sistema de IA destinado à utilização como componente de segurança de um produto ou de uma outra IA, além dos sistemas utilizados em um dos domínios abaixo (anexo III), desde que cumpram o requisito de representar risco significativo de dano para a saúde, segurança ou direitos fundamentais das pessoas ou para o meio ambiente:

EU AI Act (Parlamento Europeu)	I- Biometria e sistemas baseados em biometria: (a) sistemas utilizados para identificação biométrica, com exceção dos casos proibidos pelo art. 5º; (b) sistemas utilizados para fazer inferências sobre características pessoais com base em dados biométricos, incluindo reconhecimento de emoções (com exceção dos casos proibidos); II - gestão e funcionamento de infraestruturas críticas; III - educação e formação profissional; IV- emprego, gestão de trabalhadores e acesso ao emprego por conta própria; V- Acesso a serviços privados e a serviços e prestações públicas essenciais, como assistência médica, habitação, electricidade, aquecimento/arrefecimento e internet; VI - Manutenção da ordem pública; VII- Gestão da migração, do asilo e do controle das fronteiras; VIII- Administração da justiça e processos democráticos.
AIDA Canadense	A lista de exemplos ainda está em processo de desenvolvimento, mas já incluem: - Sistemas de triagem que afetam o acesso a serviços ou emprego; - Sistemas biométricos usados para identificação e inferência; - Sistemas que podem influenciar o comportamento humano em grande escala; -Sistemas críticos para a saúde e segurança.
Projeto de lei chileno	Sistemas de IA utilizados nos seguintes domínios: 1. Identificação biométrica remota em tempo real ou posterior de pessoas em espaços privados. 2. A utilização na gestão do abastecimento de água, electricidade e gás. 3. A atribuição e determinação do acesso aos estabelecimentos de ensino e a avaliação dos alunos. 4. A seleção e contratação de pessoas para empregos. 5. A atribuição de tarefas e a monitorização e avaliação do desempenho e comportamento dos trabalhadores. 6. A avaliação das pessoas para acesso a benefícios e serviços de assistência pública. 7. A avaliação da solvência das pessoas ou o estabelecimento da sua classificação de crédito. 8. Utilização em situações de emergência e catástrofe, como no envio ou estabelecimento de prioridades para o envio de serviços de intervenção (por exemplo, bombeiros ou ambulâncias). 9. A sua utilização para determinar o risco de pessoas cometerem crimes ou repetirem a sua prática, bem como o risco para potenciais vítimas de crimes. 10. A sua utilização em qualquer fase da investigação e interpretação de fatos que possam constituir crime no âmbito de um julgamento. 11. A sua utilização para gestão da migração, asilo e controle de fronteiras. 12. Da mesma forma, serão classificados como sistemas de IA de alto risco aqueles que apresentem o risco de causar danos à saúde e à segurança, ou o risco de ter repercussões negativas para os direitos fundamentais, cuja gravidade e probabilidade sejam equivalentes ou superiores aos riscos. repercussões negativas associadas aos sistemas de IA indicados no primeiro parágrafo deste artigo.

Normativa	Lista de critérios qualitativos e quantitativos para definição de atividades de alto risco
PL 2338/23	Critérios para avaliação se um sistema é de alto risco: I – a implementação ser em larga escala, levando-se em consideração o número de pessoas afetadas e a extensão geográfica, bem como a sua duração e frequência;II – o sistema puder impactar negativamente o exercício de direitos e liberdades ou a utilização de um serviço; III – o sistema tiver alto potencial danoso de ordem material ou moral, bem como discriminatório; IV – o sistema afetar pessoas de um grupo específico vulnerável; V – serem os possíveis resultados prejudiciais do sistema de inteligência artificial irreversíveis ou de difícil reversão; VI – um sistema de inteligência artificial similar ter causado anteriormente danos materiais ou morais; VII – baixo grau de transparência, explicabilidade e auditabilidade do sistema de inteligência artificial, que dificulte o seu controle ou supervisão; VIII – alto nível de identificabilidade dos titulares dos dados, incluindo o tratamento de dados genéticos e biométricos para efeitos de identificação única de uma pessoa singular, especialmente quando o tratamento inclui combinação, correspondência ou comparação de dados de várias fontes; IX – quando existirem expectativas razoáveis do afetado quanto ao uso de seus dados pessoais no sistema de inteligência artificial, em especial a expectativa de confidencialidade, como no tratamento de dados sigilosos ou sensíveis. (Art. 18)
EU AI Act (Comissão Europeia)	Para que o sistema seja considerado de alto risco, ele precisa cumprir dois requisitos: (a) Ser utilizado em uma das 8 áreas descritas no anexo III; e (b) Apresentar risco de dano à saúde, segurança ou de criação de efeitos adversos em direitos fundamentais: em termos de gravidade e probabilidade de ocorrência, apresenta risco maior ou equivalente aos riscos de dano ou efeitos adversos criados pelos sistemas de alto risco já listados no anexo III. Para fazer essa análise, analisa-se os seguintes critérios: - A finalidade pretendida do sistema de IA; - A extensão em que o sistema será utilizado ou se espera que seja utilizado; - A medida em que a utilização de um sistema de IA já causou danos à saúde e segurança ou impacto adverso nos direitos fundamentais ou deu origem a preocupações significativas em relação à materialização de tais danos ou impactos adversos, conforme demonstrado por relatórios ou alegações documentadas apresentadas às autoridades nacionais competentes; - A extensão potencial de tais danos ou impactos adversos, especialmente em termos da sua intensidade e da sua capacidade de afetar uma pluralidade de pessoas; - Até que ponto as pessoas potencialmente prejudicadas ou afetadas negativamente dependem do resultado produzido, pois, por razões práticas ou jurídicas, não é razoavelmente possível optar pela exclusão desse resultado; - Até que ponto as pessoas potencialmente prejudicadas ou afetadas negativamente se encontram numa posição vulnerável em relação ao utilizador de um sistema de IA,

	por exemplo, devido a um desequilíbrio de poder, de conhecimentos, de circunstâncias econômicas, sociais ou de idade; - Até que ponto o resultado produzido com um sistema de IA é facilmente reversível, pelo que os resultados que têm impacto na saúde ou na segurança das pessoas não devem ser considerados facilmente reversíveis; - Até que ponto a legislação da União em vigor prevê: (i) medidas eficazes de reparação em relação aos riscos colocados por um sistema de IA, com exclusão de pedidos de indenização; (ii) medidas eficazes para prevenir ou minimizar substancialmente esses riscos. (Art. 7º)
EU AI Act (Parlamento Europeu)	Para que o sistema seja considerado de alto risco, ele precisa representar risco significativo de danos à saúde e à segurança, ou um impacto adverso nos direitos fundamentais, no ambiente, ou na democracia e no Estado de direito. Esse risco é, no que diz respeito à sua gravidade e probabilidade de ocorrência, equivalente ou superior ao risco de danos ou de impacto adverso representado pelos sistemas de IA de alto risco do anexo III. Para fazer essa análise, analisa-se os seguintes critérios: - A finalidade pretendida do sistema de IA; - As capacidades e funcionalidades gerais do sistema, independentemente do seu propósito pretendido; - A extensão em que o sistema será utilizado ou se espera que seja utilizado; - A natureza e a quantidade de dados processados e usados pelo sistema; - Até que ponto o sistema de IA atua de forma autônoma; - Até que ponto a utilização de um sistema de IA já causou danos à saúde e à segurança, teve um impacto adverso nos direitos fundamentais, no ambiente, na democracia e no Estado de direito ou deu origem a preocupações significativas em relação à probabilidade de tais danos ou impactos adversos; - A extensão potencial de tais danos ou impactos adversos, especialmente em termos da sua intensidade e da sua capacidade de afetar uma pluralidade de pessoas ou que afeta desproporcionalmente um grupo específico de pessoas; - A medida em que as pessoas potencialmente prejudicadas ou afetadas negativamente dependem do resultado produzido, e esse resultado é puramente acessório no que diz respeito ao relevante, em particular porque, por razões práticas ou legais, não é razoavelmente possível optar pela exclusão desse resultado; - O potencial uso indevido e malicioso do sistema de IA e da tecnologia que o sustenta; - Até que ponto há um desequilíbrio de poderes ou as pessoas potencialmente prejudicadas ou afetadas negativamente se encontram numa posição vulnerável em relação ao utilizador de um sistema de IA, por exemplo, devido a um status, situação de autoridade, conhecimento ou circunstâncias econômicas, sociais ou de idade; - A extensão da disponibilidade e utilização de soluções e mecanismos técnicos eficazes para o controle, confiabilidade e correção do sistema de IA; - A magnitude e a probabilidade de benefícios da implementação do sistema de IA para indivíduos, grupos ou sociedade em geral, incluindo possíveis melhorias na segurança dos produtos; - A extensão da supervisão humana e a possibilidade de um ser humano interferir para anular uma decisão ou recomendações que possam causar danos potenciais;

	- Até que ponto o resultado produzido com um sistema de IA é facilmente reversível, pelo que os resultados que têm impacto na saúde ou na segurança das pessoas não devem ser considerados facilmente reversíveis; - Até que ponto a legislação da União em vigor prevê: (i) medidas eficazes de reparação em relação aos danos causados por um sistema de IA, com exclusão de pedidos de indemnização por danos diretos ou indiretos; (ii) medidas eficazes para prevenir ou minimizar substancialmente esses riscos. (Art. 7º)
AIDA Canadense	O Governo canadense lista os seguintes critérios para determinar quais os sistemas de IA seriam considerados de alto impacto: - Evidência de riscos de danos à saúde e segurança, ou risco de impacto adverso sobre os direitos humanos, com base tanto na finalidade pretendida quanto nas possíveis consequências não intencionais; - A gravidade dos danos potenciais; - A escala de uso; - A natureza dos danos ou impactos adversos que já ocorreram; - Até que ponto, por razões práticas ou legais, não é razoavelmente possível optar pela exclusão desse sistema; - Desequilíbrios das circunstâncias económicas ou sociais, ou idade das pessoas afetadas; e - O grau em que os riscos são adequadamente regulamentados por outra lei.
Canada: Directive on Automated Decision-Making + Algorithmic Impact Assessment tool	A Diretiva lida com IA em sua forma de decisões automatizadas Nesse contexto, decisões automatizadas de alto impacto são aquelas que afetam: direitos individuais ou de comunidades; igualdade, dignidade, privacidade e autonomia; saúde e bem-estar de indivíduos e grupos; e a sustentabilidade contínua de um ecossistema. Os critérios para identificar se as decisões são de alto impacto são: (i) dificuldade de reversão de seus resultados; (ii) se seus resultados são contínuos.
Projeto de lei chileno	Critérios para avaliar se um sistema de IA é de alto risco: - Se apresenta risco de causar danos à saúde e à segurança ou o risco de ter repercussões negativas para os direitos fundamentais, cuja gravidade e probabilidade sejam equivalentes ou superiores aos riscos de prejuízos ou repercussões negativas associadas à lista de exemplos de sistemas de IA de alto risco do art. 4º. (Art. 4º)

Nota-se certa convergência nos projetos de lei sobre IA ao redor do mundo na proposição de uma lista de exemplos de sistemas de IA de alto risco, a possibilidade de sua atualização e o enquadramento de novos sistemas (não originalmente listados) a partir da definição de critérios qualitativos e quantitativos. Tais critérios são mais bem desenvolvidos em determinados projetos, como o PL 2338/2023 e nas versões do EU AI Act, mas é importante que sejam definidos para que as legislações que regularão a IA não sejam estáticas e possam sobreviver ao decurso do tempo e dos rápidos avanços da tecnologia.

Apesar da convergência, há pontos ainda em discussão como a classificação de sistemas de IA utilizados para identificação biométrica, como a utilização de sistemas de identificação biométrica remota em tempo real em espaços acessíveis ao público, categorizado como prática proibida pela versão do Parlamento Europeu do EU AI Act. Nas demais propostas, como no PL 2338/23, versão do Conselho da UE da proposta europeia e na proposta de lei chilena, a proibição ocorre apenas nos casos de utilização desses sistemas para fins de segurança pública, com a listagem de algumas exceções.

a.3.3) *Risco Baixo (Residual)*

A categorização de risco baixo ou moderado representa o nível interventivo da regulação menos pesado, o que faz com que as obrigações relacionadas sejam mais brandas para os agentes de IA. Tal classificação pode ser alcançada a partir de elementos de identificação qualitativos e quantitativos ou pela sua definição como categoria residual, que pode ser seguida de exemplos, seja no próprio texto de lei ou na exposição de motivos. Nesse último caso, não há previsão de critérios diretos, mas indiretos de classificação, por meio de técnica legislativa de exclusão – isto é, casos que não se enquadram nos demais níveis mais intensos de risco são qualificados como residual.

Em algumas regulações, como o PL 2338/2023, o risco baixo representa uma categoria residual, pois é formado por todos os sistemas de IA que ficam de fora da classificação de risco excessivo/inaceitável ou de alto risco, não estando expressamente mencionado no texto de lei. Nesses casos, o projeto prevê estruturas de governança e processos internos como obrigatórios para todos os sistemas de IA, incluindo os de nível residual, para garantir a segurança dos sistemas e o atendimento dos direitos de pessoas afetadas. Tais obrigações estão listadas no art. 19 e incluem, por exemplo: (i) medidas de transparência quanto ao emprego de sistemas de inteligência artificial na interação com pessoas naturais, o que inclui o uso de interfaces ser humano-máquina adequadas e suficientemente claras e informativas; (ii) transparência quanto às medidas de governança adotadas no desenvolvimento e emprego do sistema de inteligência artificial pela organização; (iii) medidas de gestão de dados adequadas para a mitigação e prevenção de potenciais vieses discriminatórios; dentre outros.

No caso da proposta da União Europeia, por exemplo, a exposição de motivos define que os encargos regulamentares mais intensos serão destinados apenas quando um sistema de IA é suscetível de representar riscos elevados para os direitos fundamentais e a segurança. Para outros sistemas de IA de risco não elevado, considerado como uma categoria de risco limitada, a proposta impõe unicamente obrigações de transparência, a exemplo do fornecimento de informações para sinalizar a utilização de um sistema de

IA que interage com seres humanos. Porém, é válido ressaltar que o art. 52 (4) determina que tais obrigações de transparência são também aplicáveis para os demais sistemas de alto risco, enquadrados no título III da proposta.

Ademais, apesar de não expressamente mencionado no texto da proposta de regulamento, alguns estudos desenvolvidos pela Comissão Europeia e pelo Parlamento Europeu mencionam também um nível de risco baixo ou mínimo, que não estaria sujeito a obrigações extras no âmbito da proposta, podendo ser desenvolvidos e utilizados na UE sem cumprir quaisquer obrigações legais adicionais. Porém, a proposta prevê a criação de códigos de conduta para incentivar os fornecedores de sistemas de IA de risco não elevado a aplicarem voluntariamente os requisitos obrigatórios para sistemas de IA de risco elevado[63].

Normativa	Risco residual ou listagem de exemplos?	Exemplos	Obrigações
PL 2338/23	Risco residual (presumido – sem menção expressa)	-	Estruturas de governança definidas no art. 19, incluindo, por exemplo: medidas de transparência, de gestão de dados para mitigação e prevenção de vieses discriminatórios e de segurança da informação desde a concepção até a operação do sistema.
PL 2338/23	Risco limitado – exemplos	Sistemas que interagem com humanos (ou seja, chatbots), sistemas de reconhecimento de emoções, sistemas de categorização biométrica e sistemas de IA que geram ou manipulam conteúdo de imagem, áudio ou vídeo (ou seja, deepfakes)	Apenas obrigações mínimas de transparência
	Residual (baixo ou mínimo)	-	Sem obrigações.
AIDA	Risco residual presumido (não é expressamente mencionado)	-	Não especificado.

63 https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/698792/EPRS_BRI(2021)698792_EN.pdf.

Por se tratar de um nível menos arriscado para direitos fundamentais, recebe menos atenção dos documentos analisados. Como regra, a maior parte das propostas de regulação trazem o risco baixo como categoria residual, a partir da classificação por exclusão de todos os sistemas de IA não classificados nos graus de riscos mais intensos. Porém, mesmo se tratando de uma categoria menos focada pela regulação, há ainda a previsão de obrigações mínimas, especialmente ligadas à transparência.

a.4) Abordagem *ex-ante* e *ex-post* de risco

Com a abordagem baseada em riscos expandida também para o campo de regulação da IA, a atividade de classificação de risco passa também a ser necessária, como visto nos tópicos anteriores. Como já mencionado anteriormente por Hood et al (2001), a regulação de risco – e consequentemente sua classificação – varia de um domínio para outro e pode mudar ao longo do tempo[64]. Os princípios da OCDE sobre IA exigem que os agentes sejam responsáveis pelo bom funcionamento dos seus sistemas de IA, de acordo com o seu papel, contexto e capacidade de ação[65]. Assim, somado ao fato de que a IA é uma tecnologia complexa e em rápida evolução, tal exercício classificatório dos riscos neste cenário não é trivial, exigindo sempre uma análise contextual.

No campo da IA, além de extremamente contextual, a análise do risco pode se dar de forma prévia (*ex-ante*) ou a posteriori (*ex-post*). No primeiro caso, a avaliação pode se dar tanto a partir da criação de uma lista de exemplos de riscos inaceitáveis e altos (ou pela criação de critérios para sua classificação), mas também pelo exercício de avaliação dos sistemas de IA desde a fase de concepção do produto e/ou serviço, que deve ser alocada para a cadeia de agentes envolvidos. No caso do PL 2338/23 do Brasil, tal análise é prevista no art. 13 que determina que todo sistema de IA passe por avaliação preliminar para classificação de seu grau de risco – o que será fundamental para a alocação de obrigações para cada um dos atores.

Segundo Kaminski, a regulamentação do risco normalmente foca sua atenção em medidas *ex-ante* e subutiliza as ferramentas pós-comercialização[66]. Porém, a abordagem baseada em riscos com sua avaliação *ex-ante* e *ex-post* torna o processo regulatório baseado em risco mais completo, por meio de um movimento de aprendizagem, já que permite a reclassificação dos sistemas em momento posterior, caso haja alguma mudança significativa ao longo de sua aplicação. Isso faz com que a regulação ocorra de forma dinâmica, não estática e colaborativa. Essa abordagem é encontrada, por exemplo, na

64 HOOD *et al*, 2001, p. 3.

65 OECD. OECD AI Principles overview. Disponível em: https://oecd.ai/en/ai-principles.

66 KAMINSKI, 2022, p. 72–73.

União Europeia (EU AI Act em todas as suas versões), no Brasil (PL 2338/23), no Chile (Projeto de Lei 15869-19) e no Canadá com a previsão de critérios para reclassificação da lista de exemplos de riscos inaceitáveis e altos, além da criação de uma base de dados publicamente acessível de IAs, geralmente obrigatória para os casos de alto risco – o que permite a participação de toda a sociedade no monitoramento e avaliação desses riscos após a implementação dos sistemas.

Normativa	Previsão *ex-ante*	Previsão *ex-post*	Previsão de criação de base de dados publicamente acessível de IAs de alto risco
PL 2338/23	Sim	Sim	Sim
Proposta Europeia (*AI Act*)	Sim	Sim	Sim
Projeto de Lei 15869-19 do Chile	Sim	Sim	Sim
Ferramenta de Avaliação de Impacto Algorítmico do Canadá	Sim	Sim	Sim

Logo, por mais que a regulação de risco varie de acordo com a área em que ela é aplicada, é certo que sua análise e a classificação dos riscos deve sempre envolver uma atuação contextual, que pode ser feita tanto *ex-ante* ou *ex-post* à implementação dos sistemas de IA. A regulação e classificação do risco de forma prévia e posterior é exteriorizada, geralmente, em ferramentas de avaliação de impacto, que, como será visto no tópico seguinte, devem ser contínuas, atualizáveis de tempos em tempos (ou no caso de alterações significativas nos sistemas) e com a participação pública significativa de todos os setores da sociedade.

EIXO 2 – Avaliações de impacto algorítmico – AIA

As avaliações de impacto, já conhecidas da seara ambiental e de proteção de dados pessoais em suas espécies de relatórios de impacto ao meio ambiente e relatório de impacto à proteção de dados pessoais, são instrumentos de governança surgidos para analisar as possíveis consequências de uma iniciativa sobre interesses sociais relevantes. E, a partir dessa análise, apoiar um processo decisório informado sobre se deve se realizar a iniciativa e, caso afirmativo, sob quais condições. São aplicadas em situações nas quais

há uma incerteza sobre eventos futuros, como na emergência de novas tecnologias[67]. Por isso, as avaliações de impacto são mecanismos para gerar evidências sobre a tomada de decisões e para proteção de certas preocupações da sociedade[68].

Desde logo, cabe destacar que o instrumento de avaliação de impacto se difere de outras atividades organizacionais, a exemplo de uma avaliação de conformidade regulatória (que pode ser feita *ex-post*, ainda que não seja o ideal), em razão de seu caráter precaucionário e preventivo. Seu objetivo é a identificação de riscos e aplicação de medidas de mitigação eficientes antes da implementação de uma determinada tecnologia, seguindo um escrutínio público que desengatilha um controle social e em rede de governança[69].

Não por outra razão, uma das principais aspirações em jogo é fazer com que as avaliações de impacto – do campo ambiental, como no caso de avaliação de uma construção de rodovia próxima à mata ciliar de um rio, e chegando à seara da proteção de dados – criem um procedimento em que todas as partes interessadas possam entender e influir em um determinado processo de tomada de decisão. Trata-se de uma questão de justiça procedimental[70], sendo que o que está em jogo não diz respeito apenas ao resultado justo, mas, também, que o processo percorrido para chegar nele também o seja.

No cenário da crescente aplicação de sistemas de inteligência artificial para automatização de decisões do nosso cotidiano, isso relaciona-se com o que se convencionou a chamar de "devido processo informacional"[71]. Isto é, uma forma de concretização do contraditório e da ampla defesa e, por conseguinte, de contenção sobre ações que interferem indevidamente em liberdades públicas - e.g., policiamento preditivo - e direitos individuais - e.g., liberdade de expressão no cenário de moderação de conteúdo, por meio de maior controle sobre os procedimentos que são realizados.

Ainda, é importante ressaltar que a condução do instrumento de avaliação de impacto deve ser vista não como um fardo ou mera obrigação para o fornecedor, mas como uma oportunidade. Dada a natureza dos produtos/serviços de IA e seus recursos e escala, o modelo de avaliação proposto pode ajudar significativamente as empresas e outras entidades a desenvolver IA centrada no ser humano e eficaz, mesmo em contextos desafiadores[72]. Com isso, gerar mais confiança não só na tecnologia, mas, também, nas trocas econômicas em seu entorno.

Por fim, é essencial destacar que, mais importante do que a mera previsão de um

67 KLOZA *et al*, 2019.

68 KLOZA *et al*, 2017.

69 BIONI, Bruno Ricardo. *Regulação e Proteção de Dados Pessoais – O Princípio da Accountability*. São Paulo: Editora Forense, 2022. 320p.

70 KLOZA *et al*, 2019.

71 BIONI; MARTINS, 2020.

72 MANTELERO, 2022.

instrumento de avaliação de impacto algorítmico é que ele seja minimamente procedimentalizado para que se torne uma efetiva ferramenta de devido processo informacional e prestação de contas. A título de exemplo, o brasileiro PL 21/20 conceituou no art. 2º, VI o que seria um relatório de impacto de inteligência artificial, porém, sem dar maiores detalhes quanto aos seus objetivos, prazos e parâmetros mínimos, o que gera insegurança jurídica. Assim, para além da previsão legal de condução de um AIA, é essencial que haja também a definição de parâmetros mínimos de metodologia, critérios, etapas e, eventualmente, previsões sobre a necessidade de publicação e revisão periódica. Nesse ponto, o PL 2338/2023 avança em comparação aos demais, em similaridade com o que é feito no AI EU Act, Canadá e demais instrumentos internacionais, como veremos ao longo dos próximos tópicos.

b.1) Metodologia, critérios e momento de realização

Como ensinamento prévio retirado da figura do Relatório de Impacto à Proteção de Dados Pessoais (RIPD)[73], a falta de uma procedimentalização mínima, isto é, sistematização do que deve conter nesta ferramenta (a exemplo de prazos, critérios e metodologia escolhida), traz dificuldades para sua concretização. Ao mesmo tempo, tal tipo de parametrização não pode ser demasiadamente prescritiva para não enrijecer uma ferramenta que deve ser tão dinâmica quanto é o desenvolvimento de tecnologias de IA e outras técnicas de tratamento de dados. Consequentemente, é bem vindo que a futura regulação de IA apresente uma sistematização mínima, como sendo uma espécie de fundação, para uma edificação sólida de avaliação de impacto algorítmico. Dito de outra forma, um piso e não um teto para modelagem desta importante ferramenta.

Um estudo preliminar sobre AIA do Comitê Ad Hoc de Inteligência Artificial do Conselho da Europa constatou[74], ao avaliar as estruturas gerais de avaliações de impacto de direitos humanos, a tendência de que esses instrumentos se concentrem em impactos adversos de determinada iniciativa sobre esses direitos, o que também acontece na maioria dos atuais modelos de avaliação de impacto de sistemas de IA.

Para o CAHAI, as avaliações de impacto de IA devem ser desenvolvidas de acordo com esta abordagem, porém, isso não significa que o uso da IA gere apenas impactos adversos, já que a tecnologia tem muitas vantagens e pode criar um enorme impacto

73 BIONI, Bruno Ricardo; ZANATTA, Rafael A. F.; RIELLI, Mariana Marques. Contribuição à Consulta Pública da Estratégia Brasileira de Inteligência Artificial, Data Privacy Brasil Research, disponível em: https://www.dataprivacybr.org/wp-content/uploads/2020/06/E-BOOK-CONTRIBUIC%CC%A7A%CC%830-DPBR-INTELIGE%CC%82NCIA-ARTIFICIAL-FINAL.pdf; BIONI, Bruno; EILBERG, Daniela Dora; CUNHA, Brenda; SALIBA, Pedro; VERGILI, Gabriela. Proteção de dados no campo penal e de segurança pública: nota técnica sobre o Anteprojeto de Lei de Proteção de Dados para segurança pública e investigação criminal. São Paulo: Associação Data Privacy Brasil de Pesquisa, 2020.

74 CAHAI, 2022, p. 4.

benéfico para a humanidade. No entanto, de acordo com o CAHAI[75], a função específica primordial das avaliações de impacto em direitos humanos, que deveriam basear as AIA, deveria ser detectar possíveis riscos de violação de direitos humanos decorrentes de um determinado sistema de IA, e não contra balanceá-los com possíveis impactos benéficos decorrentes de tal aplicação.

Desta forma, o equilíbrio entre benefícios e riscos não faria parte necessariamente da metodologia de avaliação de impacto, mas auxiliaria, posteriormente, no julgamento de oportunidade quanto à implantação (ou não) de sistemas de IA. Por exemplo, em certos casos, os tomadores de decisão, como autoridades públicas, podem concluir que os impactos benéficos compensam os impactos adversos e, portanto, decidir usar tal aplicação para uma determinada finalidade.

Já no âmbito das Recomendações da UNESCO sobre ética de sistemas de IA, a função da ferramenta de AIA seria a de "identificar e avaliar benefícios, preocupações e riscos dos sistemas de IA, bem como prevenção, mitigação e reparação de riscos e medidas de monitoramento", tendo como base os impactos em direitos humanos e nas liberdades fundamentais, especialmente aos direitos das pessoas em situações precárias e de vulnerabilidade, aos direitos laborais e ao meio-ambiente[76].

Somado a isso, a partir de análise comparativa de avaliações de impacto em diferentes áreas realizada pelo d.pia.lab, é possível estabelecer elementos que constituem uma boa prática para avaliações de impacto adaptáveis para diferentes áreas[77]. Nessa linha, estabeleceu-se um método genérico para avaliação de impacto que consiste em 10 passos agrupados em 5 fases. São elas: (i) preparação; (ii) avaliação/análise; (iii) recomendações; (iv) etapas contínuas; (v) revisão[78]. Quando transplantada para o contexto de IA, compreende-se que a ferramenta serve para identificar, descrever e analisar tanto as

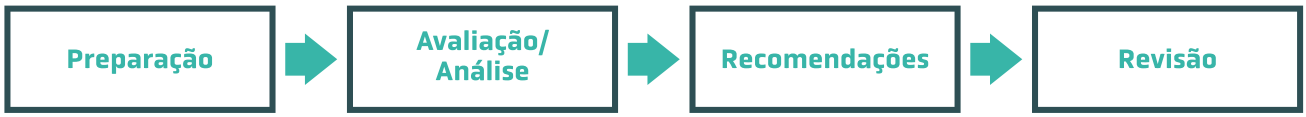
75 Texto original: "*Obviously, this does not imply that the use of AI generates adverse impacts only. AI has many advantages and can create a huge beneficial impact for mankind. It may even assist in the enjoyment, protection and strengthening of human rights, and this positive contribution should not be neglected. However, the specific function of HRIA is to detect possible risks of infringement for human rights arising from a given AI system, and not to balance them against possible beneficial impacts arising from such an application. Balancing benefits against risks is not part of the assessment methodology but would rather be performed later as part of a judgement of opportunity as to whether deploy such application. For instance, in certain cases public authorities could conclude that the beneficial impacts offset adverse impact and hence decide using such application for a given purpose. If in this case one or more human rights are curbed (which the HRDRIA can help assess) it is essential that this occurs in a manner that is justified through an approach that is both proportionate and necessary in a democratic society, for instance in the interest of national security or another legitimate public interest*". CAHAI. Human Rights, Democracy and Rule of Law Impact Assessment of AI systems. Policy Development Group (CAHAI-PDG). Strasburg: 21 de maio de 2021. p. 4.

76 UNESCO. Recommendation on the Ethics of Artificial Intelligence. Adotado em 23 novembro de 2021 e publicado em 2022. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000381137.

77 KLOZA, D., *et al.* (2017). Data protection impact assessments in the European Union: complementing the new legal framework towards a more robust protection of individuals. **d.pia.lab Policy Brief**, (1/2017), 1–4. https://doi.org/10.31228/osf.io/b68em, https://doi.org/10.5281/zenodo.5121575.

78 KLOZA, D., *et al.* (2019). Towards a method for data protection impact assessment: Making sense of GDPR requirements. d.pia.lab Policy Brief, 1(2019), 1–8. https://doi.org/10.31228/osf.io/es8bm, https://doi.org/10.5281/zenodo.5121534.

possíveis consequências do sistema em análise como as possíveis soluções para endereçar tais consequências.



Processos contínuos

De forma similar, o relatório "Avaliação de Impacto de Direitos Humanos, Democracia e Estado de Direito de Sistemas de IA" do CAHAI definiu quatro etapas mínimas para o desenvolvimento desta ferramenta, de forma a contemplar a identificação de direitos relevantes, avaliar os impactos nestes direitos (incluindo, como critérios, o escopo e escala da aplicação e o potencial de pessoas impactadas), mecanismos de governança e avaliações constantes[79].

Relatório sobre prestação de contas em IA produzido pela OCDE[80] em 2023 também ressaltou a existência de, pelo menos, 4 etapas para gerenciamento de riscos de sistemas de IA, que seria formado por: definição (escopo, contexto, atores envolvidos e critérios de análise), avaliação (identificação dos riscos individuais e coletivos a partir de gravidade x probabilidade), tratamento dos riscos (medidas de mitigação) e governança (monitoramento e revisão). Essa orientação está de acordo com outros frameworks internacionais vindos, por exemplo, do NIST, ISO 31000 e outros documentos da própria OCDE[81].

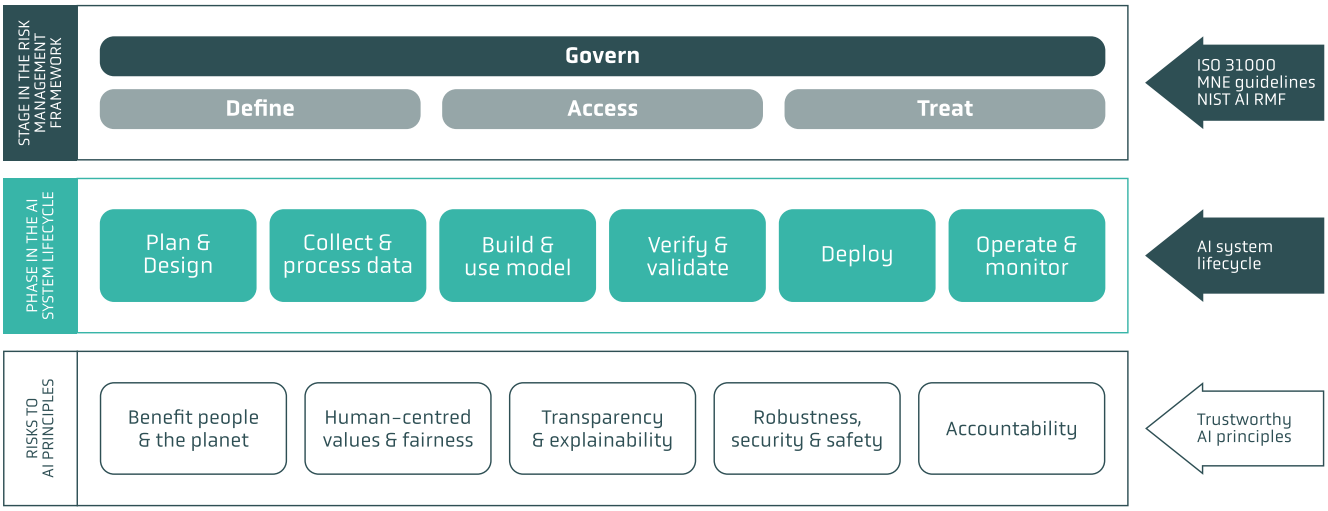
Recentemente, a OCDE também publicou "Orientações comuns para promover a interoperabilidade na gestão de riscos de IA", em que confirma que os principais modelos e quadros de gestão de riscos se alinham a essas quatro etapas. Embora a o público-alvo, o âmbito do risco, o segmento do ciclo de vida de IA, a terminologia específica utilizada

79 CAHAI. Human Rights, Democracy and Rule of Law Impact Assessment of AI systems. Conselho da Europa, CAHAI--PDG (2021)5. Strasbourg, 21 maio 2021. Disponível em: https://rm.coe.int/cahai-pdg-2021-05-2768-0229-3507-v-1/1680a291a3.

80 OECD. Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI. Publicado em 23 Feb. 2023. Disponível em: https://www.oecd-ilibrary.org/science-and-technology/advancing--accountability-in-ai_2448f04b-en.

81 OECD, 2023a.

e a própria ordem das etapas possam variar entre os documentos existentes, os modelos geralmente procuram alcançar os mesmos resultados (e.g. IA responsável, ética e confiável) por meio de processos similares de gestão de riscos em quatro etapas[82].



(OCDE, 2023b)

De forma similar, no âmbito brasileiro, o PL 2338/2023 define também uma metodologia de, pelo menos, quatro etapas representadas pela preparação, cognição do risco, mitigação destes riscos e monitoramento (art. 24).

Nesse contexto, de acordo com Alessandro Mantelero (2022), há pelo menos três fatores essenciais que devem ser considerados em uma análise de risco: (i) identificação deste risco; (ii) probabilidade do risco se concretizar; (iii) gravidade do risco identificado. Para a identificação, recomenda-se a inclusão de direitos de forma ampla como categorias potencialmente afetadas, de forma a garantir a proteção integral de pessoas naturais e dos diferentes grupos impactados frente aos possíveis riscos desencadeados pelo uso do sistema de IA, além do meio-ambiente. Logo, como mencionado, no contexto de proposição de um modelo híbrido baseado em riscos e direitos, o risco se relaciona aos potenciais prejuízos a direitos e liberdades fundamentais de pessoas naturais[83], considerando limitações e restrições, independentemente da concretização de danos materiais.

82 OECD. Common guideposts to promote interoperability in AI risk management. 07 nov. 2023. Disponível em: https://www.oecd-ilibrary.org/science-and-technology/common-guideposts-to-promote-interoperability-in-ai-risk-management_ba602d18-en.

83 GOMES, Maria Cecília. Relatório de impacto à proteção de dados: uma breve análise de sua definição e papel na LGPD. **Revista da AASP**, n. 144, 2019. p. 10-11.

Assim, a análise do risco formal geralmente envolve algum tipo de matemática que envolve elementos de probabilidade de um evento acontecer e a gravidade do dano potencialmente causado por esse evento[84]. Isto é, o impacto esperado dos riscos identificados é avaliado considerando tanto a probabilidade quanto a gravidade das consequências negativas esperadas, usando uma escala variável, que geralmente é realizada em quatro etapas (baixa, média, alta, muito alta/excessiva/inaceitável). Essa gradação, no entanto, é variável, a depender da matriz de risco adotada, que pode possuir gradações diferentes, indo da mais simples (três níveis: baixo, moderado ou alto) até os mais complexos (com quatro a cinco níveis de risco, por exemplo)[85].

Mantelero (2022) considera a probabilidade como a combinação de dois elementos: a probabilidade de consequências adversas e a exposição (número potencial de pessoas em risco), enquanto a gravidade é avaliada considerando a natureza do potencial prejuízo no exercício de direitos e suas consequências, o que inclui a verificação também do esforço necessário para superar os potenciais prejuízos e reverter os efeitos adversos. Nessa linha, por exemplo, o § 1º do art. 24 do PL 2338/2023, que dispõe sobre a avaliação de impacto para sistemas de alto risco, obriga que esta ferramenta considere e registre tanto a identificação do risco - que deve englobar tanto os riscos conhecidos e previsíveis como os que podem razoavelmente dele se esperar - como a probabilidade e a gravidade das consequências adversas.

Conforme será visto em tabela comparativa abaixo, a maior parte das avaliações propostas focam em riscos a direitos individuais, difusos, coletivos e individuais homogêneos[86] dos atingidos pelos sistemas de IA, de forma ampla, com o intuito de garantir

84 KAMINSKI, 2022, p. 8.

85 TV Senado. Comissão de juristas promove debates sobre regulação da inteligência artificial (2ª parte)– 29/04/22. Publicado em 29 abr. 2022. Fala da professora Maria Cecília Gomes. Disponível em: https://www.youtube.com/watch?v=P_yWp-2ZIZs&t=51s. Acesso em 21 jul. 2023.

86 Os direitos coletivos em sentido amplo são uma conquista social importante e tiveram sua consagração com a Constituição Federal de 1988 e os demais regramentos do microssistema (processual) coletivo, como a Lei de Ação Civil Pública e o Código de Defesa do Consumidor. Esses direitos podem ser divididos em direitos difusos, coletivos (stricto sensu) e individuais homogêneos, como previsto no parágrafo único do art.81 do CDC. Nesse contexto, os interesses

proteção integral não apenas aos direitos humanos, mas também valores éticos, sociais, democráticos e de Estado de Direito.

Nesse sentido, Mantelero (2022) sustenta que os sistemas de IA carregam consigo uma complexidade que exige que as avaliações de impacto sejam desenvolvidas a partir de um modelo misto de análise do seu impacto ético e social juntamente com as dimensões legais, como as dos direitos humanos. Para tanto, ele defende a necessidade de uma abordagem multistakeholder e centrada no ser humano, combinando a universalidade dos direitos humanos com a dimensão local dos valores sociais. De forma similar, o CAHAI propõe uma análise abrangente de direitos humanos, democracia e Estado de Direito[87] e a UNESCO menciona que a avaliação deve ter como referência não apenas pessoas físicas ou grupos/comunidades afetados, mas também o meio-ambiente[88].

Assim, considerando que os riscos variam, a depender do sistema de IA a ser utilizado, e que uma avaliação de impacto não é um processo trivial, tanto em sua realização quanto para análise, essa exigência fica geralmente restrita aos sistemas de IA de alto risco, sem prejuízo de sua realização como boa prática para os sistemas de IA de risco mais baixo, como será visto abaixo. É isso que está previsto no PL 2338/23 e no EU AI Act, assim como a AIDA canadense. Porém, além da classificação de alto risco como critério de destrave da obrigação de avaliações de impacto, há algumas iniciativas que associam outros critérios, como a natureza da organização, a exemplo de sistemas de IA utilizados pelo poder público, como reforçado pela Ferramenta de Avaliação de Impacto Ético de IA fornecida no âmbito das recomendações da UNESCO[89].

Neste ponto, considerando que a deflagração da obrigação de elaboração da AIA geralmente está associada ao grau de risco do sistema em causa, é necessário que haja um processo de avaliação deste grau pelos atores da cadeia produtiva de IA. A título exemplificativo, destaca-se o art. 13 do PL 2338/2023 ao prever que, previamente à colocação no mercado ou utilização em serviço, todo sistema de IA deve se submeter a uma avaliação preliminar para a classificação de seu risco, conforme critérios definidos nos

---

ou direitos difusos são entendidos como os transindividuais, de natureza indivisível, de que sejam titulares pessoas indeterminadas e ligadas por circunstâncias de fato, como no caso do direito a um meio ambiente sadio. Já os interesses ou direitos coletivos em sentido estrito são os transindividuais, de natureza indivisível de que seja titular grupo, categoria ou classe de pessoas ligadas entre si ou com a parte contrária por uma relação jurídica base, como acontece no caso de consumidores de serviços públicos essenciais. Nesse caso, é possível determinar quem são os titulares, já que há uma relação jurídicas entre as pessoas atingidas. Por fim, os interesses ou direitos individuais homogêneos são aqueles decorrentes de um evento com origem comum, como no caso de consumidores lesados por um produto defeituoso – aqui, há a possibilidade tanto de ajuizamento de uma ação individual como coletiva; Conselho Nacional do Ministério Público. Portal de Direitos Coletivos. Disponível em: https://www.cnmp.mp.br/direitoscoletivos/; GAJARDONI, Fernando da Fonseca. **Direitos Difusos e Coletivos I: Teoria Geral do Processo Coletivo**. São Paulo: Saraiva, 2012.

87 https://rm.coe.int/cahai-pdg-2021-05-2768-0229-3507-v-1/1680a291a3.

88 UNESCO. Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence. Publicado em 2023. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000386276/PDF/386276eng.pdf.multi.

89 UNESCO, 2023.

artigos relacionados ao risco excessivo e alto. Nesse âmbito, o projeto brasileiro define também a obrigatoriedade de registro e documentação desta avaliação para fins de responsabilização e prestação de contas a posteriori, impedindo que o processo de classificação de riscos seja deixado apenas a cargo dos atores regulados, permitindo, inclusive, que a autoridade competente determine a reclassificação de sistemas e eventual penalização por análises fraudulentas.

Já no contexto europeu, a versão do AI Act do Parlamento não prevê expressamente a realização de uma avaliação prévia, mas deixa subentendido ao prever alguns critérios para definição dos riscos inaceitáveis e altos, sem, contudo, determinar mecanismos de fiscalização por parte das autoridades competentes, o que pode acabar por tornar dificultar a implementação eficiente e harmônica do texto de lei[90].

Ademais, para além do seu escopo, é importante destacar que a AIA deve ser entendida como um processo contínuo e não instantâneo de um momento no tempo[91]. Como será visto, parece ser um consenso entre os documentos de que a AIA seja realizada, pelo menos, antes de o sistema ser efetivamente posto à disposição do público, seja como serviço ou como produto. Essas previsões vão ao encontro do que defendeu Maria Cecília Gomes em audiência pública realizada no âmbito da CJSUBIA em abril de 2022. Segundo ela, essa avaliação precisa ser feita no momento do desenvolvimento da IA e os riscos precisam ser avaliados antes de se concretizarem[92]. Assim, percebe-se que a figura da AIA é inerentemente preventiva, pautado por uma lógica de regulação *ex-ante*, ao invés desta documentação ser um diagnóstico de futuros eventos adversos[93], quando o risco já se concretizou após a tecnologia já ter sido disponibilizada no mercado[94].

Assim, a obrigação de realização de avaliações de impacto algorítmico, previamente à disponibilização do serviço ou colocação do produto no mercado, pode trazer benefícios para as organizações a partir de uma mudança para o pensamento antecipatório e ex ante. As organizações passam a ser capazes de refletir sobre as consequências de suas iniciativas, bem como sobre os meios para minimizar ou, às vezes, até evitar consequências negativas e não intencionais antes que elas ocorram, o que levam-nas a angariar a confiança do público a médio e longo prazo – e consequentemente, ganho reputacional[95].

90 Access Now. EU Trilogues: The AI Act must protect people's rights. Publicado em 12 jul 2023. Disponível em: https://www.accessnow.org/press-release/eu-trilogues-ai-act/; Access Now. Joint statement: EU legislators must close dangerous loophole in AI Act. Publicado em 07 set. 2023. Disponível em: https://www.accessnow.org/press-release/eu-trilogues-ai-act/https://www.accessnow.org/press-release/joint-statement-eu-legislators-must-close--dangerous-loophole-in-ai-act/.

91 CAHAI, 2021, p. 4.

92 Ibid.

93 KAMINSKI, 2022, p. 19.

94 https://rm.coe.int/cahai-pdg-2021-05-2768-0229-3507-v-1/1680a291a3.

95 KLOZA, Dariusz *et al* Data protection impact assessments in the European Union: complementing the new legal framework towards a more robust protection of individuals. d.pia.lab Policy Brief, 2017.

Porém, para além da sua importância preventiva, trata-se também de ferramenta que consiste em um processo sistemático que começa razoavelmente cedo no ciclo de vida de uma única iniciativa (como, no caso, sistemas de IA), antes de sua implantação, continua ao longo de seu ciclo de vida e – à medida que a sociedade muda, os perigos evoluem e o conhecimento cresce – é revisitado quando necessário[96]. Por isso, é possível referir às avaliações de impacto como um "instrumento vivo"[97] que influenciará continuamente todas as partes de desenvolvimento dos sistemas de IA.

Nesta linha, a UNESCO também sustenta que a AIA deve ser um documento vivo, preenchido progressiva e iterativamente em diferentes fases do ciclo de vida de IA, incluindo, por exemplo, as seguintes etapas: (i) concepção, desenvolvimento e pré-aquisição do sistema; (ii) aquisição, quando a AIA pode ajudar tanto na seleção de um fornecedor como na formulação de obrigações contratuais; (iii) após a implementação do sistema, quando a AIA deve ser revista em intervalos regulares, especialmente porque as respostas pode mudar ao longo do tempo à medida que a tecnologia evolui[98].

No caso do PL 2338/2023, há expressa previsão de atualizações periódicas da AIA, que deve fazer parte de todo o ciclo de vida dos sistemas de IA de alto risco (art. 25 e art. 24 parágrafo quarto). Assim, mesmo tratando-se de uma ferramenta preponderantemente realizada em momento anterior ao lançamento da tecnologia no mercado, a sua atualização ao longo do ciclo de vida da IA é indispensável. Não apenas porque novas técnicas surgem no decorrer do tempo, mas também porque incidentes podem informar a atualização de todo o processo de gerenciamento de risco para torná-lo ainda mais resiliente. Há, pois, também uma vertente *ex-post* desta ferramenta.

Além disso, no caso da regulação baseada no risco, trata-se de um instrumento inserido no processo regulatório de aprendizagem que deve ser dinâmico e iterativo. Em outras palavras, a gestão de riscos não é feita apenas como mera verificação *ex-ante* a ser concluída, mas um processo híbrido que deve ser repetido de tempos em tempos e alterado conforme os riscos e os conhecimentos sobre as tecnologias mudam, permitindo que o risco seja recalibrado em momento *ex-post*. É isso que é observado, por exemplo, na GDPR, EU AI Act, PL 2338/23 e outros documentos listados na tabela abaixo.

Nesse contexto, não há uma resposta única para quem deve ser o ator responsável por deflagar a obrigação de renovação ou atualização da avaliação de impacto algorít-

96 MANTELERO, 2022.

97 Em Kloza *et al* (2017), os autores mencionam a avaliação de impacto como um "instrumento vivo" ("*living instrument*") para explicar o fato de que a ferramenta necessita constantemente de reflexão, uma vez que a avaliação da tecnologia começa cedo (antes de sua implantação), continua ao longo de seu ciclo de vida (já implementada) e, a medida que a sociedade avança, os perigos aumentam e o crescimento evolui, é necessária a sua revisão, de forma a influenciar o design da própria tecnologia.

98 UNESCO, 2023, p. 8.

mico, nem a periodicidade ideal de realização. Porém, parece haver uma tendência na direção de sua revisão de tempos em tempos, principalmente no caso em que há uma alteração dos riscos ou das circunstâncias em que a tecnologia se encontra. A título de exemplo, no PL 2338/2023 há a atribuição para a futura autoridade competente de IA de definir a periodicidade da revisão da AIA, enquanto o projeto de lei sobre ferramentas de decisões automatizadas do estado estadunidense da Califórnia define sua atualização anual.

Legislação	Há menção sobre a avaliação de impacto ser um documento contínuo?	Trecho de menção	Há previsão de algum tipo de supervisão estatal com poderes para tanto?	Trecho de menção
PL 2338/2023	Sim	Art. 19, § 1º As medidas de governança dos sistemas de inteligência artificial são aplicáveis ao longo de todo o seu ciclo de vida, desde a concepção inicial até o encerramento de suas atividades e descontinuação. Art. 24, § 4º Caberá à autoridade competente a regulamentação da periodicidade de atualização das avaliações de impacto, considerando o ciclo de vida dos sistemas de inteligência artificial de alto risco e os campos de aplicação, podendo incorporar melhores práticas setoriais. Art. 25. A avaliação de impacto algorítmico consistirá em processo iterativo contínuo, executado ao longo de todo o ciclo de vida dos sistemas de inteligência artificial de alto risco, requeridas atualizações periódicas. § 1º Caberá à autoridade competente a regulamentação da periodicidade de atualização das avaliações de impacto.	Sim	Art. 24, § 4º Caberá à autoridade competente a regulamentação da periodicidade de atualização das avaliações de impacto, considerando o ciclo de vida dos sistemas de inteligência artificial de alto risco e os campos de aplicação, podendo incorporar melhores práticas setoriais. Art. 25, § 1º Caberá à autoridade competente a regulamentação da periodicidade de atualização das avaliações de impacto.
EU AI Act (Versão do Parlamento Europeu)[99]	Sim	Article 9 (1) A risk management system shall be established, implemented, documented and maintained in relation to high-risk AI	Sim	(85a) Given the rapid technological developments and the required technical expertise in conducting the assessment of high-risk AI

99 A versão original do EU AI Act não previa a figura da avaliação de impacto, mas apenas um processo de gerenciamento de riscos nos casos de sistemas de IA de alto risco (art. 9). Porém, na última versão da proposta, publicada pelo Parlamento Europeu, além do gerenciamento de riscos, previu também a obrigatoriedade de elaboração pelo implantador (*deployer*) de uma avaliação de impacto de direitos humanos para sistemas de IA de alto risco (art. 29a).

EU AI Act (Versão do Parlamento Europeu)		systems, throughout the entire lifecycle of the AI system. The risk management system can be integrated into, or a part of, already existing risk management procedures relating to the relevant Union sectoral law insofar as it fulfills the requirements of this article. (2) The risk management system shall consist of a continuous iterative process run throughout the entire lifecycle of a high-risk AI system, requiring regular review and updating of the risk		systems, the Commission should regularly review the implementation of this Regulation, in particular the prohibited AI systems, the transparency obligations and the list of high-risk areas and use cases, at least every year, while consulting the AI office and the relevant stakeholders.
GDPR	Sim	Art. 37 (11) Se necessário, o responsável pelo tratamento procede a um controlo para avaliar se o tratamento é realizado em conformidade com a avaliação de impacto sobre a proteção de dados, pelo menos quando haja uma alteração dos riscos que as operações de tratamento representam.	Sim	Art. 37 (11) Se necessário, o responsável pelo tratamento procede a um controlo para avaliar se o tratamento é realizado em conformidade com a avaliação de impacto sobre a proteção de dados, pelo menos quando haja uma alteração dos riscos que as operações de tratamento representam.
Directive on Automated Decision-Making (Canadá)	Sim	6.1.3 Reviewing and updating the Algorithmic Impact Assessment on a scheduled basis, including when the functionality or scope of the automated decision system changes.	Sim	Subject to the necessary delegations, the Chief Information Officer of Canada is responsible for: (...) 8.2 Developing and maintaining the Algorithmic Impact Assessment and any supporting documentation.
Algorithmic Impact Assessment tool (Canadá)	Sim	3.1 When to complete the AIA: The AIA should be completed at the beginning of the design phase of a project. The results of the AIA will guide the mitigation and consultation requirements to be met during the implementation of the automated decision system as per the directive. The AIA should be completed a second time, prior to the production of the system, to validate that the results accurately reflect the system that was built	Sim	3.2 What to consider when completing an AIA: The Office of the Chief Information Officer (OCIO) at the Treasury Board of Canada Secretariat (TBS) is responsible for maintaining the AIA tool and overseeing departmental compliance with the Directive on Automated Decision-Making.

Algorithmic Accountability Act EUA	Sim	Sec. 2 (12) IMPACT ASSESSMENT.—The term "impact assessment" means the ongoing study and evaluation of an automated decision system or augmented critical decision process and its impact on consumers.	Sim	To direct the Federal Trade Commission to require impact assessments of automated decision systems and augmented critical decision processes, and for other purposes.
Assembly Bill 331 on Automated Decision Tools (California)	Sim	22756.1. (a) On or before January 1, 2025, and annually thereafter, a deployer of an automated decision tool shall perform an impact assessment for any automated decision tool the deployer uses	Sim	Section 22756.7. (a) Within 60 days of completing an impact assessment required by this chapter, a deployer or a developer shall provide the impact assessment to the Civil Rights Department.
NIST	Sim	Risk management should be continuous, timely, and performed throughout the AI system lifecycle dimensions.	-	-
Conselho da Europa - CAI	Sim	Art. 15 (2) (e) Ensure that the risk and impact management processes are carried out iteratively throughout the design, development, use and decommissioning of the artificial intelligence system;	Sim	Article 25 – Effective oversight mechanisms 1. Each Party shall establish or designate one or more effective mechanisms to oversee and supervise compliance with the obligations in the Convention, as given effect by the Parties in their domestic legal system. 2. Each Party shall ensure that such mechanisms exercise their duties independently and impartially and that they have the necessary powers, expertise and resources to effectively fulfill their tasks of overseeing compliance with the obligations in the Convention, as given effect by the Parties in their domestic legal system.

OCDE	Sim	"AI actors should, based on their roles, the context, and their ability to act, apply a systematic risk management approach to each phase of the AI system lifecycle on a continuous basis to address risks related to AI systems, including privacy, digital security, safety and bias"	-	-
UNESCO	Sim	Os Estados-membros e as empresas devem implementar medidas adequadas para acompanhar todas as fases do ciclo de vida dos sistemas de IA, incluindo o funcionamento dos algoritmos utilizados para tomada de decisão, os dados, bem como os atores de IA envolvidos no processo, especialmente em serviços públicos e onde for necessária a interação direta do usuário final, como parte da avaliação de impacto ético.	-	Essas avaliações também devem ser multidisciplinares, multiparceiros, multiculturais, pluralistas e inclusivas. As autoridades públicas devem ser obrigadas a monitorar os sistemas de IA implementados e/ou utilizados por essas autoridades, introduzindo mecanismos e ferramentas adequados.

Normativa	Casos obrigatórios de elaboração da AIA	Metodologia	Momento de realização	Há critérios de análise?
PL 21/20	Não há menção à AIA, apenas a avaliação de impacto regulatório	-	-	-
PL 759/23	Quando o sistema for considerado de alto risco pela avaliação preliminar.	Metodologia baseada em riscos e direitos. Definição de, no mínimo, 4 etapas: preparação, cognição do risco; mitigação dos riscos encontrados e monitoramento.	Prévio com a possibilidade/ obrigatoriedade de revisões periódicas.	Sim, no § 1º do art. 24.

Canada's Artificial Intelligence and Data Act (AIDA)	Para sistemas de IA de alto risco.	Metodologia baseada em riscos (não é expresso). Apenas determina a necessidade de identificação, avaliação e mitigação de riscos de danos ou resultados enviesados previamente.	Previamente.	-
Canada's Algorithmic Impact Assessment tool	Para sistemas de decisão automatizada no âmbito da Administração Pública.	Metodologia baseada em risco. Apenas determina a necessidade de identificar os riscos e mitigá-los. Mas sua realização está disponível em um questionário aberto no Portal do Governo.	No começo da fase de design do projeto e posteriormente antes da produção do sistema para validar os resultados previamente obtidos.	Sim, há tabelas com perguntas orientadoras para a identificação das áreas de risco e para a identificação das possíveis medidas de mitigação, além de o tópico 3.2 trazer pontos que devem ser considerados no momento de realização da AIA.
Algorithmic Accountability Act EUA	Para sistemas de decisão automatizados e processos de decisão críticos ampliados (processos, procedimentos ou outras atividades que utilizam decisões automatizadas para tomar decisões críticas) das entidades cobertas, de acordo com a seção 2(7).	Não determina uma metodologia precisa, mas parece ser baseada em riscos.	Antes e depois da implantação do sistema.	Sim, definidos na Seção 4.
Assembly Bill 331 on Automated Decision Tools (California)	Ferramentas de decisões automatizadas que preenchem os critérios da Section 22756.1.	Metodologia baseada no risco	Anualmente e, assim que possível, a qualquer atualização significativa.	Sim, definidos na 22756.1. (a) (b)
EU AI Act (Versão do Parlamento Europeu)	Sistemas de IA de alto risco	Metodologia baseada no risco	Antes de sua utilização e em qualquer outro momento em que o implementador considerar que há novos critérios de análise.	Sim, definidos no artigo 29a (1).

CAHAI - Human Rights, Democracy and Rule of Law Impact Assessment of AI systems	Não menciona a figura de uma avaliação de impacto, mas fala de gestão e avaliação de riscos	Metodologia baseada em riscos	-	-
CAI - Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law (resumo consolidado)	Não define os casos obrigatórios	Metodologia baseada no riscos (art. 15, 2)	Deve ser realizado de forma iterativa ao longo da concepção, desenvolvimento, utilização e desmantelamento do sistema de IA.	Sim, definidos no art. 15 (2)
OCDE[100]	Não menciona a figura de uma avaliação de impacto, mas fala de gestão e avaliação de riscos.	Metodologia baseada em riscos	Antes ("AI in the lab") e depois ("AI in the field") de seu uso/ implantação.	Independentemente do número de níveis de risco, critérios típicos para determinar o nível de um sistema de IA incluem sua escala (gravidade dos impactos adversos (e probabilidade), escopo (amplitude de aplicação, como o número de indivíduos que são ou serão afetados) e opcionalidade (grau de escolha quanto a estar sujeito aos efeitos de um sistema de IA).
UNESCO - Recomendação sobre a Ética da Inteligência Artificial + Ferramenta de Avaliação de Impacto Ético de IA	Sessão específica sobre "avaliação de impacto ético", em que define que "os Estados-Membros devem criar marcos para a realização de avaliações de impacto, como avaliação de impacto ético, para identificar e avaliar os benefícios, as preocupações e os riscos dos	Define que os Estados-Membros devem adotar um marco normativo que estabeleça procedimento, especial para autoridades públicas.	Preferencialmente antes do lançamento da tecnologia no mercado, mas aplicado em todo o seu ciclo de vida.	Não há previsão de critérios especificamente, mas fala-se em "identificar impactos nos direitos humanos e nas liberdades fundamentais, especialmente, mas não apenas, nos direitos de pessoas marginalizadas e vulneráveis ou pessoas em situações vulneráveis, direitos

100 https://www.oecd-ilibrary.org/docserver/2448f04b-en.pdf?expires=1691001152&id=id&accname=guest&checksum=4B452E3AB7BD695B35EF6D45563DC6B6; https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1691002504&id=id&accname=guest&checksum=C786D4EDC16B1C3E6620F6617CCF0ADB.

	sistemas de IA, bem como medidas adequadas de prevenção, mitigação e monitoramento de riscos", mas não define em quais hipóteses haveria essa obrigatoriedade.			do trabalho, meio ambiente e ecossistemas e suas implicações éticas e sociais, e facilitar a participação dos cidadãos".

Nota-se, portanto, que um dos fios-condutores em comum das regulações de IA é a previsão de uma procedimentalização mínima da ferramenta de avaliação de impacto algorítmico. Diferentemente das diretrizes da UNESCO e do Conselho da Europa, o PL 2338/23, assim como a proposta europeia e canadense, é ainda tímido acerca do componente dos possíveis efeitos adversos sobre direitos sociais como relacionados a trabalho e meio ambiente. Ainda assim, é a única proposta brasileira interoperável com todas as demais no sentido de prever e empregar densidade normativa mínima para seu florescimento ao prever um piso metodológico, momento de análise e possível revisão e critérios.

b.2) Publicidade

Além da definição metodológica, um dos aspectos essenciais das avaliações de impacto, inclusive no âmbito da IA, é a definição da obrigatoriedade (ou não) de sua publicização. Como supramencionado, Kaminski, em seus estudos sobre as formas de regulação de risco, aborda esta regulação enquanto atrelada à ideia de supervisão democrática ou de prestação de contas democrática[101]. Em outras palavras, a avaliação do risco, dentro do processo de construção de uma avaliação de impacto, serviria de instrumento para discussão pública desses riscos, que passariam a ser compartilhados com toda a sociedade.

Nesse âmbito, a possibilidade de acesso do público às análises ou aos principais resultados do processo de avaliação de impacto algorítmico permitiria que o gerenciamento dos riscos dos sistemas de IA fossem sujeitos ao escrutínio público, o que garantiria que os agentes de IA prestassem contas não apenas com as autoridades competentes para supervisão, mas com toda a sociedade, principalmente os indivíduos impactados por ela, podendo servir, inclusive, como base potencial de futuras intervenções políticas substantivas (como a atualização de uma eventual listagem de sistemas de IA de alto risco, por exemplo).

De acordo com a Recomendação sobre a Ética da Inteligência Artificial da UNESCO, as avaliações de impacto (ético) devem ser transparentes e abertas ao público, quando for apropriado. Nesse sentido, a publicização das análises da avaliação de impacto ou de suas principais conclusões estaria alinhada à ideia de uma transparência qualificada[102], facilitando os processos de prestação de contas pelos agentes da cadeia produtiva de IA e a posterior fiscalização por eventual autoridade competente e pelos indivíduos e grupos impactados pelo sistema, inclusive reduzindo as assimetrias de informação e poder. Por esta linha, todas as partes interessadas podem adquirir o entendimento e a capacidade de influenciar os processos de tomada de decisão de criação e implementação de sistemas de IA, a partir da ideia de justiça procedimental e de devido processo informacional, o que também dá maior legitimidade ao processo[103].

A publicização da avaliação de impacto algoritmo pode se dar tanto a partir da disponibilização de todo o seu conteúdo como de suas principais conclusões[104]. Além disso,

101 KAMISNKI, 2022, p. 36.

102 BIONI; LUCIANO, 2019, p. 3.

103 DARIUSZ, Kloza. *Privacy Impact Assessment as a Means to Achieve the Objectives of Procedural Justice*, Jusletter IT. Die Zeitschrift für IT und Recht, disponível em: https://cris.vub.be/files/49868387/Kloza_2014_PIA_as_a_Means_to_Achieve_the_Objectives_of_Procedural_Justice.pdf; CITRON, Danielle, PASQUALE, Frank. The Scored Society: Due Process for Automated Predictions. Washington Law Review, Vol. 89, 2014.

104 No caso do PL 2338/2023, essa publicidade também pode se dar em relação ao processo prévio de análise representado pela avaliação preliminar do art. 13 no que tange aos sistemas de IA no âmbito do poder público, independentemente do grau de risco (art. 21).

os casos de obrigatoriedade de publicização de tal ferramenta podem variar de acordo com, por exemplo, o grau de risco de determinado sistema de IA, contexto de uso ou com o tipo de agente de IA (setor público ou privado). A título exemplificativo, no campo da proteção de dados, tanto o GDPR como a LGPD não definem a obrigatoriedade de publicação dos resultados da avaliação ou relatório de impacto, respectivamente, o que é apenas sugerido no primeiro caso e pode ser feito como boa prática de *accountability* do agente regulado tanto na LGPD como no GDPR.

Nas atuais propostas regulatórias para a IA mundialmente, no que tange à publicização da AIA, destaca-se o projeto de lei 5116 do legislativo do Estado norte-americano de Washington, que busca estabelecer critérios para a compra e utilização de sistemas de decisão automatizada pelo Estado. Diferentemente do que acontece no GDPR e na LGPD, a seção 5 determina expressamente a publicação da avaliação de impacto pelo órgão competente (intitulado pela proposta de "Escritório de Revisão de Prestação de Contas Algorítmica") em seu site oficial, além de estabelecer a abertura de processo para recebimento de comentários do público antes de sua aprovação pelo prazo não superior a 30 dias[105].

Também nos Estados Unidos, o *Algorithmic Accountability Act* de 2022, destinado a sistemas de tomada de decisão automatizada, segue uma lógica diferente, já que a seção 4 (c) expressamente dispõe sobre a não obrigatoriedade de divulgação da avaliação de impacto, apesar da previsão de que seja enviado, antes da implementação e continuamente, um relatório resumido destas avaliações para a entidade competente de fiscalização e supervisão (seção 3 (b) (1) (d) (e)). Desta forma, apesar da ausência de disponibilização pública do conteúdo ou resumo da avaliação para o público em geral, esta iniciativa regulatória cria certa responsabilização pública ao determinar o compartilhamento do resumo da análise para o órgão competente[106].

De forma similar ao projeto de Washington, apesar de criar limites referentes aos segredos industrial e comercial, o projeto de lei brasileiro 2338/2023 expressamente prevê no art. 26[107] a publicização das conclusões da avaliação de impacto, determinando um

105 Texto da Seção 5 (3): "(3) An agency intending to develop, procure, or use an automated decision system for implementation after January 1, 2024, must submit an algorithmic accountability report to the applicable algorithmic accountability review office and obtain approval or conditional approval prior to any use of the automated decision system. The algorithmic accountability review office must post the algorithmic accountability report on the algorithmic accountability review office's public website and invite public comment on the algorithmic accountability report for a period of no less than 30 days".

106 Kaminski (2022, p. 73 e 74) critica a falta de divulgação do núcleo central das avaliações de impacto, de forma a mantê-lo apenas interno às empresas, não sendo divulgado para reguladores, partes interessadas, especialistas ou para o público. Nesta "ausência de responsabilidade pública", a autora questiona se as entidades reguladas irão realmente mitigar os riscos de seus sistemas, mantendo-se cética em relação a isso.

107 Texto do art. 26 do PL 2338/2023: "Art. 26. Garantidos os segredos industrial e comercial, as conclusões da avaliação de impacto serão públicas, contendo ao menos as seguintes informações: I – descrição da finalidade pretendida para a qual o sistema será utilizado, assim como de seu contexto de uso e escopo territorial e temporal; II – medidas de mitigação dos riscos, bem como o seu patamar residual, uma vez implementada tais medidas; e III – descrição da participação de diferentes segmentos afetados, caso tenha ocorrido, nos termos do § 3º do art. 24 desta Lei".

conteúdo mínimo a ser disponibilizado ao público, o que inclui, por exemplo, a descrição da finalidade do sistema, seu contexto de uso e escopo territorial e temporal; medidas de mitigação de riscos adotadas; e a descrição dos diferentes segmentos afetados. Ainda, considerando a assimetria de poder nas relações envolvendo Estado e sociedade, o projeto prevê, no inciso VI do art. 21, medidas extras para o Poder Público no que se refere à publicização das avaliações preliminares de sistemas de IA desenvolvidos, implementados ou utilizados nesse contexto, independentemente do grau de risco.

Juntando-se ao PL 2338/2023 (Brasil) e ao projeto de lei 5116 (EUA - Washington), o EU AI Act, em sua versão de junho de 2023 publicada pelo Parlamento Europeu, também incorporou esta obrigação de publicidade da avaliação de impacto (no caso da proposta, de direitos fundamentais). No caso europeu, essa regra é restrita ao resumo da avaliação de impacto nos casos de sistemas implantados por autoridade pública ou determinadas organizações consideradas "controladoras de acesso" ("gatekeepers") pelo Regulamento (EU) 2022/1925 (Regulamento dos Mercados Digitais) quando consideradas implantadoras (*deployer*). Tais informações devem ser postadas em uma base de dados pública de sistemas de IA de alto risco.

Ainda no caso do PL 2338, há também a previsão de criação de uma base de dados brasileira relativa a sistemas de inteligência artificial, que poderá conter tanto a documentação de autoavaliação dos sistemas de IA, quanto as avaliações de impacto de IA de alto risco. Existem ao menos dois objetivos que são cumpridos pela criação da base de dados. O primeiro é transparência e redução da assimetria informacional em relação aos possíveis impactados pelos sistemas de IA, que tem acesso facilitado às informações relevantes para avaliar riscos aos direitos individuais e coletivos, e ter conhecimento sobre quais sistemas de IA afetam sua vida e dia-a-dia. O segundo é para os próprios fornecedores, que podem se utilizar da base de dados para verificar as melhores práticas em relação à confecção de relatório de impacto, promovendo uma cultura de compartilhamento e *benchmarking*[108].

A criação de uma base de dados acessível ao público com informações sobre sistemas de IA não é uma inovação brasileira, já que é proposta também, como mencionado previamente, nos projetos do *Artificial Intelligence Act* da União Europeia, no projeto de lei 15869-19 do Chile e do *Algorithmic Accountability Act* dos Estados Unidos.

108 WRIGHT *et al.*, 2014, p.165.

Normativa	Previsão de publicização da AIA	O que deve ser publicado?	Onde	Previsão de banco de dados público	Onde
PL 2338/2023	Não	-	-	Não	-
PL 759/23	Não	-	-	Não	-
PL 2338/23	Sim	Para todos os sistemas de IA de alto risco: as principais conclusões da AIA; Para sistemas de IA do poder público: todas as avaliações preliminares, independentemente do grau de risco.	Art. 26; Art. 21.	Sim	Art. 43.
Directive on Automated Decision-Making + Algorithmic Impact Assessment tool	Sim	Para sistemas de decisões automatizadas no âmbito da Administração Pública: publicização dos resultados finais em formato acessível em inglês e francês no Portal Aberto do Governo.	Página do Governo do Canadá; Art. 6.1.4 da Diretiva	Sim	Página do Governo do Canadá; Art. 6.1.4 da Diretiva
Washington SB 5116 - 2021-22	Sim	-	Seção 5 (3)	Sim	Seção 6
Algorithmic Accountability Act EUA	Sim, em parte	Apenas determina o envio de relatório resumido das avaliações para a Comissão competente de fiscalização e supervisão.	Seção 3 (b) (1) (d) (e) e seção 4 (c)	Sim	Seção 6
Assembly Bill 331 on Automated Decision Tools (California)	Sim, em parte	Apenas determina o envio da avaliação de impacto para o Departamento de Direitos Civis	Seção 22756.7. (a)	Não	-
EU AI Act (Versão do Parlamento Europeu)	Sim	Apenas a publicação do sumário dos resultados da avaliação nos casos de sistemas implantados por autoridade pública ou determinadas organizações consideradas "controladoras de acesso" ("gatekeepers") pelo Regulamento (EU) 2022/1925 (Regulamento dos Mercados Digitais)	Artigo 29a (5)	Sim	Artigo 51 e 60

Conselho da Europa - CAI[109]	Sim	Quando apropriado, publicação de informações sobre os esforços para identificar, avaliar, mitigar e prevenir riscos e impactos adversos empreendidos.	Artigo 15 (2) (g)	Não	-
OCDE[110]	Sim	Quando apropriado (não há maiores definições), mas inclui exemplos do que pode ser publicado: quais mecanismos de governança foram utilizados, como os riscos são monitorados e revistos, quais mecanismos existem para reparação, entre outros.	Página 50 e 51 do Relatório "Advancing accountability in AI"	Não	-
UNESCO[111]	Sim - as avaliações de impacto devem ser transparentes e abertas ao público, quando for apropriado.	Quando for apropriado (não define).	Página 26, parágrafo 53 das Recomendações	Não menciona	-
Blueprint for an AI Bill of Rights	Sim	Sempre que possível, fornecidos de forma clara e legível por máquina, utilizando linguagem simples.	Página 5 e 28	Não menciona	-

109 Consolidated Working Draft of the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law – Committee on Artificial Intelligence (CAI). Strasburg, 7 de julho de 2023.

110 OECD. Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI. 23 Feb. 2023. Disponível em: https://www.oecd-ilibrary.org/science-and-technology/advancing-accountability-in-ai_2448f04b-en.

111 UNESCO. Recommendation on the Ethics of Artificial Intelligence. 16 May 2023. Disponível em: https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.

b.3) Modelo participativo-democrático

Quando falamos de avaliação de impacto para determinadas tecnologias, serviços ou produtos, o comando legal pode trazer diferentes níveis de participação e engajamento públicos. Este envolvimento das partes interessadas pode trazer vários benefícios para o processo de avaliação (por exemplo, aumentar sua qualidade, credibilidade e legitimidade) e para o resultado (por exemplo, o processo de tomada de decisão ser mais bem informado e representativo)[112]. Isso faz com que eventuais decisões tomadas com base nas avaliações de impacto não sejam fruto de uma análise restrita a um grupo seleto de atores interessados, especialmente internos à organização, o que poderia gerar enviesamento e discriminação, mas de agentes diversos, inclusive externos, especialmente considerando os potencialmente impactados pela adoção da tecnologia[113].

Segundo a UNESCO, a elaboração da AIA requer o envolvimento de uma série de indivíduos, representantes e comunidades potencialmente afetados, o que pode ser feito por meio de consultas multilaterais proporcionais à escala e âmbito do sistema, à sua urgência e aos impactos esperados[114]. Esta participação diversa e ampla de agentes externos ao agente regulado permite que este receba críticas e apontamentos de possíveis impactos que não foram pensados antes do lançamento do produto/tecnologia, além de permitir um relacionamento transparente entre impactados (presentes ou futuros) e o agente, sendo um instrumento de redução da assimetria de informação[115].

Dito isso, é fundamental que a participação seja efetiva, e para tanto, deve ser disponibilizada documentação preliminar da avaliação de impacto aos *stakeholders*, para que possam proceder a sua própria avaliação[116]. Assim, garante-se o máximo possível que o processo de elaboração da avaliação de impacto (e não apenas seu resultado) seja justo, o que cria legitimidade, já que as pessoas tendem a confiar mais nas decisões quando sabem que não foram tomadas à porta fechada, mas envolvendo pessoas como elas, além de especialistas[117].

No campo de IA, a participação pública nos processos de avaliação de impacto é ainda mais importante nas hipóteses que envolvem a necessidade de tomada de decisões difíceis, o que acontece nos casos de sistemas de alto risco, já que potencialmente relacio-

112 KLOZA *et al*, 2019.

113 BIONI, Bruno; EILBERG, Daniela Dora; CUNHA, Brenda; SALIBA, Pedro; VERGILI, Gabriela. Proteção de dados no campo penal e de segurança pública: nota técnica sobre o Anteprojeto de Lei de Proteção de Dados para segurança pública e investigação criminal. São Paulo: Associação Data Privacy Brasil de Pesquisa, 2020. p. 8–9.

114 UNESCO, 2023, p. 43.

115 WRIGHT *et al*., 2014, p. 160.

116 WRIGHT *et al*., 2014, p. 170.

117 ECNL; Society Inside. Framework for Meaningful Engagement. Disponível em: https://ecnl.org/sites/default/files/2023-03/Final%20Version%20FME%20with%20Copyright%20%282%29.pdf.

na-se com implicações relevantes para a direitos fundamentais, seja de grupos marginalizados ou da sociedade em geral[118]. Nesse ponto, é ainda mais eminente a necessidade de que a participação não seja apenas um item de *checklist*, mas significativa.

Para tal, de acordo com um estudo sobre Participação Pública Significativa desenvolvido pela *European Center for Not-for-Profit Law Stichting*[119], o processo participativo deve considerar três elementos essenciais para ser, de fato, significativo. São eles: (i) propósito compartilhado, isto é, o propósito deve ir além do interesse do próprio órgão convocador, mas inclui interesses dos potencialmente afetados ou um propósito geral de interesse público; (ii) processo confiável, ou seja, inclusivo, aberto, justo e respeitoso, com o mínimo de barreiras à entrada; e (iii) impacto visível, no sentido de que o envolvimento das partes terá o poder de contribuir significativamente para a tomada de decisões ou introduzir alterações na governança da organização, do produto ou do serviço de IA para alinhá-la com o interesse público.

Quanto mais significativa e inclusiva, o envolvimento das partes interessadas é particularmente eficaz, tanto para compreender potenciais problemas ou oportunidades de produtos ou serviços que utilizam IA, como para identificar possíveis impactos, implicações, benefícios e danos específicos, positivos ou adversos, desses produtos ou serviços sobre os direitos humanos individuais e coletivos, especialmente considerando a inclusão de grupos marginalizados e já vulneráveis[120], o que permite a criação de sistemas mais adequados para as realidades sociais-alvo e com maior controle por parte da população.

Para melhor compreender tal dinâmica, podemos pegar emprestado do campo da defesa dos direitos das pessoas com deficiência a frase *"nada sobre nós sem nós"* para ressaltar a importância da participação significativa da sociedade na avaliação das ferramentas de IA, inclusive para possibilitar controle democrático e social dos agentes[121]. Isso vai ao encontro do princípio da governança multiparticipativa, que deve ser colocado em prática para que a sociedade não seja só um agente passivo da tecnologia, mas que possa atuar em seu desenvolvimento, especialmente nos casos de tecnologias que supostamente lhe impactará.

De acordo com Kaminski (2023, p. 79), em razão de os riscos dos sistemas de IA serem de diferentes níveis de desconhecimento, inquantificáveis e socialmente contestáveis, a participação dos diferentes atores e partes interessadas no processo de avaliação da tecnologia é um aspecto crucial para a adequada regulação do risco de IA. Ademais,

118 Ibid.

119 ECNL; Society Inside. Framework for Meaningful Engagement. Disponível em: https://ecnl.org/sites/default/files/2023-03/Final%20Version%20FME%20with%20Copyright%20%282%29.pdf.

120 Ibid.

121 COSTANZA-CHOCK, Sasha. Design Practices: "Nothing about Us without Us". Design Justice, publicado em 26 fev. 2020. Disponível em: https://designjustice.mitpress.mit.edu/pub/cfohnud7/release/4.

esta inclusão no processo de avaliação dos impactos da tecnologia, a partir do risco em que ela foi enquadrada, permite também que esses grupos façam a defesa mais ativa de seus direitos, o que está de acordo com uma regulação que segue abordagem baseada em riscos e em direitos, sendo uma ferramenta importante para as democracias modernas[122].

Para além de especialistas técnicos, é fundamental que haja participação daqueles que são potencialmente os mais prejudicados pela aplicação de determinado sistemas de IA, pois são elas que saberão descrever os reais riscos e impactos relacionados à realidade prática. Essa inclusão é, inclusive, necessária para reequilibrar a dinâmica desbalanceada de poder entre as organizações que constroem tecnologias automatizadas e as pessoas que as utilizam e são afetadas por elas[123].

Assim, para o escopo de uma avaliação de impacto em IA, é indispensável entender as partes interessadas (*stakeholders*) de forma ampla, incluindo agentes internos ou externos ao agente de IA, como o público (leigos), tomadores de decisão, especialistas, entidades da sociedade civil, pesquisadores da academia e todos aqueles que podem estar (hoje) ou estarão (no futuro) impactados ou impactando o sistema de IA em causa, especialmente grupos vulneráveis e minorias sociais.

Ademais, para fins de prestação de contas, é essencial também o registro da participação desses stakeholders e das sugestões dadas para melhoria do sistema de IA, permitindo a consulta posterior para verificação da efetividade da participação e por outros fornecedores interessados e com possíveis impactos semelhantes em seus sistemas de IA. Quando for necessária a renovação da avaliação de impacto, deve haver novamente a participação pública, ainda que de maneira simplificada, a depender do nível de mudanças ocorridas entre a consulta inicial e o momento da renovação.

Como veremos nas tabelas abaixo, a participação pública efetiva durante todo o processo de elaboração de avaliações de impacto algorítmico para sistemas de IA, especialmente quando de alto risco, é uma necessidade defendida em diferentes leis nacionais, propostas de regulação e sugestões de entidades internacionais, de forma a caminharmos em direção à criação de processos de avaliação de impacto inclusivos e permeáveis à participação pública e cidadã.

Em âmbito internacional, o CAHAI do Conselho da Europa, em estudo de avaliação sobre a avaliação de impacto em Direitos Humanos, Democracia e Estado de Direito, constatou que o engajamento da comunidade é essencial para o sucesso dessa ferramen-

122 BAROCAS, Solon; VECCHIONE, Briana; LEVY, Karen. Algorithmic Auditing and Social Justice: Lessons from the History of Audit Studies. EAAMO '21, October 5–9, 2021, –, NY, USA. Disponível em: https://dl.acm.org/doi/pdf/10.1145/3465416.3483294. p. 2.

123 Data & Society. Algorithmic Impact Methods Lab. Data & Society Announces the Launch of its Algorithmic Impact Methods Lab. Nova York, 10 mai. 2023. Disponível em: https://datasociety.net/algorithmic-impact-methods-lab.

ta quando aplicável no contexto de IA. Para tanto, é importante que haja a definição de mecanismos eficientes para a identificação dos *stakeholders* dentro das comunidades, da forma mais inclusiva possível, e, a partir disso, produzir participação significativa ativa nos processos de avaliação dos sistemas[124].

A necessidade de participação pública (interna e externa) diversa, incluindo especialistas, sociedade civil e comunidades afetadas (inclusive as sem conhecimento técnico), em todo o ciclo de vida de IA, é também ressaltada em outros documentos internacionais, como no âmbito da OCDE e da UNESCO, assim como por frameworks estadunidenses publicados pelo *White House Office of Science and Technology Policy*[125] e pelo *National Institute of Standards and Technology* (NIST)[126].

No cenário brasileiro, no que tange aos projetos de lei 5051/19, 21/20, 871/21 e 579/23, devido à característica de generalidade, não há previsão de quaisquer instrumentos de participação pública democrática ao longo do ciclo de vida dos sistemas de IA. Já no PL 2338/23, há expressa menção a essa participação pública qualificada em diferentes momentos do texto sugerido (§3º do art. 24, §2º do art. 25, art. 26, III e §2º, alínea c, do art. 30). A possibilidade de supervisão democrática, por exemplo, é presente no art. 26, ao impor a publicação das conclusões da avaliação de impacto, no §2 do artigo 25, ao determinar a participação pública na atualização de avaliação de impacto, a partir de consulta às partes.

No Brasil, esta participação deve ser o mais inclusiva possível, incluindo não apenas especialistas, mas vozes de diferentes grupos sociais, especialmente de grupos vulneráveis (desde negros até povos tradicionais) para que também sejam considerados os aspectos culturais, o conhecimento e outras características distintivas nestas avaliações, de forma a evitar o reforço da condição de sub-representação, o apagamento étnico e epistemicídio[127].

A previsão de participação pública democrática nos processos de avaliação de impacto dos sistemas de IA também aparecem em projetos de regulação ou regulações já em vigor vindos da União Europeia, dos Estados Unidos e do Canadá.

124 CAHAI. Human Rights, Democracy and Rule of Law Impact Assessment of AI systems. Strasburgo, 11 mar. 2021. Conselho da Europa, CAHAI-PDG (2021)02. p. 15.

125 Blueprint for an AI Bill of Rights. Disponível em: https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.

126 Artificial Intelligence Risk Management Framework (AI RMF 1.0). Disponível em: https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf.

127 Epistemicídio é o termo criado por Boaventura de Sousa Santos para explicar processos de invisibilização e ocultação das contribuições sociais e culturais não assimiladas pelo conhecimento ocidental, como fruto de estruturas coloniais-capitalistas e de dominação imperialista especialmente de povos africanos e indígenas.SANTOS, Boaventura de Sousa. Construindo as Epistemologias do Sul: Antologia Essencial. Volume I: Para um pensamento alternativo de alternativas. Coleção Antologias do Pensamento Social Latino-Americano e Caribenho, 1ª Ed, 2018.

Normativa	Previsão de participação pública	Em quais termos?	Onde?	Obrigação reforçada para o poder público?	Onde?	Previsão de auditoria externa?	Como?	Onde?
PL 21/20	Não	-	-	-	-	Não	-	-
PL 759/23	Não	-	-	-	-	Não	-	-
PL 2338/23	Sim	(i) Não define os envolvidos, menciona "diferentes segmentos sociais afetados" e "partes interessadas"; (ii) Prevê essa participação apenas no momento de atualização	Art. 24, § 3º Art. 25, § 2º	Sim	Art. 21, I e IV	Sim	Caberá à autoridade competente regulamentar	Art. 23, parágrafo único
GDPR UE	Sim	Se for adequado, o responsável pelo tratamento solicita a opinião dos titulares de dados ou dos seus representantes sobre o tratamento previsto.	Art. 35 (9)	Sim	Considerando 93	Sim	Menciona a possibilidade de existência de auditorias, mas não as regulamenta.	Art. 28 (3) (h); 39 (1) (b); 47 (1) (j) e 58 (1) (b)
California AB-2261 (2019-2020)	Não	Antes de finalizar e implementar seu "relatório de prestação de contas", autoridades públicas devem considerar as questões	Section 1798.335. (e) (f) (g)	Sim	Section 1798.335.	Section 1798.365. (a)	Previsão de auditoria externa feita pelo Auditor do Estado da Califórnia	Section 1798.370

California AB-2261 (2019-2020)		trazidas pelo público em período de revisão pública e comentários, além de reuniões de consulta comunitária durante o período de revisão pública. Tal obrigação também está presente nas revisões bienais feitas ao relatório de prestação de contas.						
Canada: Directive on Automated Decision-Making + Algorithmic Impact Assessment tool	Sim	Apenas menciona a necessidade de consulta às partes interessadas internas e externas, incluindo consultores jurídicos e de privacidade; equipes de política digital; e especialistas no assunto de outros setores.	Página do Governo do Canadá que descreve a ferramenta	Não	A diretiva é aplicável para autoridades públicas	Sim	Apenas menciona a possibilidade de auditoria externa, nos casos aprovados pelo governo canadense.	Seção 6.2.5.2 e Section 6.2.5.3 da Diretiva
Washington SB 5116 - 2021-22	Sim	Menciona o recebimento de comentários públicos + determina que o relatório de	Seção 5 (3) e (6) (j) (v)	Não	Projeto direcionado para o poder público.	Sim	Prevê a possibilidade de realização de auditorias por agências ou	Seção 3(4) (b)

California AB-2261 (2019-2020)		prestação de contas algorítmica deve incluir descrição de qualquer envolvimento público ou comunitário realizado, além de quaisquer planos futuros de envolvimento público ou comunitário em conexão com o sistema de decisão automatizado.					terceiros independentes para compreender os impactos de sistemas de tomada de decisão automatizada, incluindo possíveis distorções e imprecisões ou impactos díspares;	
Algorithmic Accountability Act EUA	Sim	Determina a consultar significativamente (inclusive por meio de design participativo, auditoria independente ou solicitação ou incorporação de feedback) com partes interessadas internas relevantes (como funcionários, equipes de ética e equipes de tecnologia responsáveis) e partes interessadas externas inde-	Seção 3, (b) (1) (g) e Seção 4 (a) (2)	Não	-	Sim	Menciona a possibilidade de auditoria independente	Seção 3, (b) (1) (g)

Algorithmic Accountability Act EUA		pendentes (como representantes e defensores de grupos impactados, civis sociedade e defensores, e especialistas em tecnologia) com a frequência necessária.						
EU AI Act (Versão do Parlamento Europeu)	Sim	Prevê o envolvimento de representantes das pessoas ou grupos que possam ser afetados pelo sistema de IA de alto risco, com o objetivo de receber contribuições para a avaliação de impacto, tendo um período de seis semanas para que os interessados respondam. Excepciona dessa obrigação as pequenas e médias empresas.	Art. 29a (4)	Não	-	Sim	Prevê a realização de auditorias independentes, especialmente no caso da análise de cumprimento das regras do sistema de gestão de qualidade.	Considerando 60h, Art. 29 (5), Art. 70 (1) (b), Anexo VII (5.3)

Conselho da Europa - CAI[128]	Sim	Integrar a perspectiva de todos os atores interessados, incluindo quaisquer pessoas que podem ter seus direitos potencialmente afetados pelo design, desenvolvimento, uso ou descontinuação do sistema de IA.	Art. 15 (2) (c) e art. 19	Não	-	Não	-	-
OCDE[129]	Sim	Consulta a partes interessadas (internas e externas), incluindo a sociedade civil e comunidades afetadas (até mesmo sem conhecimento técnico), para obter feedback e conhecimentos para alimentar as avaliações de impacto e risco, bem como gerir esse risco	Página 52 do Relatório "Advancing accountability in AI"	Não	-	Sim	Em geral, as auditorias de IA envolvem cientistas e engenheiros de dados, modelos e sistemas e especialistas em governança, tanto internos como externos. Nesse tópico, reforçam que as características das equipes	Menções à auditoria e à sua necessidade são feitas por todo o Relatório "Advancing accountability in AI", destaque para página 24 e 47 e seguintes

128 Consolidated Working Draft of the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law – Committee on Artificial Intelligence (CAI). Strasburg, 7 de julho de 2023.

129 OECD. Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI. 23 Feb. 2023. Available at: https://www.oecd-ilibrary.org/science-and-technology/advancing-accountability-in-ai_2448f04b-en.

		em cada parte do processo. A consulta deve ocorrer em todas as fases do ciclo de vida do sistema de IA. O formato, custo e frequência das comunicações e consultas devem ser avaliados com base no contexto.					de auditores (como gênero, país e histórico cultural) impactam a avaliação da justiça dos resultados do sistema de IA, o que justifica a defesa pela diversidade das equipem que fazem essas auditorias. Diferentes níveis de acesso poderiam permitir auditorias e análises adaptadas a um sistema específico de IA e ao seu contexto.	
UNESCO[130]	Sim	Devem ser transparentes e abertas ao público, quando for apropriado (não define).	Página 26, parágrafo 53.	Não	-	Não	-	-

130 UNESCO. Recommendation on the Ethics of Artificial Intelligence. 16 May 2023. Available at: https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.

Blueprint for an AI Bill of Rights	Sim	Sistemas automatizados devem ser desenvolvidos com consulta de diversas comunidades, partes interessadas e especialistas do domínio para identificar preocupações, riscos e impactos potenciais do sistema.	Página 15	Não	-	Sim	Não há detalhamento, mas prevê a realização de avaliações independentes por terceiros.	Mencionado algumas vezes, como nas páginas 20, 21, 24, 38 e 57.
Artificial Intelligence Risk Management Framework (AI RMF 1.0) - NIST	Sim	Identificar e gerir os riscos e potenciais impactos da IA requer um amplo conjunto de perspetivas e intervenientes em todo o seu ciclo de vida. Idealmente, os intervenientes na IA representarão uma diversidade de experiências, conhecimentos e antecedentes e compreenderão equipes demograficamente e disciplinarmente diversas.	Página 9-10 e 29-31	Não	-	Sim	Apenas menciona a possibilidade de auditoria	Página 16 e 35

		Especialistas, usuários, atores de IA externos à equipe que desenvolveu ou implantou o sistema de IA e comunidades afetadas são consultados para apoiar avaliações, conforme necessário.						
Executive Order on Safe, Secure and Trustworthy Development and Use of AI (EUA)	Sim	No âmbito do uso de IA pelo poder público federal, passa a ser obrigatória a implementação de práticas mínimas, vindas do framework do NIST ou do Blueprint, de gestão de riscos de IA que impacte direitos ou a segurança das pessoas, expressamente mencionando a realização de consulta pública.	Seg. 10, 10.1, (b) (iv)	Sim	Seg. 10, 10.1, (b) (iv)	Sim	No âmbito do uso de decisões automatizadas para implementação de benefícios sociais, é garantida a realização de auditoria.	Sec. 7, 7.2, (b) (ii) (E)

EIXO 3 – IA Generativa

Em novembro de 2022, com o lançamento do ChatGPT pela *Open AI*, a IA generativa rapidamente ganhou o discurso público, o foi demonstrado pelo crescimento exponencial de pesquisa e investimento nesta área a partir deste período[131]. Junto com seus muitos benefícios, como ganhos de produtividade e eficiência e auxílio para solução de desafios sociais, vieram também os riscos, o que tornou urgente a discussão a respeito de sua regulação.Um primeiro desafio para endereçarmos as propostas de regulação da IA generativa é defini-la. Assim como existem múltiplas definições do que seria a Inteligência Artificial, a IA generativa padece da mesma falta de uma definição consensual.

Na proposta de terminologia firmada entre os Estados Unidos e a Europa, no documento *"EU-U.S Terminology and Taxonomy for Artificial Intelligence"*[132], que tem como objetivo "informar as visões para o gerenciamento de risco da Inteligência Artificial e Inteligência Artificial Confiável ("*Trustworthy*") nos dois lados do Atlântico e avançar iniciativas de colaboração na definição de padrões em órgãos internacionais relacionados a IA"[133], não há uma definição do termo IA Generativa, apenas de modelos grandes de linguagem (*large language model* ou LLM):

> Uma classe de modelos de linguagem que usam algoritmos de deep-learning e são treinados em bases de dados extremamente grandes que podem ter tamanho de múltiplos terabytes. LLMs podem ser classificados em dois tipos: generativos ou segregatórios. LLMs Generativos são modelos que tem como output texto, como a resposta a uma questão ou ainda escrever um ensaio sobre um tópico específico. Eles são tipicamente não supervisionados ou semi-supervisionados em relação ao seu modelo de aprendizagem e fazem uma previsão de qual a resposta para uma determinada tarefa. Modelos segregatórios são modelos de aprendizado supervisionado que focam usualmente em classificar texto, como determinar se um texto foi feito por um ser humano ou por uma inteligência

131 OECD. G7 Hiroshima Process on Generative Artificial Intelligence (AI): Towards a G7 Common Understanding on Generative AI. Relatório preparado para a presidência japonesa de 2023 e para o grupo de trabalho digital e tecnológico do G7. Publicado em 7 set. 2023. Disponível em: https://www.oecd-ilibrary.org/deliver/bf3c0c60-en.pdf?itemId=%2Fcontent%2Fpublication%2Fbf3c0c60-en&mimeType=pdf.

132 Comissão Europeia. EU-U.S. Terminology and Taxonomy for Artificial Intelligence. Publicado em 31 maio 2023. Disponível em: https://digital-strategy.ec.europa.eu/en/library/eu-us-terminology-and-taxonomy-artificial-intelligence.

133 Ibid, p.1.

artificial[134].

No contexto chinês, em julho de 2023, o *Cyberspace Administration of China* (CAC) publicou algumas regras para a IA generativa em documento nomeado "Medidas Provisórias para a Gestão de Serviços Generativos de Inteligência Artificial (*Interim Measures for the Management of Generative Artificial Intelligence Services*), cujos efeitos tornaram-se válidos a partir de 15 de agosto de 2023. No artigo 22 (1) do documento há a definição de "tecnologia de inteligência artificial generativa" como *"modelos e tecnologias relacionadas que têm a capacidade de gerar conteúdo como texto, imagens, áudio e vídeo"*[135].

Já no cenário brasileiro, os atuais projetos de lei não trazem definições ou regras para as IAs generativas. O PL 2338/23, apesar de não trazer uma conceituação menciona o termo "sistemas de inteligência artificial de propósito geral", que devem incluir em sua avaliação preliminar (a respeito do seu grau de risco) as finalidades ou aplicações indicadas (art. 13, §1º). Como já mencionado anteriormente nesse relatório, a categorização dos riscos da inteligência artificial no PL 2338/23 em excessivo e alto se dá a partir das finalidades e contexto de aplicação.

No cenário europeu, a primeira versão da proposta do AI Act foi publicada em abril de 2021, antes do lançamento dos mais famosos modelos de LLM, como o GPT 3.5 e GPT 3 da OpenAI. A primeira versão da proposta a iniciar as discussões sobre tais modelos foi a adotada pelo Conselho da UE em 6 de dezembro de 2022[136], que inseriu a definição de "IA de propósito geral" no artigo 3º (1b):

> Sistema de IA de uso geral é um sistema de IA que - independentemente da forma como é colocado no mercado ou colocado em serviço, inclusive como software de código aberto - é destinado pelo fornecedor a desempenhar funções de aplicação geral, como reconhecimento de imagem e de voz, áudio e geração de vídeo, detecção de padrões, resposta a perguntas, tradução e outros; um sistema de IA de uso geral pode ser usado em uma pluralidade de contextos e ser integrado em uma pluralidade de outros sistemas de IA[137].

134 Ibid, p.9. Tradução própria.

135 http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm. Tradução própria.

136 Council of the EU. Artificial Intelligence Act: Council calls for promoting safe AI that respects fundamental rights. Press Release, publicado em 6 dec. 2022. Disponível em: https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/.

137 Tradução própria: https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf.

Mais recentemente, na versão do documento do Parlamento Europeu (PE) de maio de 2023, a emenda 169 define sistemas de inteligência artificial de propósito geral como "um sistema de IA que pode ser usado e adaptado para uma vasta gama de aplicações para as quais não foi intencionalmente e especificamente designado"[138]. A proposta europeia do PE (2023) introduziu também a emenda 99, que traz a definição de modelos de IA fundacionais (*Foundation Models*):

> Modelos fundacionais são um desenvolvimento recente, no qual modelos de IA são desenvolvidos a partir de algoritmos criados para otimizar um output em relação a sua generalidade e versatilidade. Esses modelos são com frequência treinados em uma ampla gama de fontes de dados e grandes quantidades de dados para alcançar uma ampla variedade de tarefas ("*down-stream tasks*", tarefas que dependem de um output anterior de outra tarefa ou processo), incluindo tarefas para as quais não foram desenvolvidos ou treinados de maneira específica. O modelo fundacional pode ser unimodal ou multimodal, treinado através de vários métodos como aprendizado supervisionado ou aprendizado reforçado. Sistemas de IA com um propósito específico ou de propósito geral podem ser a implementação de um modelo fundacional, o que quer dizer que cada modelo fundacional pode ser reutilizado em inúmeros outros sistemas de IA ou sistemas de IA de propósito geral. Esses modelos têm importância crescente para diversas aplicações e sistemas (aplicações e sistemas "*downstream*")[139].

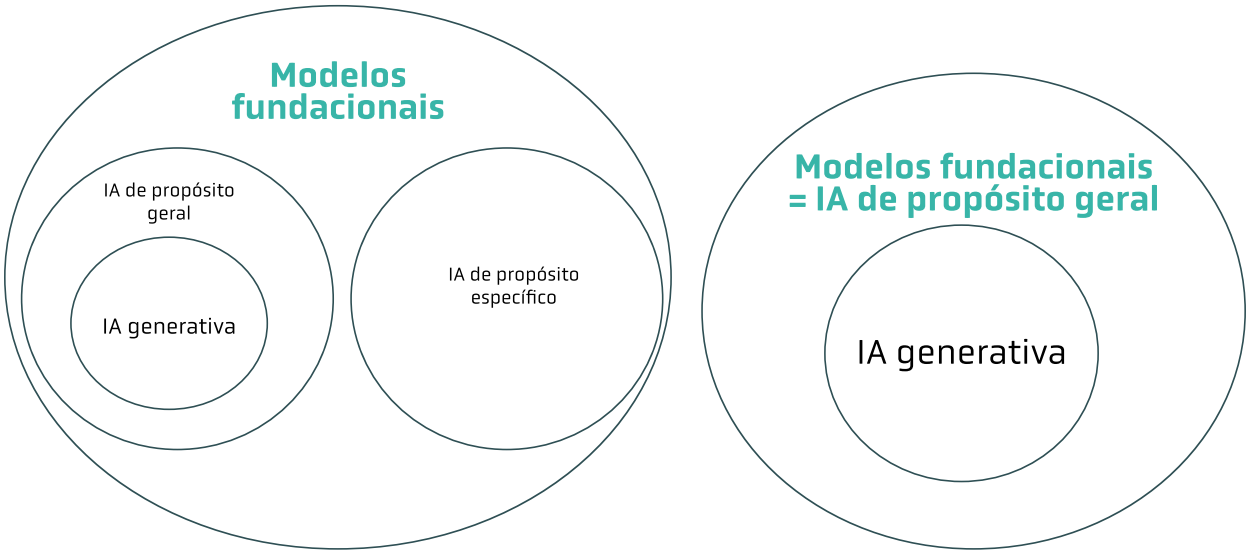
Como se observa nas definições adotadas pelo Parlamento, não fica clara a relação entre os termos "IA de propósito geral" e "IA de modelo fundacional", já que, segundo a definição, tanto sistemas de IA com um propósito específico quanto com propósito geral podem resultar da utilização de um modelo fundacional. Já os modelos generativos de IA, de acordo com a classificação europeia, são um tipo de modelo fundacional: "(...) sistemas de IA com propósito específico de gerar, com diferentes graus de autonomia, conteúdo como texto complexo, imagens, áudio ou vídeo" (emenda 399)[140]. Nesse caso, podemos entender que há uma relação de gênero e espécie, sendo o gênero, modelos fundacionais,

138 Tradução própria, p. 113 – https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf.

139 Tradução livre a partir de https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf. p.74-75.

140 Tradução livre a partir de https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf, p.200.

e a espécie, modelos generativos de IA.



Possibilidades de entendimento dos conceitos trazidos pela versão do PE do EU AI Act

Assim, a partir da legislação europeia, poderíamos adotar a definição de modelo fundacional como englobando tanto a IA de propósito geral quanto a IA generativa. Em sentido semelhante, Hacker *et al*[141], tratando da regulação de modelos de IA como o ChatGPT, equipara os termos "modelos fundacionais", "modelos grandes de linguagem" (LLMs) ou "modelos grandes de IA generativa" (LGAIMs) – este último escolhido como o termo adotado pelo referido artigo.

A partir daqui, tratamos, como Hacker *et al*[142], das designações utilizadas de modelo de IA como equiparáveis (IA generativa, modelos fundacionais, modelos grandes de linguagem – LLMs, modelos grandes de IA generativa – LGAIMs), porque, mesmo que não designem exatamente o mesmo fenômeno, suas características comuns levam a semelhantes considerações em relação à regulação, o que é também seguido pela OCDE e os países do G7[143].

De maior relevância, além do desafio de conceituação para regulação, que é comum à definição da própria IA, um segundo desafio decorre do fato de que a IA Generativa tensiona a regulação baseada em risco, modelo predominante nas tentativas de regulação de IA globalmente, como exposto neste relatório. Isso porque tal modelo regulatório é

141 HACKER *et al.*, 2023, p. 1113.

142 Ibid.

143 OECD. Initial policy considerations for generative artificial intelligence. Publicado em 18 set. 2023. Disponível em: https://www.oecd-ilibrary.org/deliver/fae2d1e6-en.pdf?itemId=%2Fcontent%2Fpaper%2Ffae2d1e6-en&mimeType=pdf; OECD. G7 Hiroshima Process on Generative Artificial Intelligence (AI): Towards a G7 Common Understanding on Generative AI. Relatório preparado para a presidência japonesa de 2023 e para o grupo de trabalho digital e tecnológico do G7. Publicado em 7 set. 2023. Disponível em: https://www.oecd-ilibrary.org/deliver/bf3c0c60-en.pdf?itemId=%2Fcontent%2Fpublication%2Fbf3c0c60-en&mimeType=pdf.

eminentemente contextual, isto é, depende de qual situação específica a IA será aplicada para, então: a) avaliar eventuais riscos a direitos de pessoas impactadas; e b) de acordo com os riscos, se graduar as obrigações decorrentes.

No caso de LLMs, os modelos se prestam a diferentes finalidades, que podem não ser possíveis de antever quando desenvolvidos, desafiando, então, a regulação baseada no risco, que foca em regular os usos da tecnologia e abordar seus impactos em contextos específicos e a partir de uma cadeia de agentes envolvidos menos complexa e dinâmica. Algumas disposições podem ser adicionadas em uma tentativa de mitigar ou solucionar tais desafios postos ao modelo regulatório baseado no risco.

Em primeiro lugar, apesar de não resolver o desafio por si só, a previsão de "inteligência artificial de propósito geral", dentro de uma regulação baseada no risco, estabelecendo regras específicas para tal modelo, é um primeiro passo em direção à sua regulação. Como mencionado, a versão do AI Act adotada pelo Conselho da UE em 6 de dezembro de 2022 já trazia a conceituação de tais modelos, o que foi aprimorado na versão do PE de junho de 2023, que passou a incluir também modelos fundacionais e generativos. No contexto brasileiro, os projetos de lei em tramitação no Congresso Nacional em sua maioria não fazem menção a essas figuras, com exceção do PL 2338/23, que menciona no § 1º do art. 13 "sistemas de inteligência artificial de propósito geral", sem, contudo, definir o termo, como já explicitado anteriormente.

Além da inclusão da ideia de "IA de propósito geral" (e suas variações), é possível pensarmos na regulação destes modelos de IA por meio da regulação baseada no risco ao incluirmos, dentro da ideia de risco, não apenas aqueles conhecidos e previsíveis, mas também os riscos que podem razoavelmente ser esperados, a partir do princípio da precaução. Nesse sentido, mesmo não havendo ampla certeza quanto à existência de certos riscos, essa incerteza e desconhecimento não podem ser usados como escusa para não se empregar medidas para evitar que eles aconteçam. Essa previsão pode ser encontrada tanto no contexto brasileiro, no PL 2338, como no europeu, na última versão do EU AI Act do Parlamento.

No Brasil, o projeto de lei 2338/23 estabelece que, caso a IA de propósito geral seja utilizada para uma das finalidades listadas como de alto risco pelo art. 17, esse sistema deverá cumprir com uma série de obrigações de governança, incluindo a elaboração de uma avaliação de impacto, prevista entre os artigos 22 e 26. Dentro dessa avaliação, o fornecedor da IA de propósito geral deverá considerar e registrar, dentre outros elementos, os "riscos conhecidos e previsíveis associados ao sistema de inteligência artificial à época em que foi desenvolvido, bem como os riscos que podem razoavelmente dele se esperar" (§1º do art. 24). Nesse sentido, o projeto brasileiro está alinhado ao princípio da precaução,

inclusive prevendo que, nos casos de sistemas de IA que possam ter como consequência impactos irreversíveis ou de difícil reversão, que a avaliação de impacto considere também as evidências incipientes, incompletas ou especulativas (§ 2º do art. 24).

O Código de Conduta Voluntário em Desenvolvimento e Gerenciamento de Sistemas de IA Generativa Avançados Responsáveis, anunciado em setembro de 2023 pelo ministro de Inovação, Ciência e Indústria do Canadá, também traz a previsão de "riscos razoavelmente esperados" dentre as questões que devem ser analisadas pelos desenvolvedores e gerenciadores em avaliação dos impactos adversos dos sistemas de IA generativos para cumprimento do princípio da segurança[144].

Já na União Europeia, a versão do Conselho incluiu o Título Ia específico para sistemas de IA de propósito geral em que há a definição das regras aplicáveis aos fornecedores dessa tecnologia no art. 4b, estendendo a eles certas obrigações de sistemas de IA de alto risco (apesar de mencionar a necessidade de um ato de execução que especificasse essa aplicação aos sistemas de propósito geral), sem mencionar os "riscos razoavelmente esperados". A inserção desse termo foi feita apenas pela versão do Parlamento Europeu em um novo artigo também especificamente criado para modelos fundacionais. De acordo com o art. 28b (2), o fornecedor destes modelos deve demonstrar, por meio de projeto, teste e análise apropriados, a identificação, redução e mitigação de riscos razoavelmente previsíveis à saúde, segurança, direitos fundamentais, meio ambiente e democracia e o estado de direito antes e durante desenvolvimento. Alinhado a isso, a Seção C do Anexo III, que trata das obrigações de transparência de sistemas fundacionais, determina que os fornecedores desses sistemas devem disponibilizar e manter em registro informações sobre os riscos razoavelmente previsíveis e as medidas que foram tomadas para mitigá-los, bem como riscos remanescentes não mitigados com uma explicação sobre o motivo pelo qual eles não podem ser mitigados.

Assim, a partir da introdução dessa ideia de "riscos razoavelmente esperados" o que o PL brasileiro fez, e que posteriormente foi também incluído na regulação europeia[145] e no código canadense, foi tentar fazer com que, mesmo não sendo possível prever todos os casos de riscos associados ao sistema de IA fundacional, que seus fornecedores lidem com os riscos que razoavelmente seriam dele esperados, mesmo que esses riscos

144 Governo do Canadá. Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems. Setembro de 2023. Disponível em: https://ised-isde.canada.ca/site/ised/en/voluntary-code-conduct-responsible-development-and-management-advanced-generative-ai-systems.

145 Em 14 de junho de 2023, o Parlamento Europeu adotou algumas emendas ao texto do EU AI Act. Dentre elas, destaca-se a emenda 102 que introduziu o considerando 60h específico para modelos fundacionais; emenda 263 que anexou o termo "razoavelmente" aos riscos previsíveis na etapa de identificação e análise dos riscos do sistema de gestão de riscos; emenda 399 que introduziu o artigo 28b para criar obrigações para o fornecedor de modelos fundacionais; e emenda 771 que criou a Seção C do Anexo VIII para incluir riscos razoavelmente previsíveis dentro da descrição das capacidades e limitações de modelos fundacionais que precisam ser fornecidas e registradas. Para mais informações, acessar: https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf.

não venham a se concretizar. Tal disposição se alinha ao princípio da precaução.

Apesar de a proposta europeia não mencionar o princípio da precaução, ele pode ser extraído do art. 28b (2) e do novo Considerando 60-G. Este último estabelece que, em razão da complexidade e impacto inesperado dos sistemas de IA fundacionais, além da falta de controle dos provedores posteriores de IA sobre o desenvolvimento dos LGAIMs, deve haver um compartilhamento justo de responsabilidades ao longo da cadeia de valor de IA, o que faz com que esses modelos sejam sujeitos a medidas proporcionais e mais requisitos e obrigações específicos, a exemplo da obrigação de avaliar e mitigar possíveis riscos e danos e a de implementar medidas de gestão de dados, incluindo a avaliação de enviesamentos.

Dito isso, um ponto de diferenciação entre o PL 2338/23 e a versão do PE do AI Act é em relação ao elemento que deflagra a obrigação de avaliar riscos e mitigá-los para as IAs de propósito geral. Enquanto o modelo brasileiro prevê a obrigatoriedade de elaboração de uma avaliação de impacto algorítmico para as finalidades de sistemas de IA de alto risco (art. 17), onde os LGAIMs podem se enquadrar ou não, a última versão do texto europeu prevê, desde logo, obrigações específicas para esses modelos no novo art. 28b, incluindo a mencionada avaliação e mitigação de riscos razoavelmente esperados, independentemente do nível de risco.

Para Hacker[146], a versão do PE do AI Act traz avanços significativos para a regulação dos LGAIMs. Porém, pela interpretação do autor do art. 28b(2)(a), todos esses modelos teriam que implementar avaliações de risco e medidas de mitigação para os riscos razoavelmente previsíveis à saúde, segurança, direitos fundamentais, meio ambiente, democracia e estado de direito, com o envolvimento de especialistas independentes, o que faria com que, na prática, sejam equiparáveis às IAs de alto risco. Esta classificação de alto risco "presumida", para o autor, inviabilizaria tais modelos na prática.

Contudo, mesmo que certos aspectos da avaliação de risco sejam alterados na regulação ideal, considerando sua onerosidade no caso desses modelos, é importante destacar que eles já apresentam hoje riscos específicos e relevantes, não só para os direitos humanos como para a economia e sociedade. Esses riscos podem ser, de forma não exaustiva, divididos em: (i) riscos relacionados aos consumidores; (ii) geração de desinformação; (iii) riscos de restrição da competição econômica no mercado; (iv) discriminação, (v) sustentabilidade ambiental; e (vi) propriedade artística e intelectual, especialmente no que tange aos direitos autorais[147]. Isso, por si só, justificaria que os LGAIMs passassem por avalia-

146 HACKER *et al.*, 2023, p. 1115.

147 Estudo da OCDE de setembro de 2023 analisou alguns desses riscos, que já podem ser sentidos na realidade prática, a exemplo da amplificação de desinformação, reforço de práticas discriminatórias e enviesadas, questões de direitos de propriedade intelectual e impacto no mercado de trabalho; OECD. Initial policy considerations for generative artificial intelligence. Publicado em 18 set. 2023. Disponível em: https://www.oecd-ilibrary.org/deliver/fae2d1e-

ções de risco e consequente mitigação.

Ademais, o Considerando 60-G da versão do PE do AI Act esclarece que os requisitos e obrigações específicos para os sistemas de IA fundacionais não equivalem a considerar esses modelos como sistemas de IA de alto risco, mas que possuem a função de garantir que haja um elevado nível de proteção dos direitos fundamentais, saúde e segurança, meio-ambiente, democracia e Estado de Direito.

Além dos elementos mencionados, a regulação de sistemas de IA de propósito geral também pode ser melhor alcançada a partir de modelos regulatórios baseados no risco por meio de uma melhor definição dos agentes envolvidos na cadeia produtiva desses sistemas, destrinchando as obrigações de cada um. Em análise comparativa entre o PL 2338/23 e a última versão do AI Act, vinda do Parlamento Europeu, nota-se clara complexificação da rede de agentes envolvidos no caso europeu. Enquanto o projeto brasileiro menciona apenas os agentes de IA (fornecedor e operador), a proposta europeia prevê fornecedores, distribuidores, importadores, operadores e outros terceiros e, especificamente para os modelos fundacionais, contempla o fornecedor, novos fornecedores e os demais agentes envolvidos na cadeia de valor dos sistemas.

Elemento de destaque da versão vinda do PE foi a proposta de uma cooperação entre os agentes envolvidos com os LGAIMs, que já existia na versão do Conselho para os fornecedores (Article 4b(5)), mas foi melhorada e complexificada nesta última atualização. De acordo com o Considerando 60-F, por exemplo, para os modelos fundacionais fornecidos como um serviço (como aqueles por meio de API), estipula-se, como regra, que os fornecedor original deve cooperar com os fornecedores posteriores (downstream providers) ao longo do tempo durante o qual esse serviço é fornecido e suportado, de forma a permitir a mitigação de risco apropriada. Somado a isso, destaca-se também que a proposta europeia do PE, no Considerando 60-G, aborda o elemento de incerteza na evolução dos modelos fundacionais de IA e como isso impacta a definição das responsabilidades dos agentes.

Assim, diante da complexidade dos modelos e desse cenário de incerteza, o dispositivo destaca a necessidade de esclarecer o papel dos atores que contribuem para o desenvolvimento desses sistemas, especialmente os fornecedores (originais e posteriores). Como a tecnologia envolve alta complexidade e pode ter impactos inesperados, especialmente porque permite a sua utilização variada, inclusive para funcionalidades não inicialmente pensadas pelo fornecedor original, o texto europeu prevê a falta de controle dos fornecedores posteriores e estabelece obrigações de governança mais intensas para os fornecedores originais.

6-en.pdf?itemId=%2Fcontent%2Fpaper%2Ffae2d1e6-en&mimeType=pdf.

Em linha com tal disposição, o Considerando 60-H constata que, dada a natureza dos modelos fundacionais, falta experiência em avaliação de conformidade e os métodos de auditoria de terceiros ainda estão em desenvolvimento. Consequentemente, definem a obrigação de a Comissão Europeia e a autoridade europeia específica de IA, que será criada, terão a responsabilidade de monitorar e avaliar periodicamente a estrutura legislativa e de governança de tais modelos no contexto da UE.

Apesar de não trazer um maior destrinchamento da cadeia de agentes envolvidos ou de incluir "riscos razoavelmente previsíveis", em razão de sua natureza e escopo[148], a Ordem Executiva de "Desenvolvimento e Uso Seguro, Protegido e Confiável de Inteligência Artificial" do presidente norte-americano Joe Biden, publicada em 30 de outubro de 2023, aborda pontos interessantes sobre a governança dos modelos fundacionais por parte do governo federal. Ressalta-se a grande preocupação da administração Biden com riscos de segurança relacionados à produção de conteúdo sintético, como faz a IA Generativa, que pode ser negativamente utilizada para produção de *deep-fake* e *deep nudes*[149], por exemplo.

Dentre as medidas anunciadas, está: (i) a promoção de capacidades de identificação e rotulação de conteúdo sintético produzido por IA, inclusive com a possibilidade de auditar tais sistemas, o uso de marca d'água e a proibição de que a IA generativa produza material de abuso sexual infantil ou de imagens íntimas não consensuais de pessoas reais; (ii) obtenção de informações sobre IAs fundacionais de uso duplo cujo modelo seja amplamente aberto[150] (e.g. quando o modelo está publicado na Internet), já que podem representar riscos significativos para a segurança, incluindo a solicitação de contribuições do setor privado, academia e sociedade civil sobre riscos, benefícios e abordagens políticas e regulatórias aplicáveis a esses modelos (inclusive informações sobre mecanismos

148 A Ordem Executiva é significativa, mas não tem a profundidade e o detalhamento que uma legislação poderia trazer. Embora tenha o poder de iniciar ações e definir prioridades para as agências federais, não possui força vinculante para o setor empresarial, apesar de sua inegável influência. Isso porque, ao estabelecer padrões e requisitos para a IA que adquire, o governo pode moldar e direcionar práticas de mercado, uma vez que as empresas terão que se adaptar a essas regras caso queiram firmar contratos com o governo. De toda forma, a falta de coercibilidade e a presença de mecanismos de governança eficientes para empresas pode levar a uma aplicação não uniforme e à carência de conformidade; Center for AI and Digital Policy. "World Cup" of AI Policy News edition. CAIDP Update 5.42 – AI Policy News (Nov. 6, 2023). Disponível em: https://www.linkedin.com/posts/center-for-ai-and-digital-policy_caidp-update-542-ai-policy-news-nov-activity-7127339609293824000-fCHi/.

149 Deep nude é a prática de utilização de sistemas de IA para geração de conteúdo de nudez falsa, geralmente tendo por base uma foto da vítima vestida; LOPES, Larissa. Já ouviu falar na prática do Deep Nude? Jusbrasil, publicado em outubro 2023. Disponível em: https://www.jusbrasil.com.br/artigos/ja-ouviu-falar-na-pratica-do-deep-nude/1979706886.

150 Tal posicionamento da Administração Biden parece alinhar-se à teoria de Irene Solaiman, em cujo artigo defende um framework de avaliação de IAs generativas de acordo com seu grau de abertura/acesso (totalmente fechadas, gradual, acesso hospedado, acesso em API ou em nuvem, acesso baixável e totalmente aberto); SOLAIMAN, Irene. The Gradient of Generative AI Release: Methods and Considerations. Fevereiro de 2023. Disponível em: https://arxiv.org/abs/2302.04844.

para gerenciamento de riscos e benefícios)[151].

Ademais, a Ordem possui uma seção específica para promoção do uso responsável e seguro de IA generativa no Governo Federal (section 10, 10.1, (f)). Após desencorajar a imposição de proibições ou bloqueios gerais amplos para a utilização de IA generativa por agências federais, o documento ressalta, por exemplos: (i) a necessidade de limitação do acesso, conforme necessário, a serviços específicos de IA generativa com base em avaliações de risco específicas; (ii) estabelecimento de diretrizes e limitações sobre o uso apropriado de IA generativa; e (iii) incentivo ao emprego de práticas de gestão de riscos, como treinamento de funcionários e cumprimento dos requisitos de manutenção de registros, segurança cibernética, confidencialidade, privacidade e proteção de dados.

Por fim, ainda no que se refere à IA Generativa, discute-se também a possibilidade de serem LLMs auditados. Nesse tópico, Luciano Floridi et al (2023) defendem um modelo de auditoria ou de avaliação de risco desses modelos em três camadas: de governança, do modelo e da aplicação. Segundo os autores, primeiramente, os fornecedores de tecnologias de LLMs se submeteriam a auditorias de governança que avaliariam procedimentos organizacionais, estruturas de responsabilização internas e sistemas de gestão da qualidade para verificação, por exemplo, dos níveis de robustez. Em seguida, os LLMs passariam para as auditorias do modelo, avaliando as suas capacidades e limitações após a formação inicial, mas antes da implementação em aplicações concretas específicas, com fins de verificação do desempenho, segurança da informação e veracidade. Por fim, os produtos e serviços criados com base nos LLMs passariam por auditoria de aplicação contínua para avaliação da conformidade legal e de seu impacto nos utilizadores, grupos e no ambiente natural ao longo do tempo. Essas camadas atuariam na informação e complementação uma das outras, de forma a contribuir para a boa governança de sistemas complexos, incluindo até mesmo LLMs.

151 Section 4, 4.5 do Executive Order.

PRINCIPAIS PROPOSTAS DE REGULAÇÃO DE IA NO BRASIL SOBRE IA GENERATIVA				
PL 5051/19	PL 21/20	PL 872/21	PL 759/23	PL 2338/23
Não menciona.	Não menciona.	Não menciona.	Não menciona.	Apenas menciona "sistemas de inteligência artificial de propósito geral"

PROPOSTAS DE LEI QUE JÁ PREVEEM ALGUM DISPOSITIVO SOBRE IA GENERATIVA			
Normativa	Parâmetros		Menção na normativa
PL 2338/2023 (Brasil)	**Definição**	-	-
	Cadeia de agentes envolvidos	Fornecedor e operador (agentes de IA)	Art. 4º, incisos II, III e IV e art. 13, § 1º
	Previsão de riscos razoáveis por parte da cadeia de agentes	Sim	Art. 3º, XI (aplicação geral) e art. 24, § 1º, a e § 2º (no caso de IA de alto risco)
	Obrigações	Não há obrigações específicas para IA generativa, mas obrigações para sistemas de IA de acordo com o grau de risco	Capítulo IV (Governança de Sistemas de IA)
Proposta de Regulamento de Inteligência Artificial da União Europeia (*EU AI Act Proposal*) - versão do Conselho da Europa	**Definição**	IA de propósito geral (*general purpose AI system*)	-
	Cadeia de agentes envolvidos	Fornecedor (*provider*)	Art. 4º, incisos II, III e IV e art. 13, § 1º
	Previsão de riscos razoáveis por parte da cadeia de agentes	Não	Art. 3º, XI (aplicação geral) e art. 24, § 1º, a e § 2º (no caso de IA de alto risco)
	Obrigações	Aplica-se as regras de sistemas de IA de alto risco, mas depende de um ato de execução que especificaria como essas regras seriam aplicadas para sistemas de IA de propósito geral, à luz das suas características, viabilidade técnica, especificidades da cadeia de valor da IA e da evolução do mercado e tecnológica. Há exceções para essa regra	Capítulo IV (Governança de Sistemas de IA)

Proposta de Regulamento de Inteligência Artificial da União Europeia (*EU AI Act Proposal*) - versão do Parlamento Europeu	**Definição**	Modelo fundacional	-
		Inteligência artificial de propósito geral	Art. 3º, (1) (c)
		IA generativa	Art. 3º, (1) (d)
	Cadeia de agentes envolvidos	Operador (fornecedor, implantador, representante autorizado, importador e distribuidor). No caso de IA generativa, especificamente fala-se em fornecedor, novo fornecedor e demais atores da cadeia de valor de IA	Art. 28b (4)
	Previsão de riscos razoáveis por parte da cadeia de agentes	Sim	Considerando 60h, artigo 9º (2) (a), artigo 28b (2) (a) e Anexo VIII, Seção c (6)
	Obrigações	Diferentes regras, independentemente de ser fornecido como um modelo independente ou incorporado em um sistema ou produto de IA, ou fornecido sob licenças gratuitas e de código aberto, como um serviço, etc. Exemplos: (i) demonstrar a identificação e mitigação de riscos razoavelmente previsíveis, inclusive com a inclusão de experts nessas avaliações; (ii) incorporação de datasets apenas quando submetidos por medidas de governança de dados; (iii) projetar e desenvolver o modelo de fundação, fazendo uso de padrões aplicáveis para reduzir o uso de energia; (iv) manter documentação técnica e instruções inteligíveis para o uso; (v) estabelecimento de um sistema de gerenciamento de qualidade; (vi) registro no banco de dados da UE do art. 60; (vii) obrigações de transparência, dentre outras.	Artigo 28b
***Interim Measures for the Management of Generative Artificial Intelligence Services* (China)**	**Definição**	**Tecnologia de inteligência artificial generativa**	Art. 22 (1)
	Cadeia de agentes envolvidos	Provedores e usuários de serviços de IA generativa	Art. 22 (2) e (3), respectivamente

Interim Measures for the Management of Generative Artificial Intelligence Services **(China)**	Previsão de riscos razoáveis por parte da cadeia de agentes	Não	-
	Obrigações	Há várias obrigações para fornecedores, como proteção de informações, rotular conteúdo gerado pela IA generativa, estabelecer mecanismos de reclamações e denúncias, além de obrigações tanto para fornecedores como usuários, a exemplo da adesão aos valores fundamentais socialistas, medidas de prevenção à discriminação, respeito aos direitos de propriedade inteliectual, medidas de transparência, etc.	Art. 4º e capítulos 2 e 3.
Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems **(Canadá)**	Definição	-	-
	Cadeia de agentes envolvidos	Desenvolvedores (*developers* - fazendo a diferenciação também para os *downstream developers*) e gerentes (*managers*)	Tabela no site oficial do Código de Conduta
	Previsão de riscos razoáveis por parte da cadeia de agentes	Sim	Medidas a serem tomadas de acordo com o Código de Conduta - Princípio de Segurança
	Obrigações	Define uma lista de medidas que devem ser tomadas de acordo com princípios, separando-as entre as que devem ser seguidas por *deployers* e *managers*, além de variar se é um caso de sistemas generativos avançados para o uso público ou não. Exemplos: implementação de um *framework* compreensivo de gerenciamento de risco, divulgação de informações, cooperação entre agentes de IA generativa, métodos de teste, etc.	Tabela no site oficial do Código de Conduta

***Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence - Joe Biden* (EUA)**	Definição	Modelo fundacional de dupla utilização (*dual-use foundation model*) e IA Generativa (*Generative AI*)	Section 3, (k) e (p)
	Cadeia de agentes envolvidos	-	-
	Previsão de riscos razoáveis por parte da cadeia de agentes	Sim (não há o termo de forma expressa, mas pode ser do conceito de "*AI red-teaming*"[152])	Section 3, (d)
	Obrigações	Obrigações para agências do governo federal que incluem, por exemplo, realização de avaliação de risco para IA Generativa (que pode gerar limite de acesso para uma IA Generativa), práticas de gerenciamento de risco, treinamento de pessoal, criação de um guia para uso de IA generativa no trabalho, realização de consultas públicas sobre modelos fundacionais básicos de uso duplo amplamente disponíveis, entre outros.	Section 4, (i) (A) (B); Section 4, 4.4, (ii) (A) (B); Section 4, 4.5, (a), (iv); Section 4, 4.6; Section 8, (b), (i) (A); Section 10, 10.1, (b) (viii) (A), (B), (C); Section 10, 10.1, (f)

152 (d) O termo "*red-teaming de IA*" ("equipe vermelha de Inteligência Artificial") significa um esforço de teste estruturado para encontrar falhas e vulnerabilidades em um sistema de IA, muitas vezes em um ambiente controlado e em colaboração com criadores de IA. A equipe vermelha de Inteligência Artificial é mais frequentemente realizada por "equipes vermelhas" ("red teams") dedicadas que adotam métodos adversários para identificar falhas e vulnerabilidades, como resultados prejudiciais ou discriminatórios de um sistema de IA, comportamentos imprevistos ou indesejáveis do sistema, limitações ou riscos potenciais associados ao mau uso do sistema. (Tradução própria)

5. Particularidades nacionais para a regulação de IA à brasileira

Em qualquer construção de ambiente regulatório, é imprescindível que a regulação considere as particularidades do contexto em que ela será implementada. O Brasil, como país do sul global[153], é atravessado por questões sociais, raciais, de gênero, de colonialidade e de território, o que é reforçado pela ampla utilização de sistemas de IA que carregam em si majoritariamente vozes e padrões masculinos, ocidentais, europeus, de branquitude e de riqueza[154].

Por exemplo, o país ainda apresenta intensas violações de direitos humanos, liderando índices de LGBTQfobia, desigualdade, racismo e violência de gênero[155]. Ainda hoje, podemos falar de um colonialismo digital, em que grande parte do mundo majoritário/sul global, incluindo o Brasil, encontra-se em uma posição de colônia, utilizada para obtenção de mão de obra barata e mineração extrativista de dados e de matéria-prima bruta, enquanto também é posta em condição de mercado consumidor de tecnologias emergentes vindas do norte global, especialmente de grandes empresas monopolistas de tecnologia[156].

Considerando esse cenário atual de violências, desigualdades e opressões, o país não pode adotar uma postura de importação irrefletida de modelos regulatórios de contextos de norte-global, uma vez que possuímos circunstâncias (políticas, econômicas e sociais), identidades, características e problemas que nos são próprios – e que devem ser considerados na construção regulatória, o que demanda que a regulação brasileira de inteligência artificial seja pensada por meio de nossas peculiaridades. Dessa forma, apesar de ser bem-vinda a incorporação de dispositivos estrangeiros que façam sentido para

153 O termo Sul Global foi utilizado pela primeira vez em 1969 pelo ativista político Carl Oglesby. Ele foi cunhado para substituir de forma mais neutra expressões como "países subdesenvolvidos" ou "terceiro mundo", que tinham conotações negativas, já que reforçavam os estereótipos sobre as comunidades pobres e as representam como ícones da pobreza, escondendo suas histórias de opressão e exploração contínua. Porém, nos últimos tempos, a expressão "sul-global" também passou a ser vista como pejorativa, já que acabam por ser imprecisa, homogeneizar grupos, além de criar certo determinismo geográfico, como se os países do Hemisfério Sul fossem fadados a ser pobres e não terem expectativas de desenvolvimento. Assim, talvez a melhor expressão seja "Mundo Majoritário", já que esses países representam, de fato, a maioria da humanidade; HEINE; Jorge. O Sul Global está em ascensão – mas o que é exatamente o Sul Global? Interesse Nacional, publicado em 10 de jul. 2023. Disponível em: https://interessenacional.com.br/edicoes-posts/o-sul-global-esta-em-ascensao-mas-o-que-e-exatamente-o-sul-global/. Acesso em 25 jul. 2023; Demetriodor. O 'Sul Global' é um termo terrível. Não use! Publicado em 11 nov. 2018. Disponível em: re-design.dimiter.eu/?p=969; ARUN, Chinmayi. AI and the Global South: Designing for Other Worlds. In: DUBBER, M.; PASQUALE, F; DAS, S. Oxford Handbook of Ethics of AI. 2019.

154 AI Manyfesto, 2021.

155 SILVA, Tarcizio. Regular a inteligência artificial no Brasil pode mitigar o racismo algorítmico. Folha de São Paulo, publicado em 03 de jul. 2023. Disponível em: https://www1.folha.uol.com.br/blogs/politicas-e-justica/2023/05/regular-a-inteligencia-artificial-no-brasil-pode-mitigar-o-racismo-algoritmico.shtml#:~:text=Novo%20projeto%20de%20lei%20avan%C3%A7ou,combate%20aos%20danos%20do%20racismo&text=Os%20impressionantes%20saltos%20t%C3%A9cnicos%20nos,maravilha%20sobre%20as%20tecnologias%20digitais. Acesso em 21 jul. 2023.

156 FAUSTINO; LIPPOLD, 2023.

nossa realidade, é ainda mais vital a criação de outros específicos para lidarmos com as nossas particularidades.

Nesse contexto, o PL 21/20 e o 759/23 não avançam nessa temática, já que apresentam uma redação genérica, majoritariamente principiológica e com baixa carga coercitiva, uma vez que não preveem ferramentas de governança eficientes para lidar com os problemas e riscos da IA, principalmente no cenário em que esses riscos reforçam discriminações estruturais e vividas em camadas de opressão interseccionais, como no caso brasileiro. A título de exemplificação, o termo "não-discriminação" é apenas citado como fundamnto e princípio para o uso responsável da IA no Brasil no PL 21/20 e não é sequer mencionado no texto do PL 759, 872 e 5051.

Avanços mais significativos podem ser encontrados no texto do PL 2338/2023. Em primeiro lugar, o PL reconhece as desigualdades e assimetrias estruturais do contexto brasileiro ao adotar expressamente as definições de discriminação direta e indireta (art. 4º, VI e VII) vindas da Convenção Interamericana contra o Racismo, que, desde 2022, tem status de emenda constitucional em território nacional, trazendo um reforço da proteção contra discriminações em diferentes pontos do texto[157].

Um desses momentos, ocorre com a previsão de um rol de direitos para indivíduos potencialmente afetados pela IA no art 5º, com destaque para os direitos de correção de vieses discriminatórios (diretos, indiretos, ilegais ou abusivos – art. 12º), à informação (art. 7º), à explicação (art. 8º) e à contestação (art. 9º). Nesse ponto, o PL reforça a não vedação à adoção de critérios de diferenciação quando isso ocorre em razão de objetivos ou justificativas razoáveis e legítimas à luz dos direitos fundamentais (PU do art. 12), como é o caso de ações afirmativas, como é o caso de cotas raciais, por exemplo.

Além do foco no combate à discriminação, o texto tem como outro ponto de atenção: a proteção de grupos (hiper)vulneráveis em diferentes momentos. Para citar exemplos, dentre os critérios para atualização da lista dos sistemas de alto risco e risco excessivo pela futura autoridade de IA está o fato de "o sistema ser discriminatório" (art.18 c) e de "o sistema afetar pessoas de um grupo vulnerável específico" (Art.18 d), além de a metodologia para a Avaliação de Impacto Algorítmico destacar o possível impacto discriminatório dos sistemas (art.24, §1º, f).

Ademais, como foi explicitado em tópico anterior, o PL 2338/23 reforça a importância da participação da sociedade na avaliação e no conhecimento dos riscos dos sistemas de IA, a partir da previsão dessa participação nos processos de Avaliação de Impacto Algorítmico (art. 25, §2º) e na obrigação de publicação de suas conclusões, o que permite

157 Comissão de Juristas Responsável por Subsidiar a Elaboração de Substitutivo sobre Inteligência Artificial no Brasil (CJSUBIA). Relatório Final. 2022. Disponível em: https://www.stj.jus.br/sites/portalp/SiteAssets/documentos/noticias/Relato%CC%81rio%20final%20CJSUBIA.pdf p. 12–13.

controle público e social dos riscos.

Desta forma, para a regulação da tecnologia no país, quando comparado com o PL 5051/19, PL 21/20, PL 871/21 e com 753/23, o PL 2338/23 pode ser interpretado não como uma rivalização, mas como um avanço, já que sua construção foi pensada de forma ampla e em conjunto com a sociedade por meio de consultas e audiências públicas, multidisciplinares e multissetoriais, nacionais e internacionais, realizadas ao longo do processo da CJSUBIA em 2022, inclusive com maior inclusão de grupos possivelmente afetados pelos sistemas de IA, especialmente minorias e comunidades vulnerabilizadas[158], o que possibilitou a inclusão e/ou manutenção de aspectos regulatórios relevantes para a realidade brasileira, a exemplo da definição de discriminação direta e indireta, a previsão de proteção especial para grupos vulneráveis e a imposição de medidas de governança mais intensas nos casos de sistemas de IA de alto risco.

O PL 2338/2023 é mais afirmativo e protetivo de direitos, o que pode ser extraído do amplo rol de fundamentos e princípios previstos no art. 2º e 3º, respectivamente. O amplo rol de princípios do art. 3º também demonstra a preocupação do projeto de lei em estabelecer uma estrutura normativa protetiva de direitos forte, considerando não só princípios de IA aceitos internacionalmente, como confiabilidade e robustez; transparência, explicabilidade, inteligibilidade e auditabilidade; prestação de contas, responsabilização e reparação integral de dano; não maleficência; e participação humana, previstos em documentos da OCDE, União Europeia, *Berkman Klein Center for Internet & Society*, IEEE, G20 e entidades públicas e privadas[159], mas também outros mais aplicáveis à realidade brasileira, tais como não discriminação, justiça, equidade e inclusão; crescimento inclusivo e desenvolvimento sustentável; devido processo legal, contestabilidade e contraditório em um sentido amplo; e prevenção, precaução e mitigação de riscos sistêmicos.

Essa linguagem mais afirmativa de direitos se alinha com os valores constitucionais que irradiam da Constituição Federal de 1988 (CF/1988), assim como de outras regulações existentes que também prescrevem uma série de direitos e garantias, a exemplo da Lei Geral de Proteção de Dados (LGPD), Código de Defesa do Consumidor (CDC), Marco Civil da Internet (MCI) e dos Estatuto da Igualdade Racial, do Idoso e da Pessoa da Deficiência. Desta forma, a conciliação de uma regulação de risco e de direitos tem ressonância através dessa abertura semântica que conecta a proposta regulatória de IA para com outros diplomas de onde emergirá necessariamente uma diálogo entre todas essas fontes

158 Coalizão Direitos na Rede (CDR). Carta de Apoio ao PL 2338/2023. Publicado em 14 jun. 2023. Disponível em: https://direitosnarede.org.br/2023/06/14/carta-de-apoio-ao-pl-2338-2023/. Acesso em 18 jul. 2023.

159 Fjeld, Jessica and Achten, Nele and Hilligoss, Hannah and Nagy, Adam and Srikumar, Madhulika, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI (January 15, 2020). Berkman Klein Center Research Publication No. 2020-1, Available at SSRN: https://ssrn.com/abstract=3518482.

normativas[160]. Isso sem deixar de lado a preocupação, também constitucional, de incentivo à ciência, tecnologia e inovação[161].

A título exemplificativo, o PL 2338/23 lista como fundamento para o desenvolvimento, implementação e uso de IA no Brasil (art. 2º) a necessidade de centralidade da pessoa humana, o respeito aos direitos humanos e valores democráticos e o livre desenvolvimento da personalidade, expressões dos artigos 1º ao 7º da Constituição Federal. Tais artigos constitucionais também se manifestam pelos fundamentos de defesa da igualdade, não discriminação, pluralidade e o respeito aos direitos trabalhistas; proteção da privacidade, proteção de dados e autodeterminação informativa; e a preservação do meio ambiente e estímiulo ao desenvolvimento sustentável. Estes fundamentos que preveem um vocabulário mais afirmativo de direitos alinham-se ao estímulo ao desenvolvimento tecnológico e inovação, inclusive por meio da promoção de pesquisa; e a defesa da livre iniciativa, livre concorrência e defesa do consumidor, que também irradiam de valores constitucionais, especialmente dos artigos 170, 205 e 218 da CF/88.

A preocupação do projeto de lei em equilibrar o desenho de uma estrutura normativa protetiva de direitos à sua atuação ativa na promoção do desenvolvimento econômico e inovação do país é também visível na própria estrutura topográfica do projeto de lei. Isso porque o texto prevê, primeiro, em capítulo próprio, a sistematização dos direitos dos sujeitos potencialmente impactados pela IA e, ao final, uma seção própria com medidas para fomentar a inovação no país.

Outro ponto de atenção no PL 2338/23 em termos de particularidades brasileiras é a forma como traz obrigações de governança reforçadas para o poder público. Se, de forma geral, a relação entre Estado e cidadão é naturalmente desigual em razão do desequilíbrio de forças, inclusive no que tange ao acesso à informação, esse desbalanço de forças é ainda mais intenso em países do mundo majoritário, como o Brasil. Isso porque são países em que a população ainda demanda maior assistência e um Estado de Bem-Estar Social é vital na tentativa de combater desigualdades e alcançar a igualdade material.

Por isso, por mais que não haja um capítulo específico para o poder público no PL 2338, há dispositivos específicos para ele. No inciso III do art. 14, por exemplo, há uma

160 De acordo com a Teoria do Diálogo das Fontes, diante do pluralismo de fontes legislativas (internacionais, supranacionais e nacionais), sejam elas gerais ou especiais, com campos de aplicação convergente, é necessário que haja diálogo e coordenação entre elas, de forma a não haver sua revogação, derrogação ou ab-rogação, mas sua coordenação em prol de valores superiores, como direitos humanos e a proteção dos vulneráveis; MARQUES, Claudia Lima; BENHAMIN, Antônio Herman. A Teoria do Diálogo das Fontes e seu Impacto no Brasil: uma homenagem a Erik Jayme. Revista de Direito do Consumidor (RDC) 2340, 115 indb, 2018. Disponível em: https://revistadedireitodoconsumidor.emnuvens.com.br/rdc/article/view/1042/911.

161 Fala da professora Cláusula Lima Marques e do professor Danilo Doneda na Sessão Parlamentar de instalação da Comissão de Juristas (CJSUBIA) no Senado Federal; TV SENADO. Inteligência artificial: instalação da comissão de juristas que vai analisar o tema – 30/03/22. Realizada em 30 mar. 2022. Disponível em: https://www.youtube.com/watch?v=nXnliBi3vKY&ab_channel=TVSenado. Acesso em 21 jul. 2023.

hipótese de sistema de IA de risco excessivo direcionada particularmente ao Poder Público, que fica proibido de *"avaliar, classificar ou ranquear as pessoas naturais, com base no seu comportamento social ou em atributos da sua personalidade, por meio de pontuação universal, para o acesso a bens e serviços e políticas públicas, de forma ilegítima ou desproporcional"*. Atenção especial também é dada ao Estado para a listagem das finalidades de uso de sistemas de IA consideradas de alto risco pelo art. 17, que inclui, a título exemplificativo a avaliação de critérios de acesso, elegibilidade, concessão, revisão, redução ou revogação de serviços privados e públicos que sejam considerados essenciais", "administração da justiça", "investigação criminal e segurança pública", 'investigação por autoridades administrativas" e "gestão da migração e controle de fronteiras".

Somado a isso, o art. 21 cria medidas de governança adicionais para órgãos e entidades do poder público para contratar, desenvolver ou utilizar sistemas de IA de alto risco. Dentre elas, destaca-se a exigência de que o poder público realize consulta e audiência pública prévias sobre a possível utilização de IA de alto risco (inciso I), além de ter de garantir, de forma facilitada e efetiva, o direito de explicação e revisão humanas ao cidadão no caso de decisões que gerem efeitos jurídicos relevantes ou que impactem significativamente os interesses do afetado (inciso IV)[162]. Outro exemplo de reforço à transparência pública está na obrigação de que haja a publicização em veículos de fácil acesso de todas as avaliações preliminares das IAs desenvolvidas, implementadas ou utilizadas pelo poder público, independentemente do grau de risco (inciso VI). São medidas que tentam reduzir o desequilíbrio de forças existente entre Estado e cidadão, especialmente aquele relacionado à detenção de conhecimento e informação, de forma a garantir que a IA diminua e não amplifique distorções socioeconômicas estruturais.

Assim, enfatiza-se que o atual texto do PL 2338/23 representa um primeiro passo importante para que o Brasil regule a IA a partir da centralidade humana, sendo este sujeito, pessoa humana, aquele que vive e experiencia o Brasil de assimetrias e desigualdades estruturais (dentre elas, o racismo). Contudo, apesar de seus irrefutáveis avanços em termos de proteção de direitos, especialmente de grupos vulnerabilizados, e do combate a todas as formas de discriminação, ainda há espaço para melhorias.

Considerando o aprofundamento das desigualdades e concentração de poder econômico, político e epistêmico nos últimos anos como fruto da ampliação da utilização de sistemas de IA[163], o PL 2338/23 pode avançar em seu compromisso antirracista e anti-

162 GARROTE, Marina. Regulating Artificial Intelligence in Brazil. Center for Human Rights & Global Justice, NYU School of Law, publicado em 28 set. 2023. Disponível em: https://chrgj.org/2023/09/28/regulating-artificial-intelligence--in-brazil/.

163 SILVA, Tarcizio. Regular a inteligência artificial no Brasil pode mitigar o racismo algorítmico. Folha de São Paulo, publicado em 03 de jul. 2023. Disponível em: https://www1.folha.uol.com.br/blogs/politicas-e-justica/2023/05/regular-a-inteligencia-artificial-no-brasil-pode-mitigar-o-racismo-algoritmico.shtml#:~:text=Novo%20projeto%20de%20lei%20avan%C3%A7ou,combate%20aos%20danos%20do%20racismo&text=Os%20impressio-

discriminatório. Por exemplo, há quem sustente a inclusão de critério de avaliação de sistemas de risco excessivo ou alto o potencial de reforço às disparidades interseccionais presentes no país, além do expresso banimento de sistemas de IA[164] considerados racistas, sexistas e transfóbicos[165], especialmente em contextos sensíveis, como é o caso de sistemas de reconhecimento facial para a segurança pública ou ferramentas que avaliam a periculosidade de um indivíduo para fins judiciais.

Ademais, por mais avançado que o PL 2338/23 seja em termos de proteção de direitos e na busca pelo combate à discriminação, o texto ainda apresenta uma postura "defensiva", isto é, trazendo instrumentos de governança – necessários – para promoção da defesa contra os resultados ilegítimos ou ilegais possivelmente produzidos por sistemas de IA. Todavia, o projeto ainda não avança de forma significativa em propostas "reativas", por exemplo, por meio do estímulo à produção de bancos de dados e sistemas de IA éticos diversos, abertos e multidisciplinares em território nacional, em combinação com o fomento à educação e capacitação[166].

Nesse âmbito, o PL 2338/23 apenas menciona no inciso X do art. 2º "o acesso à informação e à educação, bem como a conscientização sobre os sistemas de inteligência artificial e suas aplicações" como um de seus fundamentos legais – enquanto o PL 21/20 e o 759 sequer mencionam. O Brasil possui bons exemplos de legislações que traçam obrigações para o poder público em direção à capacitação, conscientização e educação de forma concreta, de acordo com valores constitucionais, como fez o Marco Civil da Inter-

---

nantes%20saltos%20t%C3%A9cnicos%20nos,maravilha%20sobre%20as%20tecnologias%20digitais. Acesso em 21 jul. 2023.

164 Ibid; Coalizão Direitos na Rede. Nota Técnica do Projeto de Lei nº 2338/2023. Agosto de 2023. Disponível em: https://direitosnarede.org.br/2023/08/23/coalizao-direitos-na-rede-divulga-nota-tecnica-sobre-o-pl-2338-2023-que-busca-regular-a-ia/.

165 BUOLAMWINI, Joy; GEBRU, Timnit. Gender Shades: intersectional accuracy disparities in commercial gender classification. Cambridge: Proceedings of Machine Learning Research, vol. 81, pp.1–15, 2018; BUOLAMWINI, Joy; RAJI, Inioluwa Deborah. Actionable Auditing: investigating the impact of publicly naming biased performance results of commercial AI products. Cambridge: Association for the Advancement of Artificial Intelligence/ACM conference on Artificial Intelligence, Ethics, and Society, 2019. Disponível em: https://www.media.mit.edu/publications/ actionable-auditing-investigating-the-impact-of-publicly-naming-biased-performance-results-of-commercial-ai-products/; COSTANZA-CHOCK, Sasha. Design Justice, A.I., and escape from the matrix of domination. Camrbidge: Journal of Design and Science, jul. 2018. DOI:10.21428/96c8d426. Disponível em: https://jods.mitpress.mit. edu/pub/costanza-chock/release/4; SILVA, Mariah Rafaela; VARON, Joana. Reconhecimento Facial no Setor Público e Identidades Trans: tecnopolíticas de controle e a ameaça à diversidade de gênero em suas interseccionalidades de raça, classe e território. Uma pesquisa realizada pela Coding Rights com apoio da ONG Privacy International via financiamento do International Development Research Center (IDRC). Rio de Janeiro: jan. 2021; COSTA, Ramon; KREMER, Bianca. Inteligência Artificial e Discriminação: Desafios e Perspectivas para a Proteção de Grupos Vulneráveis diante das Tecnologias de Reconhecimento Facial. Direitos Fundamentais & Justiça | Belo Horizonte, ano 16, número especial, p. 145-167, outubro 2022. Disponível em: https://dfj.emnuvens.com.br/dfj/article/view/1316/1065.

166 SILVA, Tarcizio. Regular a inteligência artificial no Brasil pode mitigar o racismo algorítmico. Folha de São Paulo, publicado em 03 de jul. 2023. Disponível em: https://www1.folha.uol.com.br/blogs/politicas-e-justica/2023/05/regular-a-inteligencia-artificial-no-brasil-pode-mitigar-o-racismo-algoritmico.shtml#:~:text=Novo%20projeto%20de%20lei%20avan%C3%A7ou,combate%20aos%20danos%20do%20racismo&text=Os%20impressionantes%20saltos%20t%C3%A9cnicos%20nos,maravilha%20sobre%20as%20tecnologias%20digitais. Acesso em 21 jul. 2023.

net. Nesta normativa, foi criado um capítulo específico para a atuação do poder público em prol do desenvolvimento inclusivo da Internet no país. Nesse sentido, por exemplo, o art. 26 do MCI determina que, fruto do dever constitucional do Estado na prestação de educação, a capacitação da população inclua o uso seguro, consciente e responsável da internet como ferramenta para o exercício da cidadania, a promoção da cultura e o desenvolvimento tecnológico.

No contexto da regulação da IA, poderia haver uma capítulo programático no PL 2338/23 acerca dos deveres do Poder Público, em colaboração com a sociedade, a esse respeito. De acordo com o professor Tarcízio Silva, *"o Brasil tem riqueza humana, histórica e cultural para liderar a produção de tecnologias digitais éticas e combate a vieses de conhecimento em um mundo multipolar"*[167]. Com estímulo estatal, que pode vir por meio de normas programáticas na regulação da IA, é possível o investimento em capacitação da população para o uso e desenvolvimento de sistemas de IA para sua utilização segura, consciente e responsável.

Portanto, mesmo com as possibilidades de refinamentos no PL 2338, o projeto é hoje, ao menos, um via menos acidentada para que haja uma governança de IA em consonância com o contexto socioeconômico brasleiro enquanto país situado no sul global.

É fundamental que o Brasil construa uma regulação de IA que tenha similaridades com as discussões e modelos estrangeiros, para que haja convergência regulatória, mas levando em consideração as particularidades do país para que a regulação da tecnologia funcione para o contexto brasileiro e para as pessoas que aqui vivem, como foi iniciado por meio do PL 2338/2023. Parafraseando Cazuza *" já passou da hora de o Brasil mostrar a sua cara, ou só nos restará a festa pobre da IA para a qual nem sequer seremos convidados"*[168].

Nesse sentido, o PL 2338/2023 parece caminhar em direção à construção de um vocabulário mais afirmativo e protetivo de direitos[169], considerando também as particularidades do país enquanto integrante do mundo majoritário, que foram mencionadas ao longo desta seção. Essa abordagem é essencial para o avanço social de países desiguais, como é o caso brasileiro, para que a IA e sua regulação possam servir à benefício da sociedade brasileira e não reforçar suas práticas estruturais prejudiciais.

167 Ibid.

168 BIONI, Bruno; MENDES, Laura Schertel; ALMEIDA, Virgilio. Brasil pode liderar regulamentação da inteligência artificial. Folha de São Paulo, publicado em 13 jul. 2023. Disponível em: https://www1.folha.uol.com.br/ilustrissima/2023/07/brasil-pode-liderar-regulamentacao-da-inteligencia-artificial.shtml. Acesso em 18 jul. 2023.

169 Ibid.

MENÇÃO A PARTICULARIDADES BRASILEIRAS NO PL 2338/2023	
Tema	Conteúdo do artigo
Linguagem afirmativa de direitos	Rol amplo de fundamentos e princípios nos artigos 2º e 3º.
	Capítulo II específico para a previsão de direitos para as pessoas afetadas por sistemas de inteligência artificial.
Combate à discriminação e proteção de grupos vulneráveis	Art. 4º traz a definição de discriminação (VI) e discriminação indireta (IV).
	Seção IV específica para o direito à não-discriminação e à correção de vieses discriminatórios diretos, indiretos, ilegais ou abusivos – art. 12: Parágrafo único do art. 12 abre exceção para os casos de adoção de critérios de diferenciação em função de objetivos ou justificativas demonstradas, razoáveis e legítimas à luz do direito à igualdade e dos demais direitos fundamentais.
	Art. 18 traz como critérios para atualização da lista dos sistemas de IA de risco excessivo e alto risco o sistema ter alto potencial danoso de ordem material ou moral, bem como discriminatório (III) e o sistema afetar pessoas de um grupo específico vulnerável (IV).
	Art. 24, §1º, estipula que a avaliação de impacto deverá considerar e registrar, ao menos, o processo e resultado de testes e avaliações e medidas de mitigação realizadas para verificação de possíveis impactos a direitos, com especial destaque para potenciais impactos discriminatórios (f).
Proposições para a diminuição das assimetrias de poder e reforço do controle público e social de riscos	Obrigações de governança específicas e adicionais para o poder público para o desenvolvimento, contratação ou utilização de sistemas de IA de alto risco previstas no art. 21.
	Possibilidade de participação da sociedade na avaliação e no conhecimento dos riscos dos sistemas de IA, a partir da previsão dessa participação na AIA (art. 25, §2º).
	Obrigação de publicação das principais conclusões das avaliações de impacto em uma base de dados de inteligência artificial de alto risco, acessível ao público (art. 43).

Conclusão

Interoperabilidade regulatória: entre colonialismo e emancipação normativa

Há, atualmente, uma efervescência normativa pela qual cada vez mais a discussão não é se, mas como se regular inteligência artificial. Empilham-se propostas a nível local, regional e globais de *hard law* e *soft law* que se mostram não apenas difíceis de acompanhar, mas, sobretudo, de compará-las e compreender suas convergências e particularidades.

Esta publicação fez uma curadoria de mais de 20 (vinte) fontes normativas mapeadas a partir de 03 (três) eixos temáticas: (i) regulação baseada no risco; (ii) avaliações de impacto algorítmico; e (iii) IA Generativa, finalizando com um capítulo próprio sobre as particularidades nacionais para a regulação de IA à brasileira. Apesar de um fio condutor-comum em termos de uma racionalidade regulatória assimétrica e baseada em risco (*risk-based approach*), notou-se que não há uma homogeneização principalmente para fins de conciliação como uma abordagem que também seja afirmativa de direitos (*rights-based approach*). Citam-se abaixo algumas dessas variações, reprisando parte das conclusões traçadas no sumário executivo.

Do ponto de vista topográfico, varia-se como as propostas organizam não apenas conceitos, princípios, mas, principalmente, direitos com precedência à taxonomia de riscos e as boas práticas e medidas de governança. A escolha em enunciar primeiro direitos – preferencialmente em um capítulo(s) próprio(s) – denuncia que a *ratio legis* tinha como seu ponto primário, e não secundário ou mesmo terciário, de atenção à proteção da pessoa ou de grupos afetados pelos benefícios e riscos de IA. Portanto, a estruturação normativa é, também um indicativo da tão desejada conciliação de uma abordagem baseada em direitos e em riscos.

O apetite regulatório também é significativamente heterogêneo em listar quais situações apresentam riscos inaceitáveis-excessivos e de alto risco. Da diferença entre banimento e moratória de dados biométricos e inteligência artificial no campo da segurança pública, passando pela extensão da lista exemplificativa de proibição *ex-ante* de casos de uso IAs e chegando ao critérios quantitativos e qualitativos para dilatação-regressão dinâmica da carga regulatória mais ou menos intensa para proteção das pessoas ou dos grupos afetados. Essa interligação da lógica da classificação de riscos com direitos pode desembocar no reforço ou no esvaziamento da implementação de obrigações em jogo para fins não de qualquer tipo de inovação, mas uma que seja responsável.

Um ponto de mergulho que escancara tais nuances são as avaliações de impacto algorítmico. Se, por um lado, tal ferramenta é unanimidade, sendo listada praticamente

em todas as fontes normativas analisadas, por outro lado, o modo como é dissecada e minimamente procecdimentalizada é substancialmente distinto. A esse respeito, o PL 2338/23 avança não só quanto à necessidade de uma versão pública de tal documentação – em linha com EU AI Act e outras iniciativas vindas do Canadá, EUA e Chile–, mas, também e principalmente, no possível envolvimento de quem é impactado pelo lançamento da tecnologia em um determinado contexto. Uma tradição jurídica regulatória brasileira que advém do campo ambiental e do consumo, por exemplo. Há uma governança em rede para colocar o direito em movimento. No entanto, a proposta ainda é tímida quanto à gramática dos direitos potencializados e impactos que vão para além do plano individual, como os sociais relativos ao trabalho, ambiental, cultural e outros. Uma pauta de extrema importância para países da parte majoritária do mundo que sofrem com a extração precária de mão de obra (e.g, *data labeling*) e ilegal de minérios (e.g, construção de *chips*), o que é um novo tipo de colononialismo.

Assim, por trás do movimento de interoperabilidade regulatória não se deve minimizar divergências significativas no que diz respeito ao grau de "supervisão democrática" dos riscos toleráveis associados à IA. Isto é de um arranjo normativo de maior ou menor escrutínio público cujas distorções notam-se historicamente em outras experiências regulatórias. A exemplo do que aconteceu na crise financeira de 2008 em que não só o sistema regulatório, mas, principalmente, o seu *enforcement* foi capturado causando um colapso sistêmico em que o desenvolvimento social, econômico e tecnológico foi estruturalmente prejudicado por anos[170]. Nesta tragédia, não houve inovação responsável e a regulação foi certamente um destes gargalos.

Por isso, é urgente, em especial para países do Sul Global, enxergar as convergências e, principalmente, as divergências das alternativas regulatórias quanto ao seu grau de co-gerenciamento sobre os riscos dos usos de IAs na direção de uma porosidade social maior para desengatilhar uma abordagem sociotécnica emancipatória[171]. É necessário ficar vigilante no xadrez da chamada interoperabilidade regulatória porque há nele um novo tipo de colonialismo. Um mais "insidioso" e mais "ardiloso"[172] no qual direitos e supervisão democrática não devem ser esvaziados pela narrativa discursiva genérica de

170 COHEN, Julie E. **Between Truth and Power: The Legal Constructions of Informational Capitalism**. Oxford University Press, 2019.

171 A esse respeito, por exemplo, ver o trabalho desenvolvido pela Rede Latino-americana de Estudos sobre Vigilância, Tecnologia e Sociedade/LAVITS. Alguns trabalhos em destaque são: BRUNO, Fernanda. Racionalidade algorítmica & subjetividade maquínica. *IN*: SANTAELLA, Lucia (Org.). **Simbioses do Humano e Tecnologias: Impasses, Dilemas, Desafios**. São Paulo, SP: Editora da Universidade de São Paulo/IEA-USP, 2022; BRUNO, Fernanda; PEREIRA, Paula Cardoso; FALTAY, Paulo. Inteligência artificial e saúde: ressituar o problema. **Revista Eletrônica de Comunicação, Informação & Inovação em Saúde (RECIIS)**, vol. 17, nº 1, abr-jul, 2023. Disponível em: https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/3842.

172 Os termos entre aspas e a ideia defendida são derivados de SANTOS, Boaventura de Sousa. Boaventura de Sousa Santos: o Colonialismo e o século XXI. Centro de Estudos Estratégicos da Fiocruz, publicado em 06 de abril de 2018. Disponível em: https://cee.fiocruz.br/?q=boaventura-o-colonialismo-e-o-seculo-xxi.

uma regulação assimétrica baseada em risco. Caso contrário, não florescerão práticas efetivas de *accountability* para redução da assimetria informacional e, consequentemente, de poder[173].

173 Nesse sentido, de acordo com Bruno Bioni: "O que está em jogo não é apenas a capacidade de autoproteção (...) mas (...) como uma plêiade de atores irá mobilizar as suas respectivas prerrogativas para reduzir a assimetria de poder em jogo. E, com isso, experimentar um processo de codeliberação, e não de dominação informacional"; BIONI, Bruno Ricardo. Regulação e Proteção de Dados Pessoais – O Princípio da Accountability. São Paulo: Editora Forense, 2022. 320p. p. 245.

Referências bibliográficas

- Access Now. EU Trilogues: The AI Act must protect people's rights. Publicado em 12 jul 2023. Disponível em: https://www.accessnow.org/press-release/eu-trilogues-ai-act/.
- Access Now. Joint statement: EU legislators must close dangerous loophole in AI Act. Publicado em 07 set. 2023. Disponível em: https://www.accessnow.org/press-release/joint-statement-eu-legislators-must-close-dangerous-loophole-in-ai-act/.
- ACLU of Washington. How Automated Decision Systems are used in Policing. Publicado em 26 dez. 2022. Disponível em: https://www.aclu-wa.org/story/how-automated-decision-systems-are-used-policing.
- AI Decolonial Manyfesto. Disponível em: https://manyfesto.ai/index.html.
- Artificial Intelligence Risk Management Framework (AI RMF 1.0). Disponível em: https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf.
- ARUN, Chinmayi. AI and the Global South: Designing for Other Worlds. In: DUBBER, M.; PASQUALE, F; DAS, S. Oxford Handbook of Ethics of AI. 2019.
- Associação Data Privacy Brasil de Pesquisa (DPBR). Nota Técnica - Contribuições do Data Privacy Brasil ao Projeto de Lei nº 21, de 04 de fevereiro de 2020. Disponível em: https://www.dataprivacybr.org/wp-content/uploads/2021/09/dpbr_notatecnica_pl21.pdf.
- BAROCAS, Solon; VECCHIONE, Briana; LEVY, Karen. Algorithmic Auditing and Social Justice: Lessons from the History of Audit Studies. EAAMO '21, October 5–9, 2021, –, NY, USA. Disponível em: https://dl.acm.org/doi/pdf/10.1145/3465416.3483294.
- BENNETT, Colin J.; RAAB, Charles D., Revisiting the governance of privacy: Contemporary policy instruments in global perspective. Regulation & Governance,Vol. 14, Issue 3, p. 447-464, 2018.
- BERNSTEIN, Peter L. Against the Gods: The Remarkable Story of Risk. Wiley, 1996.
- BIONI, B.; ZANATTA, R.; RIELLI, M. (2020). Data Privacy Br: Contribuição à Consulta Pública da Estratégia Brasileira de Inteligência Artificial. São Paulo: Reticências Creative Design Studio. Disponível em: https://www.dataprivacybr.org/wp-content/uploads/2020/06/E-BOOK-CONTRIBUIC%CC%A7A%CC%83O-DPBR-INTELIGE%CC%82NCIA-ARTIFICIAL-FINAL.pdf.
- BIONI, Bruno Ricardo. Regulação e Proteção de Dados Pessoais - O Princípio da Accountability. São Paulo: Editora Forense, 2022. 320p.
- BIONI, Bruno; EILBERG, Daniela Dora; CUNHA, Brenda; SALIBA, Pedro; VERGILI, Gabriela. Proteção de dados no campo penal e de segurança pública: nota técnica sobre o Anteprojeto de Lei de Proteção de Dados para segurança pública e investigação criminal. São Paulo: Associação Data Privacy Brasil de Pesquisa, 2020.
- BIONI, Bruno; LUCIANO, Maria. O Princípio da Precaução para a Regulação da Inteligência Artificial: Seriam as Leis de Proteção de Dados seu Portal de Entrada. In: Frazão, Ana. Mullholand, Caitlin. Inteligência Artificial e Direito: ética, regulação e responsabilidade. São Paulo: Revista dos Tribunais, 2019.
- BIONI, Bruno; MENDES, Laura Schertel; ALMEIDA, Virgilio. Brasil pode liderar regulamentação da inteligência artificial. Folha de São Paulo, publicado em 13 jul. 2023. Disponível em: https://www1.folha.uol.com.br/ilustrissima/2023/07/brasil-pode-liderar-regulamentacao-da-inteligencia-artificial.shtml.
- BLACK, Julia. Proceduralisation and polycentric regulation. Revista Direito GV, Especial 1, pp. 099-130, 2005. p. 105-110.

- Blueprint for an AI Bill of Rights. Disponível em: https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf.
- BRADFORD, Anu. The Brussels Effect: How the European Union Rules the World. Nova York: Columbia Law School, mar. 2020.
- BRUNO, Fernanda; PEREIRA, Paula Cardoso; FALTAY, Paulo. Inteligência artificial e saúde: ressituar o problema. Revista Eletrônica de Comunicação, Informação & Inovação em Saúde (RECIIS), vol. 17, nº 1, abr-jul, 2023. Disponível em: https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/3842.
- BRUNO, Fernanda. Racionalidade algorítmica & subjetividade maquínica. *IN*: SANTAELLA, Lucia (Org.). Simbioses do Humano e Tecnologias: Impasses, Dilemas, Desafios. São Paulo, SP: Editora da Universidade de São Paulo/IEA-USP, 2022.
- BUOLAMWINI, Joy; GEBRU, Timnit. Gender Shades: intersectional accuracy disparities in commercial gender classification. Cambridge: Proceedings of Machine Learning Research, vol. 81, pp.1–15, 2018.
- BUOLAMWINI, Joy; RAJI, Inioluwa Deborah. Actionable Auditing: investigating the impact of publicly naming biased performance results of commercial AI products. Cambridge: Association for the Advancement of Artificial Intelligence/ACM conference on Artificial Intelligence, Ethics, and Society, 2019. Disponível em: https://www.media.mit.edu/publications/actionable-auditing-investigating-the-impact-of-publicly-naming-biased-performance-results-of-commercial-ai-products/.
- CAHAI. Human Rights, Democracy and Rule of Law Impact Assessment of AI systems. Conselho da Europa, CAHAI-PDG (2021)5. Strasbourg, 21 maio 2021. Disponível em: https://rm.coe.int/cahai-pdg-2021-05-2768-0229-3507-v-1/1680a291a3.
- CAHAI. Human Rights, Democracy and Rule of Law Impact Assessment of AI systems. Strasburgo, 11 mar. 2021. Conselho da Europa, CAHAI-PDG (2021)02.
- Center for AI and Digital Policy. "World Cup" of AI Policy News edition. CAIDP Update 5.42 - AI Policy News (Nov. 6, 2023). Disponível em: https://www.linkedin.com/posts/center-for-ai-and-digital-policy_caidp-update-542-ai-policy-news-nov-activity-7127339609293824000-fCHi/.
- CITRON, Danielle, PASQUALE, Frank. The Scored Society: Due Process for Automated Predictions. Washington Law Review, Vol. 89, 2014.
- Coalizão Direitos na Rede (CDR). Carta de Apoio ao PL 2338/2023. Publicado em 14 jun. 2023. Disponível em: https://direitosnarede.org.br/2023/06/14/carta-de-apoio-ao-pl-2338-2023/. Acesso em 18 jul. 2023.
- Coalizão Direitos na Rede. Nota Técnica do Projeto de Lei nº 2338/2023. Agosto de 2023. Disponível em: https://direitosnarede.org.br/2023/08/23/coalizao-direitos-na-rede-divulga-nota-tecnica-sobre-o-pl-2338-2023-que-busca-regular-a-ia/.
- Coalizão Direitos na Rede. Inteligência Artificial não pode ser regulada a toque de caixa. Publicado em 23 de setembro de 2021. Disponível me: https://direitosnarede.org.br/2021/09/23/inteligencia-artificial-nao-pode-ser-regulada-a-toque-de-caixa/; Coalizão Direitos na Rede. Brasil não está pronto para regular inteligência artificial. Publicado em 07 de dezembro de 2023. Disponível em: https://direitosnarede.org.br/2021/12/07/brasil-nao-esta-pronto-para-regular-inteligencia-artificial/.
- COHEN, Julie E. Between Truth and Power: The Legal Constructions of Informational Capitalism. Oxford University Press, 2019.
- COLOMBO, Silvana. Os mecanismos de participação popular na gestão do meio ambiente à luz do texto constitucional: aspectos positivos e negativos. Editora Unijuí: Revista Direitos Humanos e Democracia, ano 9, nº 18, jul/dez 2021.

- Comissão Europeia. EU-U.S. Terminology and Taxonomy for Artificial Intelligence. Pulicado em 31 maio 2023. Disponível em: https://digital-strategy.ec.europa.eu/en/library/eu-us-terminology-and-taxonomy-artificial-intelligence.
- Comissão de Juristas Responsável por Subsidiar a Elaboração de Substitutivo sobre Inteligência Artificial no Brasil (CJSUBIA). Relatório Final. 2022. Disponível em: https://www.stj.jus.br/sites/portalp/SiteAssets/documentos/noticias/Relato%CC%81rio%20final%20CJSUBIA.pdf.
- Conselho Nacional do Ministério Público. Portal de Direitos Coletivos. Disponível em: https://www.cnmp.mp.br/direitoscoletivos/.
- Consolidated Working Draft of the Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law - Committee on Artificial Intelligence (CAI). Strasburg, 7 de julho de 2023.
- COSTA, Ramon; KREMER, Bianca. Inteligência Artificial e Discriminação: Desafios e Perspectivas para a Proteção de Grupos Vulneráveis diante das Tecnologias de Reconhecimento Facial. Direitos Fundamentais & Justiça | Belo Horizonte, ano 16, número especial, p. 145-167, outubro 2022. Disponível em: https://dfj.emnuvens.com.br/dfj/article/view/1316/1065.
- COSTANZA-CHOCK, Sasha. Design Justice, A.I., and escape from the matrix of domination. Camrbidge: Journal of Design and Science, jul. 2018. DOI:10.21428/96c8d426. Disponível em: https://jods.mitpress.mit. edu/pub/costanza-chock/release/4;
- COSTANZA-CHOCK, Sasha. Design Practices: "Nothing about Us without Us". Design Justice, publicado em 26 fev. 2020. Disponível em: https://designjustice.mitpress.mit.edu/pub/cfohnud7/release/4.
- Council of the EU. Artificial Intelligence Act: Council calls for promoting safe AI that respects fundamental rights. Press Release, publicado em 6 dec. 2022. Disponível em: https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/
- DA SILVA, Paula Guedes Fernandes. Inteligência Artificial na União Europeia: formas de regular a tecnologia que já nos regula. In: MENDES, Gilmar Ferreira; DE MORAIS, Carlos Blanco. Governance da Ordem Jurídica em Transformação. Anais do X Fórum Jurídico de Lisboa, 2022, p. 589. Disponível em: https://www.forumjuridicodelisboa.com/2023-anais.
- DA SILVA, Paula Guedes Fernandes; et al. Avaliação de impacto algorítmico: o que é e como está regulada no PL 2.338/23 do Brasil. Migalhas, publicado em 19 out. 2023. Disponível em: https://www.migalhas.com.br/coluna/migalhas-de-responsabilidade-civil/395547/avaliacao-de-impacto-algoritmico.
- DA SILVA, Paula Guedes Fernandes; GARROTE, Marina Gonçalves. Insuficiência dos princípios éticos para normatização da Inteligência Artificial: o antirracismo e a anti-discriminação como vetores da regulação de IA no Brasil. POLITICS, setembro de 2022. Disponível em: https://politics.org.br/edicoes/insufici%C3%AAncia-dos-princ%C3%ADpios-%C3%A9ticos-para-normatiza%C3%A7%C3%A3o-da-intelig%C3%AAncia-artificial-o.
- DARIUSZ, Kloza. Privacy Impact Assessment as a Means to Achieve the Objectives of Procedural Justice, Jusletter IT. Die Zeitschrift für IT und Recht, disponível em: https://cris.vub.be/files/49868387/Kloza_2014_PIA_as_a_Means_to_Achieve_the_Objectives_of_Procedural_Justice.pdf.
- Data & Society. Algorithmic Impact Methods Lab. Data & Society Announces the Launch of its Algorithmic Impact Methods Lab. Nova York, 10 mai. 2023. Disponível em: https://datasociety.net/algorithmic-impact-methods-lab.

- Data Privacy Brasil Research. Nota Técnica - Contribuições do Data Privacy Brasil ao Projeto de Lei nº 21, de 04 de fevereiro de 2020. Disponível em: https://www.dataprivacybr.org/wp-content/uploads/2021/09/dpbr_notatecnica_pl21.pdf.
- Data Privacy Brasil. Dadocracia - Ep. 78 - Marco Legal da IA. Dadocracia, publicado em nov. 2021. Disponível em: https://open.spotify.com/episode/15BWzRa4cWVRo0jtGGPm4T?si=v7X-iVnWQ3e-eIArlGmKaUg.
- Data Privacy Brasil. Dadocracia - Ep. 78 - Marco Legal da IA. Dadocracia, publicado em nov. 2021. Disponível em: https://open.spotify.com/episode/15BWzRa4cWVRo0jtGGPm4T?si=v7X-iVnWQ3e-eIArlGmKaUg;
- Data Privacy Brasil. Dadocracia - Ep. 80 - Mais Marco Legal da IA. Dadocracia, publicado em dez. 2021. Disponível em: https://open.spotify.com/episode/0t4Rr07EwIjrdpmvzht79Z?si=OiOyUXc0T-5-kH0nr6qhzKA&nd=1;
- Demetriodor. O 'Sul Global' é um termo terrível. Não use! Publicado em 11 nov. 2018. Disponível em: re-design.dimiter.eu/?p=969;
- ECNL; Society Inside. Framework for Meaningful Engagement. Disponível em: https://ecnl.org/sites/default/files/2023-03/Final%20Version%20FME%20with%20Copyright%20%282%29.pdf.
- Estadão. "Mais importante lei de recnologia no Brasil não está sendo debatida", diz especialista. Bruno Romani, publicado em 07 dez. 2021. Disponível em: https://www.estadao.com.br/link/cultura-digital/mais-importante-lei-de-tecnologia-no-brasil-nao-esta-sendo-debatida-diz-especialista/.
- FAUSTINO, Deivison; LIPPOLD, Walter. Colonialismo Digital: por uma crítica hacker-fenoniana. 1ª ed. São Paulo: Boitempo, 2023.
- FERRAZZO, Débora; DUARTE, Francisco Carlos. Colonização jurídica na América Latina. Disponível em: www.publicadireito.com.br/artigos/?cod=f376b8ae6217d18c.
- FIGUEIRA, Paulo Sérgio Sampaio. O papel do conselho do meio ambiente nas políticas públicas ambientais. Publicado em 14 de abril de 2022. Disponível em: https://direitoambiental.com/o-papel-do-conselho-do-meio-ambiente-nas-politicas-publicas-ambientais/.
- Folha de São Paulo. Brasil apressa lei para inteligência artificial, dizem especialistas. Amanda Lemos, publicado em 18 jul. 2021. Disponível em: https://www1.folha.uol.com.br/mercado/2021/07/brasil-apressa-lei-para-inteligencia-artificial-dizem-especialistas.shtml.
- GAJARDONI, Fernando da Fonseca. Direitos Difusos e Coletivos I: Teoria Geral do Processo Coletivo. São Paulo: Saraiva, 2012.
- GARROTE, Marina. Regulating Artificial Intelligence in Brazil. Center for Human Rights & Global Justice, NYU School of Law, publicado em 28 set. 2023. Disponível em: https://chrgj.org/2023/09/28/regulating-artificial-intelligence-in-brazil/.
- GASPAR; Walter B.; DE MENDONÇA, Yasmin Curzi. A Inteligência Artificial no Brasil ainda precisa de uma estratégia. Relatório do Centro de Tecnologia e Sociedade da FGV Direito Rio. Maio de 2021. Disponível em: https://bibliotecadigital.fgv.br/dspace/bitstream/handle/10438/30500/EBIA%20pt-br.pdf?sequence=3&isAllowed=y.
- GELLERT, Raphaël. Understanding the notion of risk in the General Data Protection Regulation. Computer Law & Security Review: The International Journal of Technology Law and Practice (2017).
- GOMES, Maria Cecília O.Entre o método e a complexidade: compreendendo a noção de risco na LGPD. *In:* Temas atuais de proteção de dados. PALHARES, Felipe (Coord.). São Paulo: Thomson Reuters Brasil, 2020, pp 245-271.
- GOMES, Maria Cecília. Relatório de impacto à proteção de dados: uma breve análise de sua definição

e papel na LGPD. Revista da AASP, n. 144, 2019. p. 10-11.

- Governo do Canadá. Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems. Setembro de 2023. Disponível em: https://ised-isde.canada.ca/site/ised/en/voluntary-code-conduct-responsible-development-and-management-advanced-generative-ai-systems.
- HACKER, Philipp. Sustainable AI Regulation. Privacy Law Scholars Conference 2023. Disponível em: https://arxiv.org/abs/2306.00292.
- HACKER, Philipp; ENGEL, Andreas; MAUER, Marco. Regulating ChatGPT and other Large Generative AI Models. Fairness, Accountability, and Transparency (FAccT '23), June 12–15, 2023. Disponível em: https://dl.acm.org/doi/10.1145/3593013.3594067.
- HEINE; Jorge. O Sul Global está em ascensão – mas o que é exatamente o Sul Global? Interesse Nacional, publicado em 10 de jul. 2023. Disponível em: https://interessenacional.com.br/edicoes-posts/o-sul-global-esta-em-ascensao-mas-o-que-e-exatamente-o-sul-global/.
- HOOD, Christopher; ROTHSTEIN, Henry; BALDWIN, Robert.The Governance of Risk: Understanding Risk Regulation Regimes. Nova York: Oxford University Press, 2001. ISBN 0-19-924363-8.
- Jjeld, Jessica and Achten, Nele and Hilligoss, Hannah and Nagy, Adam and Srikumar, Madhulika, Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI (January 15, 2020). Berkman Klein Center Research Publication No. 2020-1, Available at SSRN: https://ssrn.com/abstract=3518482.
- KAMISNKI, 2022, p. 36; BOYD, Willian. Genealogies of Risk: Searching for Safety, 1930s-1970s. Ecology Law Quarterly, nº 895, 2012. Disponível em: https://scholar.law.colorado.edu/faculty-articles/143/.
- KlOZA, D., et al. (2017). Data protection impact assessments in the European Union: complementing the new legal framework towards a more robust protection of individuals. d.pia.lab Policy Brief, (1/2017), 1-4. https://doi.org/10.31228/osf.io/b68em, https://doi.org/10.5281/zenodo.5121575
- KLOZA, D., et al. (2019). Towards a method for data protection impact assessment: Making sense of GDPR requirements. d.pia.lab Policy Brief, 1(2019), 1-8. https://doi.org/10.31228/osf.io/es8bm, https://doi.org/10.5281/zenodo.5121534.
- LOPES, Larissa. Já ouviu falar na prática do Deep Nude? Jusbrasil, publicado em outubro 2023. Disponível em: https://www.jusbrasil.com.br/artigos/ja-ouviu-falar-na-pratica-do-deep-nude/1979706886
- MARQUES, Claudia Lima; BENHAMIN, Antônio Herman. A Teoria do Diálogo das Fontes e seu Impacto no Brasil: uma homenagem a Erik Jayme. Revista de Direito do Consumidor (RDC) 2340, 115 indb, 2018. Disponível em: https://revistadedireitodoconsumidor.emnuvens.com.br/rdc/article/view/1042/911.
- Ministério da Ciência, Tecnologia e Inovações (MCTI). Estratégia Brasileira de Inteligência Artificial (EBIA). Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivosinteligenciaartificial/ebia-documento_referencia_4-979_2021.pdf.
- MONTELERO, A. (2022). Beyond Data - Human Rights, Ethical and Social Impact Assessment in AI. Asser Press, Information Technology and Law Series (IT&Law), Vol. 36.
- OECD. OECD Framework for the Classification of AI systems. OECD Digital Economy Papers, No. 323, OECD Publishing, Paris, 2022. Disponível em: https://doi.org/10.1787/cb6d9eca-en.
- OECD. Advancing accountability in AI: Governing and managing risks throughout the lifecycle for trustworthy AI. Publicado em 23 Feb. 2023. Disponível em: https://www.oecd-ilibrary.org/science-and-technology/advancing-accountability-in-ai_2448f04b-en.

- OECD. G7 Hiroshima Process on Generative Artificial Intelligence (AI): Towards a G7 Common Understanding on Generative AI. Relatório preparado para a presidência japonesa de 2023 e para o grupo de trabalho digital e tecnológico do G7. Publicado em 7 set. 2023. Disponível em: https://www.oecd-ilibrary.org/deliver/bf3c0c60-en.pdf?itemId=%2Fcontent%2Fpublication%2Fbf3c0c60-en&mimeType=pdf.
- OECD. Initial policy considerations for generative artificial intelligence. Publicado em 18 set. 2023. Disponível em: https://www.oecd-ilibrary.org/deliver/fae2d1e6-en.pdf?itemId=%2Fcontent%2Fpaper%2Ffae2d1e6-en&mimeType=pdf.
- OECD. Common guideposts to promote interoperability in AI risk management. Publicado em 07 nov. 2023. Disponível em: https://www.oecd-ilibrary.org/science-and-technology/common-guideposts-to-promote-interoperability-in-ai-risk-management_ba602d18-en.
- QUELLE, Claudia. 'The 'risk revolution' in EU data protection law: We can't have our cake and eat it, too' in R Leenes, R van Brakel, S Gutwirth and P De Hert (eds), Data Protection and Privacy: The Age of Intelligent Machines (Hart Publishing, forthcoming). 2017.
- QUELLE, Claudia. Does the risk-based approach to data protection conflict with the protection of fundamental rights on a conceptual level? Tilburg Law School Research Paper , 1-36, 2015.
- SANTOS, Boaventura de Sousa. Boaventura de Sousa Santos: o Colonialismo e o século XXI. Centro de Estudos Estratégicos da Fiocruz, publicado em 06 de abril de 2018. Disponível em: https://cee.fiocruz.br/?q=boaventura-o-colonialismo-e-o-seculo-xxi.
- SANTOS Boaventura de Sousa. Construindo as Epistemologias do Sul: Antologia Essencial. Volume I: Para um pensamento alternativo de alternativas. Coleção Antologias do Pensamento Social Latino-Americano e Caribenho, 1ª Ed, 2018.
- SILVA, Mariah Rafaela; VARON, Joana. Reconhecimento Facial no Setor Público e Identidades Trans: tecnopolíticas de controle e a ameaça à diversidade de gênero em suas interseccionalidades de raça, classe e território. Uma pesquisa realizada pela Coding Rights com apoio da ONG Privacy International via financiamento do International Development Research Center (IDRC). Rio de Janeiro: jan. 2021;
- SILVA, Tarcizio. Linha do Tempo do Racismo Algorítmico. Blog do Tarcizio Silva, 2022. Disponível em: https://tarciziosilva.com.br/blog/posts/racismo-algoritmico-linha-do-tempo.
- SILVA, Tarcizio. Regular a inteligência artificial no Brasil pode mitigar o racismo algorítmico. Folha de São Paulo, publicado em 03 de jul. 2023. Disponível em: https://www1.folha.uol.com.br/blogs/politicas-e-justica/2023/05/regular-a-inteligencia-artificial-no-brasil-pode-mitigar-o-racismo-algoritmico.shtml#:~:text=Novo%20projeto%20de%20lei%20avan%C3%A7ou,combate%20aos%20danos%20do%20racismo&text=Os%20impressionantes%20saltos%20t%C3%A9cnicos%20nos,maravilha%20sobre%20as%20tecnologias%20digitais.
- SOLAIMAN, Irene. The Gradient of Generative AI Release: Methods and Considerations. Fevereiro de 2023. Disponível em: https://arxiv.org/abs/2302.04844.
- Southern Alliance for the Global Digital Compact: contribution for the promotion of digital human rights.2023. Disponível em: https://www.dataprivacybr.org/documentos/southern-alliance-for-the-global-digital-compact/.
- TRUBEK, David M.; COTRELL, Patrick; NANCE, Mark. "Soft Law," "Hard Law," and European Integration: Toward a Theory of Hybridity. Legal Studies Research Paper Series, Winsconsin, n. 1002, p. 1-42, nov. 2005. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=855447.
- TV Senado. Comissão de juristas promove debates sobre regulação da inteligência artificial (2ª par-

te)– 29/04/22. Publicado em 29 abr. 2022. Fala da professora Maria Cecília Gomes. Disponível em: https://www.youtube.com/watch?v=P_yWp-2ZIZs&t=51s. Acesso em 21 jul. 2023.

- TV Senado. Inteligência artificial: instalação da comissão de juristas que vai analisar o tema - 30/03/22. Realizada em 30 mar. 2022. Disponível em: https://www.youtube.com/watch?v=nXnliBi-3vKY&ab_channel=TVSenado.
- UNESCO. Ethical Impact Assessment: A Tool of the Recommendation on the Ethics of Artificial Intelligence. Publicado em 2023. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000386276/PDF/386276eng.pdf.multi.
- UNESCO. Recommendation on the Ethics of Artificial Intelligence. Adotado em 23 novembro de 2021 e publicado em 2022. Disponível em: https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence;
- VARON, Joana; SILVA, Mariah Rafaela. Reconhecimento facial no setor público e identidades trans: tecnopolíticas de controle e ameaça à diversidade de gênero em suas interseccionalidades de raça, classe e território. Disponível em: <https://codingrights.org/docs/rec-facial-id-trans.pdf>.
- WRIGHT, David et al. Integrating privacy impact assessment in risk management. International Data Privacy Law, v. 4, n. 2, p. 155–170, 2014.
- YANG, Zeyi. China just announced a new social credit law. Here's what it means. MIT Technology Review, publicado em 22 nov. 2022. Disponível em: https://www.technologyreview.com/2022/11/22/1063605/china-announced-a-new-social-credit-law-what-does-it-mean/.
- ZANATTA, Rafael A. F. Proteção de Dados Pessoais como Regulação de Risco: uma nova moldura técnica?. I ENCONTRO DA REDE DE PESQUISA EM GOVERNANÇA DA INTERNET, NOVEMBRO DE 2017. Disponível em: https://www.redegovernanca.net.br/public/conferences/1/anais/ZANATTA,%20Rafael_2017.pdf.