Contribuição à consulta pública da Comissão de Juristas do Senado responsável por subsidiar a elaboração de substitutivo sobre inteligência artificial no Brasil - **CJSUBIA**

Junho, 2022

IP.rec

INSTITUTO DE PESQUISA EM DIREITO & TECNOLOGIA DO RECIFE

Ficha Técnica

Realização:
Instituto de Pesquisa em Direito e Tecnologia do Recife - IP.rec

Equipe:
Coordenação:
André Lucas Fernandes

Autores:
André Lucas Fernandes
Carolina Branco
Clarissa Mendes
Laura Pereira
Lucas Santana
Raquel Saraiva
Rodrigo Alexandre

Revisão:
André Lucas Fernandes
Raquel Saraiva

Design e diagramação:
Clara Guimarães

www.ip.rec.br/ /institutoiprec contato@ip.rec.br
@institutoiprec @ip.rec

INSTITUTO DE PESQUISA EM DIREITO & TECNOLOGIA DO RECIFE

Sumário

3. Direitos e deveres

4. Accountability, governança e fiscalização

Contribuição à consulta pública da Comissão de Juristas do Senado responsável por subsidiar a elaboração de substitutivo sobre inteligência artificial no Brasil - CJSUBIA

1. Conceitos e compreensão de inteligência artificial

1.1. Objeto a ser regulado

Na justificativa da versão original do Projeto de Lei 21/2020, apresentado pelo Deputado Eduardo Bismark (PDT/CE) em fevereiro de 2020, fala-se em como as transformações causadas pela presença cada vez mais forte e mais alastrada da inteligência artificial fazem com que seja imperativa a necessidade de legislar sobre direitos e deveres envolvidos no desenvolvimento e uso dessas aplicações. A persistência do desafio que orientou o projeto que viria a se tornar a principal referência na elaboração de um Marco Regulatório de Inteligência Artificial no Brasil, diligência que é o objeto desta consulta pública, é prova simultânea de sua complexidade e importância pública.

A combinação desses fatores significa que o debate deve ser aberto, plural, interdisciplinar, e contar com representação regional e multissetorial. O reconhecimento tardio dessa condição inerente é parte fundamental do necessário esforço coletivo representado pelas audiências e consultas da Comissão, razão pela qual o Instituto de Pesquisa em Direito e Tecnologia do Recife - IP.rec elogia a iniciativa e apresenta suas contribuições, com a expectativa de que essa direção seja fortalecida ao longo de todo o processo legislativo referente à matéria.

A proposta de elaboração de um Marco Normativo para a Inteligência Artificial no Brasil precisa ser conduzida com o devido cuidado e a indispensável abertura à participação efetiva da sociedade civil. Isso porque,

se a urgência aventada pela justificativa do PL 21/20 tem razão de ser, também é verdade que mesmo este projeto foi alvo de incertezas basilares ao longo da sua apressada trajetória na Câmara dos Deputados. A primeira questão que se impõe diz respeito à própria definição de inteligência artificial e, consequentemente, do escopo de aplicação da proposição. O IP.rec considera que o amadurecimento do debate quanto **ao objeto a ser regulado** é ponto fundamental para a compreensão quanto às razões, os princípios e os procedimentos envolvidos em tal regulação.

Não há uma definição consensual do que é inteligência artificial e tampouco pode-se afirmar que essa seja a melhor terminologia para embasar um marco normativo que tenha como escopo as aplicações popularmente conhecidas como pertencentes a esse conjunto de tecnologias[1]. A ausência de uma definição balizada é um dado no campo científico especializado [2] e, por consequência, é ponto a ser considerado também pelo legislador[3]. Não por acaso, foi questão repetidamente apontada em contexto de definições internacionais de valor normativo ou referencial, como nas Recommendation of the Council on Artificial Intelligence da OCDE[4], nas Recomendações da Unesco (2021)[5] e no debate em torno da elaboração e proposição do AI Act da União Europeia [6], marco proposto pelo Conselho Europeu.

Consideramos que os textos dos projetos de lei considerados nesta consulta pública não são exitosos em estabelecer uma definição funcional para fins legais. A definição proposta no Art. 2º do PL 21/20 dialoga diretamente com parte do conceito proposto pela OCDE (2021)[7], mas inclui elementos que

[1] (CANALES, 2019; SCHUETT, 2019).
[2] (WANG, 2018).
[3] (ZIEMIANIN, 2021).
[4] (OECD, 2019)
[5] (UNESCO, 2019).
[6] (BRYSON, 2022), em artigo na revista Wired.
[7] A definição da OCDE para sistemas de IA é "An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy". A da UNESCO fala em "Therefore, this Recommendation approaches AI systems as systems which have the capacity to process data and information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning or control". Outras definições enfatizam a ideia de sistemas de informação que conseguem se adaptar a diferentes ambientes ou contextos (WANG, 2019).

prejudicam sua precisão e compreensão. É o que ocorre, por exemplo, ao estabelecer como critério a capacidade ampla e pouco objetiva de "aprender a perceber e a interpretar o ambiente externo, bem como a interagir com ele". Não se mostra claro como se dariam as ações de "perceber", "interpretar" e "interagir" no âmbito de uma aplicação de IA. Nesse sentido, consideramos que a definição do objeto a ser regulado deve ser prontamente revisitada, avaliando-se os benefícios em incluir uma definição do tipo, em primeiro lugar, e quais elementos e enfoques devem tomar parte dela, em caso afirmativo.

1.2. Aspectos sócio-técnicos da IA

O exercício de definição do objeto a ser abarcado pelo marco regulatório vincula-se diretamente aos princípios, objetivos e determinações nele estabelecidas. Nesse sentido, a proposta de se retornar às questões de terminologia e de elementos envolvidos no que se entende como sendo "inteligência artificial" deve ser acompanhada de uma abordagem que incorpore **integralmente e plenamente** os aspectos sociotécnicos do tema.

Trata-se, aqui, de qualificar a abordagem da lei para aperfeiçoar o tratamento dado às questões críticas e decisivas da crescente utilização de sistemas automatizados de decisão nas mais diferentes áreas da vida humana. Isso significa considerar que esses sistemas são formados por componentes humanos, técnicos e institucionais que se articulam em arranjos que não são necessariamente positivos, benéficos e funcionais. A presença de aspectos contextuais e sociais é parte indissociável do desenvolvimento e uso de inteligência artificial, o que não pode ser secundarizado no corpo da lei. Essa condição sociotécnica se expressa nos objetivos, valores e escolhas que são incorporados pela técnica, na finalidade, no conceito utilizado para a definição de riscos até no design das suas características e se estende até as implicações da aplicação em determinado contexto.

Assim, não há razões para adotar a premissa do benefício final ou da possibilidade indefinida de possível correção ou aperfeiçoamento,

especialmente quando há evidências contrárias a essa premissa[8]. O abandono da premissa não significa o impedimento à inovação e ao desenvolvimento técnico-científico, mas um passo na direção de uma avaliação holística quanto ao grau e a natureza dos problemas que podem ser enfrentados por normativas centradas no interesse público e no ser humano.

Hoje, a inteligência artificial faz parte da operação diária de sistemas públicos que organizam segmentos como a previdência[9] e a segurança[10]. Estudo da ONG Derechos Digitales avaliou que a implementação de sistema automatizado no Sistema Nacional de Emprego (SINE) se deu de forma opaca, sem a adoção de parâmetros mínimos de transparência e participação pública ao longo do ciclo de vida do projeto, ou seja, desde a etapa de pesquisa e diagnóstico até a avaliação e modelação do seu funcionamento. Com graves riscos prejudiciais aos direitos humanos, os problemas identificados foram vários: desde a opacidade quanto à utilização de dados pessoais por parte da empresa responsável pelo sistema até as dificuldades e enviesamento que ele derivava para grupos sociais vulneráveis.

Tais danos ocorreram mesmo em um cenário no qual o Brasil possui uma legislação avançada de proteção de dados pessoais e é destaque em relação à implementação de ferramentas de governo eletrônico e segurança digital. O diagnóstico é evidente: mesmo com procedimentos e regramentos tão necessários e importantes, há brechas e insuficiências. Na avaliação dos autores, a implementação de sistemas que atravessam diretamente o campo das políticas públicas deve contar, necessariamente, com a participação pública significativa e de procedimentos que vão da decisão pelo desenvolvimento e implementação até à avaliação e monitoramento frequente [11]. Por isso, são problemas que poderiam ser enfrentados pela incorporação prioritária, qualitativa, significativa e integral dos elementos envolvidos na hora de se optar pela importação, desenvolvimento e adoção de um sistema de

[8] (SLOANE e MOSS,2019).
[9] (GERCINA, 2022)
[10] (PEET, 2021)
[11] (FUENTES e VENTURINI, 2021)

inteligência artificial que impacta diretamente na garantia aos direitos humanos de cidadãos brasileiros.

A gravidade do tema exige que a legislação estabeleça um arcabouço legal robusto, que vá além da prescrição de recomendações de ética e boas práticas que se restringem a uma abordagem principiológica. Se há a compreensão de que é necessário firmar um marco específico para o objeto em questão, ele deve servir para suprir as lacunas e insuficiências de outros regramentos já consolidados e que podem servir de fonte normativa em determinados casos, como o Código de Defesa do Consumidor ou a própria Lei Geral de Proteção de Dados. Por isso, consideramos que a centralidade do ser humano não está suficientemente afirmada e operacionalizada na versão atual do PL 21/20. Esse diagnóstico será explorado nas próximas seções, já que ele se expressa no modelo de responsabilidade sugerido e na ausência de prerrogativas bem definidas quanto ao direito à intervenção humana, por exemplo.

1.3. Por que e como regular

A princípio, a escolha legislativa brasileira em definir norma específica para a inteligência artificial encontra correspondência no diagnóstico especializado e no debate global quanto às insuficiências para a abordagem do tema apenas em regramentos tradicionais, provenientes de áreas que tangenciam o tema[12].

Contudo, a determinação de regra específica não é automaticamente equivalente ao aperfeiçoamento da legislação. A depender das prioridades estabelecidas e do nível de amadurecimento da pauta, o debate acelerado e restrito pode contribuir para criar legislações inócuas ou danosas, que pouco acrescentam ao quadro regulatório já existente e que, inclusive, podem fragilizá-lo. A possibilidade de desvio da atividade legislativa faz com que seja

[12] (ZIEMIANIN, 2021)

indispensável o esgotamento de debate aberto, interdisciplinar, multissetorial e com representação regional.

Por isso, é preciso que as bases que justificam a elaboração de norma específica estejam bem estabelecidas e que as funções de um Marco Regulatório sejam bem compreendidas pelos seus autores. Considerando que os instrumentos já disponíveis para a regulação de IA abarcam aplicações, por exemplo, do Código de Defesa do Consumidor, entendemos que a legislação específica deve ser enfática e suficiente para dirimir dúvidas e lacunas que não são resolvidas pelo CDD ou, em outro exemplo, pela Lei Geral de Proteção de Dados. Essa demanda se dá em função tanto das características do objeto a ser regulado quanto da urgência que se muito aventa em relação ao tema, tão presente em âmbito nacional e internacional.

Disso, pode-se pontuar objetivos que justifiquem e motivem a busca por um regramento geral, como garantir segurança jurídica, estímulo à inovação e a defesa de direitos fundamentais. A partir da perspectiva da sociedade civil, entendemos que a legislação deve ter como ponto de partida a centralidade do ser humano e a determinação de obrigações e procedimentos que assegurem que o ciclo de vida dos sistemas automatizados passe por crivo social correspondente à importância que esses sistemas têm assumido na vida de toda a população.

Por essas razões, consideramos que há grave equívoco no caráter principiológico e generalista do PL 21/2020. A versão atual do projeto, apressadamente aprovada na Câmara dos Deputados, não justifica seu papel como legislação específica e tampouco expressa determinações concretas para a matéria, sem a definição de obrigações, sanções e instrumentos de fiscalização. Sem a devida robustez e amadurecimento que devem ser almejados neste processo consultivo, o Marco Regulatório pode se tornar instrumento exclusivo de segurança jurídica e estímulo irrestrito à inovação, falhando em responder adequadamente aos objetivos que mais atendem aos interesses da sociedade.

1.4. Princípios da IA

Tratando-se da implementação de Inteligência Artificial, a ética é ponto de partida, porém não se esgota em si mesma e nem satisfaz, visto que a implementação dessa tecnologia tem grande potencial modificativo sobre dinâmicas sociais e econômicas. No Estado Democrático de Direito deve haver um compromisso normativo de promoção e proteção dos direitos humanos, e por isso as tecnologias que impactam o exercício dos direitos sociais, políticos, civis, culturais e econômicos, demandam necessariamente intervenção regulatória. Para tanto, tal intenção deve também ser concretizada na fixação de princípios, os quais possuem função de condutor, um guia, um norte da busca da compreensão do significado da norma e dos institutos jurídicos.[13]

À vista disso, a OCDE estabeleceu em seu documento "Recommendation of the Council on Artificial Intelligence"[14] alguns princípios a serem seguidos no processo de regulamentação da inteligência artificial. São eles: o crescimento inclusivo, o desenvolvimento sustentável e bem-estar; os valores centrados no homem e justiça; a transparência e explicabilidade; a robustez, segurança e proteção; e a responsabilização.

Ao exigir o **crescimento inclusivo, desenvolvimento sustentável e bem-estar**, o documento propõe que a finalidade da utilização da IA deva ser a promoção de benefícios às pessoas humanas e ao planeta, citando em rol exemplificativo o estímulo da criatividade e inclusão de populações sub-representadas, a redução de desigualdades, e a proteção do meio-ambiente. Já ao apontar a necessidade de **valores centrados no ser humano e equidade**, quis garantir o respeito aos direitos fundamentais e humanos, valores democráticos, Estado de Direito e diversidade, buscando a salvaguarda de uma supervisão e/ou intervenção humana, sempre que necessário.

Quanto à **transparência e explicabilidade**, empenha-se em garantir aos usuários informações claras, precisas e facilmente acessíveis

[13] (DELGADO, 2007).
[14] (OECD, 2019)

sobre as implicações e utilizações de sistemas de inteligência artificial, exigindo também, que seu funcionamento seja explicável para o usuário, além da necessidade de identificação objetiva quando se tratar de um sistemas automatizado e não de seres humanos.

Ao mencionar a **robustez, segurança e proteção**, tem por objetivo a gestão e a avaliação dos riscos dos sistemas de IA durante toda a sua vida útil, de forma que possam funcionar de acordo com um planejamento prévio, viabilizando, dessa forma, uma possível investigação do conjunto de dados utilizados no treinamento e funcionamento de processos e decisões tomadas pela IA.

Por fim, a **responsabilização** garantirá que os atores engajados no desenvolvimento de sistemas de decisões automatizadas sejam responsabilizados de acordo com, no mínimo, os princípios já elencados, e por consequência exigirá, também, a demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas estabelecidas, além da eficácia dessas medidas..

Apesar da regulação dos dados pessoais não ser a única uma peça essencial para o bom funcionamento da IA, é importante também ressaltar os princípios já implementados na **LGPD**, fundamentais na discussão do tratamento de dados desses sistemas. São eles: a finalidade, a adequação, a necessidade, o livre acesso, a qualidade dos dados, a transparência, a segurança, a prevenção, a não-discriminação e a responsabilização e prestação de contas. Desses, destaca-se o **princípio da não discriminação**, que determina a impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos. Aqui, destaca-se que deve haver uma amplificação, para que sua leitura não seja apenas sobre a intencionalidade, mas sim sobre as consequências efetivas da implementação da tecnologia em questão.

Não obstante, é também necessário discutir a importância da incorporação do **princípio da vulnerabilidade** quando se trata das regulamentações desenvolvidas em torno de sistemas automatizados de tomada de decisão. O usuário, pessoa física, natural, deve ter aqui a sua

vulnerabilidade presumida (absoluta), enquanto a da pessoa jurídica deve ser aferida no caso concreto. Tal princípio busca proteger a parte mais frágil do polo, a fim de promover o equilíbrio de uma relação substancialmente desproporcional.

Isso posto, é fundamental que haja compreensão da importância da produção das normas guiada por esses princípios, a fim de garantir não somente a pacificação social diante das profundas mudanças de dinâmicas sociais e econômicas originadas a partir da implementação dos sistemas de IA, mas também a proteção social em relação aos consequentes riscos existentes, materializados na limitação do exercício de direitos coletivos que impactam grupos inteiros de indivíduos, potenciais vítimas de discriminação, vigilância e censura em massa por essas tecnologias.

2. Impactos da inteligência artificial

2.1. Benefícios da IA

A Inteligência Artificial evidentemente possibilita enormes vantagens, e é amplamente utilizada nos mais diversos setores. Segundo Nadimpalli[15], um dos benefícios foi o aumento do desempenho de profissionais de saúde, em que se pode utilizar de sistemas informáticos especialmente desenvolvidos para identificar pacientes de alto risco, analisar com precisão problemas fisiológicos específicos e tomar uma iniciativa mais rápida, como também auxiliar no processo da tomada de decisão e poupar tempo aos médicos, além de outros benefícios.

Setores como o de logística e transporte também utilizam inteligência artificial para auxiliar e garantir maior desempenho em seus processos. Já a área financeira e a bancária se beneficiam na melhoria de monitoramento de suas atividades e análise mais rápida de possíveis problemas. Nadimpalli[16] ainda pontua que a inteligência artificial foi colocada

[15] (NADIMPALLI, 2019)
[16] Idem.

em uso em áreas que oferecem risco à vida humana, como por exemplo na indústria de mineração, que ao utilizar veículos e máquinas que podem operar sem ser comandados por uma pessoa no subsolo é possível proteger os trabalhadores.

Conforme Gomes[17], a inteligência artificial sistematiza e automatiza tarefas intelectuais, como também abrange uma variedade de subcampos e áreas. Nesse sentido é possível exemplificar algumas aplicações, como o desenvolvimento de sistemas especialistas, que processam e interpretam também informações não numéricas; sistemas visuais, como os de reconhecimento facial e impressão digital; o processamento de linguagem natural, em que se utiliza da voz para executar comandos; o planejamento e logística, que contribui na automatização e planejamento estratégico de transportes; a construção de robôs autônomos e seus demais usos dentro da robótica.

No âmbito jurídico a inteligência artificial pode beneficiar processos que antes demandariam maior tempo, esforço e orçamento. De acordo com Felipe e Perrota[18], não se trata mais de uma escolha a ser feita, mas sim uma realidade a adoção e direcionamento das ferramentas de IA, uma vez que contribuem como instrumentos de transformação do modus operandi do trabalho jurídico, mas somente mobilizam e realizam a partir da representação de conhecimento, análise e interferências do ser humano jurista.

De acordo com Márquez Díaz[19], diferentes problemas relacionados à análise de dados massivos podem ser solucionados com o uso da IA. Por exemplo, a criação de soluções plausíveis contra a COVID-19, como monitoramento, detecção, diagnóstico e tratamento de doenças associadas a vírus, pela congruência entre tecnologias disruptivas e informações críticas ou sensíveis, possibilitando o desenvolvimento de sistemas de estudo e análise

[17] (GOMES, 2010)
[18] (FELIPE e PERROTA, 2010)
[19] (DÍAZ, 2020)

mais avançados e que facilitem a obtenção de dados relevantes para a tomada de decisões.

Entretanto, a maneira como se dá o uso dos dados pode acarretar em situações preocupantes, diante disso alguns instrumentos são considerados quando se pensa no contexto da utilização de dados. Conforme Ribeiro, a Lei Geral de Proteção de Dados por ser uma legislação que trata sobre o direito à informação, ela possui diversos dispositivos que buscam informar ao titular dos dados o que efetivamente ocorre no tratamento dos seus dados pessoais. Assim, tem-se o direito à transparência, além de instrumentos de proteção ao titular dos dados pessoais quando se trata de decisões automatizadas, nesse caso o direito à explicação e o direito à revisão.

2.1.1. Contextos com uso de dados pessoais

Os dados pessoais obtidos na captação devem levar em consideração conceitos que estejam de acordo com princípios, como na ética dos dados que se concentra em problemas éticos colocados pela recolha e análise de grandes conjuntos de dados, abordando questões relacionadas à criação de perfis, publicidade, re-identificação de indivíduos, privacidade em grupo, discriminação e transparência, entre outras[20].

Tratando-se de sistemas de IA, principalmente os que utilizam tecnologias que recaem sobre o arcabouço da aprendizagem de máquina, ou *machine learning*, a presença de conjuntos de dados é importante para a inferência de correlações e descobrimento de padrões, sendo considerados as experiências de onde o sistema vai aprender[21]. O tamanho do conjunto e os tipos desses dados são fundamentais para a performance do algoritmo, no entanto, quando o objetivo final do sistema é um produto ou serviço prestado a uma pessoa, é recorrente que se faça uso de dados pessoais.

[20] (DONEDA, 2018)
[21] (MITCHELL, 1997)

Dados pessoais podem ser compreendidos como dados relativos a uma pessoa específica de forma que a identifique ou tenha potencial de identificar[22] e atualmente já ocorrem diversas discussões sobre como fazer o uso correto, centrado no ser humano e respeitando a privacidade das pessoas para que esses sistemas realmente venham a melhorar a qualidade de vida das populações por eles servidas. Entre as diversas áreas onde sistemas assim podem ser atuantes, temos a utilização da IA, por exemplo, para facilitar a elaboração de políticas públicas em praticamente todo seu ciclo de vida[23], ter uma análise mais minuciosa e melhorar os sistemas educacionais[24] e auxiliar profissionais de saúde no diagnóstico de doenças[25].

Porém, é preciso reforçar que fazer uso de dados pessoais (muitas vezes, sensíveis[26]) deve levar em conta a privacidade dos indivíduos de onde aqueles dados foram extraídos e considerar todos os riscos atrelados a sua utilização, mais aprofundados na seção 2.2.

2.1.2. Contextos sem uso de dados pessoais

A Inteligência Artificial é um campo de conhecimentos que oferece modelos de apoio à decisão e ao controle com base em conhecimentos empíricos e teóricos, mesmo que apoiados em dados incompletos[27]. As entradas usadas pelo algoritmo, no entanto, não precisam ser necessariamente dados pessoais e vemos, na literatura e mercado, diversos exemplos dessas aplicações.

Dentre seus usos sem a utilização de dados pessoais, encontrados exemplos de sistemas de IA aplicados à agricultura, para auxiliar na produção agrícola através de dados característicos da planta e do solo, predizendo a produtividade das culturas[28]; ao mercado de ações, onde ativos

[22] (LGPD, 2018)
[23] (VERHULST; ENGIN; CROWCROFT, 2019)
[24] (UNESCO, 2019)
[25] (XING; GIGER; MIN, orgs., 2020)
[26] (LGPD, 2018)
[27] (SELLITO, 2002)
[28] (MICHELON, 2016)

comercializados diariamente, negociados em grandes quantidades, como o petróleo tipo brent, necessitam de previsões de preço mais precisas e eficientes[29]; e à área de transporte, onde cada vez mais pesquisas se voltam aos carros autônomos e políticas voltadas a suas capacidades e impactos na sociedade[30].

Entretanto, é preciso tratar essas aplicações com o mesmo rigor que aquelas que fazem uso de dados pessoais. Carros autônomos podem causar acidentes, previsões imprecisas na bolsa de valores podem causar prejuízos econômicos e más decisões na agricultura podem causar dano ao meio ambiente, por exemplo. Apesar dos avanços trazidos pela IA, é fundamental que sua adoção seja feita com estudos de impacto e cautela.

2.1.3. Pesquisa e desenvolvimento de IA a partir de experiências setoriais

Como já pudemos perceber, ferramentas de inteligência artificial podem ser utilizadas nos mais diversos setores para automatizar tanto atividades mais mecanizadas ou repetitivas quanto atividades mais complexas e que requerem uma acurácia e uma precisão maiores do que a média do ser humano para a mesma atividade.

Nesse sentido, o surgimento de lawtechs reforça o crescimento e adoção de programas de computador que permitam a melhoria de eficiência de operadores jurídicos, escritórios de advocacia, tribunais, órgãos legislativos e administrativos, afirma Maranhão[31]. Há também o surgimento de diversos centros de pesquisa para dar suporte ao advento da Inteligência Artificial, relacionados ao Direito em alguns países, como exemplo o Codex, da Universidade de Stanford; o Cirsfid, centro de Informática Jurídica da Universidade de Bologna; o programa de Sistemas Inteligentes, da Universidade de Pittsburgh; o centro de Direito e Tecnologia da Informação

[29] (IGNÁCIO, et al, 2017)
[30] (FAISAL, et al, 2019)
[31] (MARANHÃO, 2017)

do King's College. Maranhão[32] também enfatiza que as principais universidades do mundo estão adotando o ensino de lógica jurídica e lógica de programação, além de criar incubadoras de lawtechs.

Alguns estudos de caso podem ainda contribuir para o direcionamento e desenvolvimento correto da IA, em relação ao seu uso, vantagens e desvantagens, como também possíveis riscos inerentes à sociedade, governos e setores. No exemplo de Porto[33], é possível perceber detalhadamente como a adoção da IA pode impactar o modo como os processos se desenvolvem:

> A experiência pioneira realizada no Estado do Rio de Janeiro comprovou a eficácia do método no executivo fiscal, de modo que sua implantação pode gerar uma cultura de adimplemento dos tributos, com reflexos incomensuráveis para a sociedade como um todo e um impacto extremamente alto para o Judiciário. A solução do executivo fiscal implica numa redução elevada da taxa de congestionamento do Judiciário, sendo possível reduzir a mesma em até 12% (doze por cento) com a movimentação desses processos. Além disso, não se pode desconsiderar o impacto financeiro e orçamentário que essa medida irá ocasionar nos cofres públicos, em benefício de toda a comunidade.

2.2. Riscos

A inteligência artificial está invariavelmente inserida em uma cultura econômica, científica e tecnológica pautada pela inovação e, consequentemente, pela justificação do risco. A partir desse contexto, é usual identificar a associação entre os valores que marcam o campo de produção e consumo das soluções automatizadas – como a eficiência, a velocidade e a otimização –, e a ausência de limites, reflexões e responsabilidade, o que acaba por expressar uma cultura distorcida de inovação e iteração sem balizas de responsabilização e tipificação de riscos altos e inaceitáveis.

---

[32] Idem.
[33] (PORTO, 2019: 192)

A valorização histórica e contextual dos valores e práticas da aceleração não apenas impulsiona essa cultura, mas também catapultam a inteligência artificial para uma posição de crescente utilização e destaque, inclusive na execução de funções estruturantes da atividade social contemporânea, da designação de ofertas de crédito financeiro às vagas em programas sociais de emprego e renda.

A associação entre as funções desempenhadas por esses sistemas e a cultura valorativa que parece caracterizá-los é recorrentemente ilustrada em casos que explicitam os riscos envolvidos na sua utilização. Em conjunto com esses fatores, o acúmulo de evidências quanto aos efeitos excludentes e discriminatórios da operação cotidiana de inteligência artificial inviabiliza qualquer abordagem legislativa que se limite a determinações principiológicas e recomendações indeterminadas, como tem sido definido pelo PL 21/2010.

Pelo contrário, o quadro suscita a necessidade de que a regulação adotada incorpore mecanismos concretos de identificação prévia e combate aos riscos, a partir de um sistema de gradação que prioritariamente reconheça a existência de riscos altos e inaceitáveis, mas também garanta segurança jurídica e condições para a concorrência e a inovação.

São inúmeras as possibilidades de riscos em sistemas de IA. A partir do diálogo com experiências setoriais e internacionais, podemos identificar ao menos grandes categorias de risco: relacionados a dados, técnicos (riscos adversariais em IA e Machine Learning, por exemplo), confiança/ética e compliance.

Não basta, por isso, incorporar terminologias que indiquem respeito à transparência e à ética. Os riscos devem ser incorporados ao corpo da legislação, em um modelo de gradação e de aplicação granular que tutele, de maneira central, impactos sociotécnicos do âmbito dos direitos humanos. Nesse sentido, defendemos a prevalência do princípio da precaução e o entendimento da existência de riscos inaceitáveis, que não podem ser reparados e envolvem a violação de direitos que tem guarida na ordem constitucional e dela derivam, arriscando, primariamente, a vida das pessoas que são submetidas a essas ferramentas.

Um dos riscos inaceitáveis que enfatizamos aqui é o uso de reconhecimento facial na área da segurança pública, que tem se provado uma aplicação que fere liberdades individuais e, principalmente, põe em risco a vida e aumenta a desigualdade social e as vulnerabilidade de grupos que já são historicamente vulnerabilizados. Nesse sentido, defendemos o banimento do reconhecimento facial para a segurança pública.

2.2.1. Potencial discriminatório

Apesar do risco de discriminação não ser inerente aos processos de tomada de decisão algorítmica, uma vez que isso depende de uma série de fatores relacionados ao modo como o modelo de IA é construído e treinado[34], existe um grande potencial de acentuação de desigualdades e ameaça a direitos fundamentais na aplicação de sistemas automatizados.

Nesse contexto, é essencial que haja um compromisso de pensar e agir sobre os possíveis e já existentes impactos na implementação dessa tecnologia. É preciso, de antemão, questionar a suposta objetividade dos algoritmos, e principalmente a forma como esses dispositivos tecnológicos se articulam diante da constatação de problemas, que vão além da mera resolução técnica. Aplicações em setores públicos, por exemplo, envolvem uma rede sociotécnica que inclui agentes humanos e não humanos, num contexto histórico e institucional específico, e por isso é imprescindível que desde a própria concepção da tecnologia sejam contempladas tais considerações.[35]

Há inúmeros casos práticos em que ameaças à direitos fundamentais na aplicação dessa tecnologia estão presentes. Exemplo disso, temos a restrição das oportunidades de emprego a determinadas populações, como é o caso do SINE no Brasil e a potencial superestimação associada a bancos de dados que possuem principalmente informações sobre aqueles que já foram alvo de intervenção de políticas públicas no sistema Alerta Niñez, no Chile, o que pode

[34] (BRUNO, CARDOSO, FALTAY, 2021)
[35] (FUENTES e VENTURINI, 2021)

afetar a posição no ranking de crianças e adolescentes de setores socioeconômicos desfavorecidos e aumentar a estigmatização para a que podem estar sujeitos.

Não obstante, é importante ressaltar também a possibilidade do uso de tais dispositivos tecnológicos ser concentrado nas camadas de maior renda da sociedade, devido aos requisitos de sistema e dispositivo, além do acesso à internet ainda ser desigual. Além disso, é válido considerar a provável discriminação de corpos marginalizados por circunstâncias sociais ao serem analisados e categorizados por um sistema que não possui categorias para classificá-los.[36]

Portanto, é necessário um olhar atento aos aspectos sistêmicos e estruturais, sobretudo aqueles relacionados às múltiplas formas de assimetria e falta de clareza implicadas na adoção massiva de tecnologias como a inteligência artificial, que modifica não apenas a escala, mas a natureza dos problemas decorrentes do uso de tais tecnologias.[37] Dessa forma, é fundamental que os direitos à explicabilidade, transparência e revisão sejam plenamente aplicados no desenvolvimento e aplicação de sistemas automatizados, inclusive exigindo acesso à informação das especificações técnicas do modelo de IA, de forma que possa haver uma avaliação mais precisa dos potenciais discriminatórios do sistema.

3. Direitos e deveres

3.1. Transparência

A transparência é o princípio mais amplamente invocado na literatura que contempla as diretrizes éticas de inteligência artificial - ainda que haja uma significativa variância no que diz respeito à interpretação, justificação, domínio de aplicação e modo de alcançá-la[38]. No Brasil, o principal

[36] (DÍAZ, 2018)
[37] (BRUNO, CARDOSO, FALTAY, 2021)
[38] (JOBIN, et al, 2019; LOI, et al, 2021)

instrumento legal que regulamenta o tratamento de dados pessoais é a LGPD, que define o **princípio da transparência** como a "garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial".

Para que um debate robusto possa acontecer a respeito de qualquer sistema algorítmico, é preciso antes de tudo oferecer às partes interessadas informações relevantes detalhando o que esse sistema faz e como ele opera. Assim, o princípio da transparência pode se tornar um mecanismo útil para monitorar o comportamento de sistemas algorítmicos, fornecendo as pré-condições informacionais necessárias para a *accountability*[39].

No entanto, Arbix[40] reforça a necessidade de ir além de formulações abstratas de códigos, princípios e recomendações, em especial porque, na maior parte dos casos, não há mecanismos de *enforcement* e nem sempre se consegue identificar a natureza real dos problemas. Seria necessário uma articulação entre as ações técnicas e o sistema legal-regulatório, avançando na criação de padrões de precisão e de leis com foco mais apurado.

Há uma gama de fatores que podem moldar a efetividade da transparência de um sistema: tipo, escopo e confiabilidade da informação tornada disponível; quem seriam os destinatários desta informação e como pretendem usá-la; e qual seria a relação entre a entidade a publicar e o destinatário[41].

Mais do que compreender esses sistemas por completo, os formuladores de políticas precisam selecionar quais as frações de informação mais pertinentes a serem disponibilizadas. Nesse sentido, três camadas-chave merecem atenção especial[42]:

> A) Localizar os aspectos relevantes a respeito do envolvimento humano no desenho, operação e

[39] (DIAKOPOULOS, 2020: 197)
[40] (ARBIX, 2020)
[41] (DIAKOPOULOS, 2020:199)
[42] Idem.

gerenciamento de um sistema, facilitando a identificação de indivíduos que possam ser responsabilizados em cada etapa;

B) No que diz respeito aos dados utilizados e à sua qualidade, incluindo: precisão, completude, pontualidade, frequência de atualização e incerteza; representatividade de uma amostra para uma dada população de interesse; informações sobre coleta (incluindo quem a fez, com quais motivações, intenções, com que fundos; se houve consentimento);

C) Detalhes relativos aos modelos usados como metadados, tais como: características, pesos e tipos de modelos; data em que foram criados; versão. Também é possível incorporar heurística, limites, suposições, regras ou restrições que sejam úteis de divulgar, juntamente com qualquer justificativa de design a respeito de por que ou como foram escolhidos. Em alguns casos, transparência no nível do código pode ser necessária.

Por fim, a ideia de buscar por uma "transparência completa", além de dificilmente alcançável, talvez sequer seja pertinente, na medida em que pode entrar em conflito com questões de privacidade ou produzir um volume de informações que não seja viável para compreender. O mais importante é uma política de transparência cuidadosamente construída, contextualmente específica, que produza informações alinhadas às capacidades das partes interessadas em processá-las e capazes de produzir efetiva governança e *accountability* acerca de um sistema[43].

[43] Ibidem.

3.2. Explicabilidade

A explicabilidade é uma dimensão complementar à transparência, uma vez que a primeira consolida o cumprimento da segunda. Um dos pontos de partida possíveis para caracterizá-la é o artigo 13 da GDPR, que prevê ao titular o direito a informações *significativas* sobre a lógica por trás das decisões automatizadas que usem seus dados. Ainda que não use o termo exato, é a um **direito à explicação** que esse artigo se refere, instaurando as questões: Os modelos de inteligência artificial devem ser explicáveis? De que forma isso pode ser implementado? É buscando responder a essas perguntas que se desenvolveu o campo de estudos referido como IA Explicável, ou XAI[44].

A ideia de uma inteligência artificial explicável contrasta com os elementos de opacidade que permeiam um sistema de aprendizagem de máquinas, que frequentemente é referido como uma **caixa-preta.** De acordo com essa ideia, o funcionamento interno de um sistema seria inescrutável até pelos próprios designers, uma vez que se tratam de modelos complexos e baseados em milhões de parâmetros derivados de funções matemáticas[45].

O recurso à ideia de uma inescrutabilidade a esses sistemas, obscurecendo a natureza e a qualidade da informação tornada disponível, não diz respeito somente ao domínio técnico; para Pasquale, há também uma dimensão seletiva, com vistas a evitar mecanismos de regulação, assim como custos adicionais e questões incômodas para as empresas responsáveis pela coleta e processamento[46].

Não há consenso na área sobre a necessidade e a medida dessa explicabilidade. No artigo In *Defense of The Black Box*, Holm[47] faz ressalvas, argumentando que nós rotineiramente aceitamos conclusões humanas sem saber como se originaram; no que diz respeito aos sistemas, seria preciso

[44] (ASGHARI, et al, 2021; GUNNING, 206; MONTEIRO, 2018)
[45] (ALVES e ANDRADE, 2021; DIAKOPOULOS, 2020)
[46] (PASQUALE, 2015).
[47] (HOLM, 2019)

observar se, para alguns cenários específicos, essa opacidade de fato é impeditiva para que bons resultados sejam atingidos e para que tais sistemas sejam úteis.

O pesquisador Nigam Shah endossa que a interpretabilidade nem sempre é necessária para a utilidade, especialmente no campo da medicina, em que os médicos rotineiramente oferecem tratamentos sem saber como ou por quê funcionam. Nesses casos, testes criteriosos acerca dos resultados devem ser suficientes. Por outro lado, Shah reforça que há contextos em que leis e regulamentos deveriam exigir absolutamente uma explicação causal para garantir que as decisões sejam justas. Exemplos disso seriam quando modelos de IA são usados para negar às pessoas entrevistas de emprego, fiança, empréstimos, programas de saúde ou moradia. Assim, ele conclui que os desenvolvedores de modelos devem ser claros sobre por que uma informação é necessária e qual tipo de explicação é útil para determinado contexto[48].

Kate Vredenburgh, por outro lado, é mais incisiva: ela reforça a explicação como um direito, e este deve ser implementado de forma estrutural, a fim de não sobrecarregar os indivíduos com explicações complexas. Uma possibilidade para isso poderia ser através de um representante fiduciário para nossos dados[49]. Ademais, a XAI pode contribuir para a identificação de vieses nos modelos algorítmicos, que poderiam reforçar mecanismos de desigualdade e impactar na qualidade da democracia [50].

É certo que a ciência ainda procura formas mais amigáveis de concretizar o direito à explicação, tornando os sistemas que utilizam inteligência artificial mais próximos dos usuários. Mas é preciso, para fins de regulação geral sobre o tema, que haja um mínimo de garantia de transparência e explicação sobre critérios para tomada de decisão do sistema de IA, a fim de minimizar riscos e resguardar direitos dos cidadãos e cidadãs.

[48] (MILLER, 2021)
[49] (MILLER, 2020)
[50] (O'NEIL, 2021; ALVES e ANDRADE, 2021)

3.3. Revisão

Revisão é mais um aspecto da aplicação de inteligência artificial que tem ligação com a transparência e a explicabilidade. É a partir do direito à revisão que seu titular tem capacidade para requisitar a revisão de uma decisão totalmente automatizada que possa ter um impacto nos seus direitos, principalmente quando se trata de definição do seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade.

Sendo absolutamente interligado ao direito à explicação, ambos são fundamentais para a proteção dos usuários potencialmente afetados pela implementação de sistemas de IA contra possíveis decisões automatizadas de caráter discriminatório e/ou arbitrário. Exemplo disso é que tal preocupação já é uma realidade na vida dos trabalhadores brasileiros que já estão tendo que lidar com a implementação das decisões automatizadas no Sistema Nacional de Emprego, expressando sua preocupação com a falta de clareza e garantias das possíveis consequências que a medida poderia ter em relação à privacidade dos dados dos trabalhadores, bem como à equidade e à igualdade de acesso e oportunidades.[51]

Para cumprir essa função protetiva, a aplicabilidade e efetividade do direito à revisão perpassa, necessariamente, por uma atividade humana. Contrário, um pedido de revisão de decisão automatizada poderia resultar em outra decisão igualmente automatizada, o que prejudica a transparência e a concretização de um direito à explicação consistente, visto que tais direitos implicam que sua concretização seja feita, necessariamente, em linguagem inteligível para o requerente.

Sobre isso, a LGPD, no art. 20, falhou em não garantir que a revisão das decisões tomadas unicamente com base em tratamento automatizado de dados pessoais seja feita por pessoa natural, o que garantiria mais um avanço

[51] (BRUNO, CARDOSO, FALTAY, 2021)

à salvaguardas dos direitos dos titulares de dados pessoais quando são submetidos a ferramentas desta natureza.

Portanto, é fundamental compreender que o direito à revisão deve ser exigido juntamente com a condição de realização a partir de um crivo humano, de forma a garantir que a função de tal norma seja alcançada e possíveis danos aos usuários sejam devidamente observados e verdadeiramente revisados.

3.4. Direito à intervenção humana

O direito à intervenção humana é aspecto recorrentemente associado à confiabilidade e controle de tecnologias de Inteligência artificial. Quando agentes humanos fazem parte da aplicação de um sistema automatizado, eles parecem funcionar como uma barreira contra incorreções e injustiças, garantindo maior qualidade e "humanidade" a partir dos ganhos em capacidade de monitoramento, avaliação e modificação daquele sistema. Por essa razão, a presença humana é um dos elementos que distingue sistemas automatizados de sistemas semi-automatizados e é requisito mínimo em determinados segmentos, como no de carros autônomos.

É esperado, portanto, que ela seja uma das pautas de destaque quando se fala em estímulo regulatório à inteligência artificial benéfica, socialmente responsável e centrada no ser humano. Por isso, chama a atenção o fato de mecanismos de revisão, intervenção e controle humano não terem sido incluídos nos PLs avaliados.

Para além disso, destaca-se que esta inclusão deve levar em conta que os potenciais benefícios do controle humano dependem da compreensão qualificada, detalhada e concreta do que se entende por direito à intervenção humana. Isso porque existem diferentes formas pelas quais uma pessoa pode estar presente em um sistema de IA, e nem todas elas garantem condições reais de agência humana significativa[52]. É essencial que o Marco Normativo de

[52] (WAGNER, 2019; CROOTOF, 2016).

Inteligência Artificial contemple a dimensão do direito à intervenção humana e o faça a partir de uma perspectiva pautada em efetivamente constituir arcabouço legal e material para tal.

Caso essa noção não seja qualificada e desdobrada no texto da lei, especialmente em segmentos de maior risco, o ser humano pode ser inserido na cadeia de aplicação de uma IA apenas como um "carimbador", já que não possui tempo, autonomia e treinamento para exercer uma intervenção significativa. Nos casos em que a presença humana é meramente protocolar, ela pode funcionar como mecanismo de falseamento da tomada de decisão e ser válvula de escape para a responsabilização por problemas e falhas que ocorram nesse processo, especialmente quando há a preocupante intenção de se adotar um regime de responsabilidade subjetiva.

3.5. Correção de vieses

Quando se trata de sistemas técnicos, o **viés** é comumente definido como um problema estatístico: seria um efeito que distorce sistematicamente um fenômeno, tornando-se impreciso. No entanto, se um modelo estatístico pode ser suficiente para resolver um problema no domínio computacional, ele não dá conta da complexidade dos problemas sociais, que possuem uma dimensão contextual incontornável[53]. Assim, uma resolução apenas algorítmica dificilmente englobaria todo o espectro de riscos de enviesamento que abarcam um sistema de IA. Tendo isso em mente, há algumas categorias predominantes de vieses a considerar num processo de correção[54]:

O **viés sistêmico,** também denominado institucional ou histórico, resulta de procedimentos e práticas institucionais. Ele afeta, por exemplo: como organizações e times são estruturados, quem controla os processos de tomada de decisão, assim como heurísticas individuais e de grupo e vieses cognitivos e perceptivos em todo o ciclo da IA. As equipes envolvidas no desenho e desenvolvimento trazem para o processo seus próprios vieses, tanto

[53] (HAO, 2019)
[54] (SCHWARTZ, et al, 2022)

individuais como de grupo, afetando todo o enquadramento de um problema, as suposições sobre quais dados devem ser usados, que modelos devem ser desenvolvidos, onde o sistema deve se situar, entre outros aspectos.

Viés estatístico e computacional - deriva de processos técnicos imperfeitos. Por exemplo: quando a amostra não é representativa da população; quando os algoritmos são treinados em certos tipos de dados e não podem extrapolá-los; quando esses dados são representados matematicamente de forma inadequada, ou tendo como base dados errados ou heterogêneos.

Há ainda o **viés humano:** erros sistemáticos no pensamento humano, que normalmente se dão de forma implícita e inconsciente, portanto dificilmente podem ser controlados ou erradicados. Baseiam-se em princípios heurísticos e percepções sobre as informações necessárias para tomar uma decisão. Assim, são vieses que permeiam os processos de tomada de decisão individuais, grupais e institucionais.

Esse destrinchamento analítico em diferentes dimensões de vieses parte da suposição de que a inteligência artificial não opera num vácuo, e portanto soluções unicamente técnicas podem não ser suficientes para dar conta de todas as variáveis que impactam no ciclo de um sistema. Seria necessária uma análise sociotécnica, articulada à dimensão técnica, para um olhar mais apurado; além disso, a ideia de que um problema pode ser "corrigido" pode ser realista no campo da estatística, mas quando se trata de um problema social, trata-se de um processo contínuo[55].

3.6. Atributos do design técnico: segurança, robustez, resiliência, acurácia e confiabilidade

Durante todo o processo do ciclo de vida de um software, desde seu planejamento e análise de requisitos até sua manutenção e atualização, é desejado que o sistema/produto atinja a mais alta qualidade. No entanto, é

[55] (HAO, 2019; SCHWARTZ, et al, 2022)

preciso definir esse objetivo a partir de atributos e de uma forma mais pragmática. Muitas definições surgem em torno dessa discussão, mas em geral todas convergem para uma capacidade de um software de estar conforme seus requisitos[56]. Essa capacidade de entregar um serviço desejado ou se comportar conforme esperado por todos os envolvidos com aquele sistema também pode ser chamada de **dependabilidade**, da tradução direta do inglês, *dependability*[57].

Para se medir ou ao menos analisar a qualidade de um software, vários outros atributos mais específicos se fazem necessários, e a presença deles em relatórios das equipes de gestão e/ou desenvolvimento é fundamental para uma boa transparência. Entre tais atributos, temos confiabilidade, disponibilidade, resiliência, robustez, segurança de funcionamento (*safety*), segurança (*security*), dentre outros. Todos já bem conhecidos na comunidade de Engenharia de Software, mas que vêm sendo recentemente aderidos e levados em consideração para os sistemas de IA[58]. Em se tratando de IA, além de explicabilidade e enviesamento, é preciso atentar-se para as métricas de erro, como acurácia e precisão.

Confiabilidade (*reliability*) pode ser entendida como uma medida de continuidade de um serviço prestado por um sistema[59] (entende-se aqui "serviço" como o comportamento esperado do software), ou seja, o quanto o software atua sem falhas, dadas as condições (como dados específicos, condições externas ao hardware, etc) e um período de tempo definido[60]. Próxima à confiabilidade, encontra-se a **disponibilidade** (*availability*), que é uma medida sobre a prontidão de um sistema a um acesso, mais especificamente a probabilidade de um sistema estar operacional em um instante específico pré-determinado na avaliação[61].

Resiliência (*resilience*) e **robustez** (*robustness*) também são dois conceitos associados com a ideia de falha. A diferença entre esses dois

[56] (ISO/IEC/IEEE, 24765:2017)
[57] (LAPRIE, 1985)
[58] (PONS & OZKAYA, 2019)
[59] (AVIZIENIS, et al., 2004; LAPRIE, 1985)
[60] (WEBER, 2002)
[61] (AVIZIENIS, et al., 2004; WEBER, 2002; LAPRIE, 1985)

atributos é sutil e suas definições são correlacionadas; no entanto, resiliência está associada à capacidade do sistema de, dada uma condição (defeito, erro ou falha), o software conseguir se recuperar rápida e efetivamente, protegendo seus elementos críticos (como dados sensíveis ou funcionalidades mais importantes), enquanto a robustez é descrita como uma capacidade voltada a tolerar essas condições não desejadas, de forma a tratá-las para não interferir na funcionalidade do sistema[62].

Segurança pode ser entendida como dois atributos diferentes, sendo eles *safety* e *security*[63]. A diferença entre os dois termos é que *safety* está relacionada à segurança de funcionamento, na qual o sistema é capaz de realizar contramedidas para uma situação não esperada de forma que isso não cause nenhum dano aos usuários ou a outros sistemas vinculados a ele[64]. *Security*, por outro lado, é a segurança propriamente dita do sistema, a capacidade de proteger-se contra falhas maliciosas. Por conta disso, essa última definição de segurança se baseia em três propriedades: **confidencialidade**, **integridade** e **disponibilidade**[65] (a última, já descrita anteriormente). Confidencialidade é a propriedade referente ao acesso ao sistema, em que apenas um usuário autorizado pode acessar o sistema, enquanto integridade é a propriedade que existe quando apenas um usuário autorizado pode manipular dados no sistema[66]. O tópico de segurança tem ficado cada vez mais em voga no campo da Inteligência Artificial (sobretudo em Aprendizagem de Máquina) pela descoberta de várias vulnerabilidades que esses modelos têm em relação a ataques que manipulam os dados de entrada, chamados **ataques adversariais**[67].

Apesar de todas essas medidas serem discutidas há bastante tempo dentro da área de engenharia de software, apenas nos anos recentes que essa discussão chegou ao campo da inteligência artificial, sobretudo por conta do rápido crescimento que a última vem sofrendo. Já existem trabalhos sobre a

[62] (FIRESMITH, 2019a; FIRESMITH, 2019b; VALENTE, 2022)
[63] (MAZIERO, 2019)
[64] (WEBER, 2002)
[65] (MAZIERO, 2019; AMOROSO, 1994)
[66] (MAZIERO, 2019; WEBER, 2002)
[67] (SZEGEDY, et al., 2014; ILYAS, et al., 2018; GOLDWASSER, et al., 2022)

aplicação desses atributos em softwares baseados em IA, bem com sua regulação[68]. Além de todos esses atributos, medidas de erro observacional, que quantificam o quão diferente um valor observado (como uma saída de um modelo de IA) é do que era esperado. Nesse cenário, medidas como **acurácia**, **precisão**, **revocação**, entre outras, são de suma importância na análise prévia de adoção ou não de uma tecnologia desta natureza.

3.7. Segredos comercial e industrial

Uma das ressalvas ao princípio da transparência, como definido na LGPD, é diante da necessidade de resguardar um segredo comercial e industrial. A premissa, nesse caso, é a de que publicar informações detalhadas acerca de como um sistema funciona pode minar as vantagens técnicas de uma empresa no mercado, facilitando aos competidores que a imitem[69].

O argumento que serve de base para esse tipo de preocupação é o de propriedade, podendo ser implementado através de uma patente ou sigilo comercial. Embora o uso de patentes esteja aumentando - especialmente pelo Google - o sigilo traz consigo a vantagem de não precisar acompanhar a rapidez que é característica ao aprendizado de máquina[70].

Embora essa pareça ser uma preocupação restrita ao setor privado, é preciso ter em mente que muitos serviços públicos utilizam softwares comprados no mercado ou terceirizam todo o processo para empresas[71]. No Brasil, esse tipo de parceria entre os setores já é amplamente adotada nos nas áreas de emprego[72], educação, transporte, controle de fronteiras e segurança pública[73], tendo assim um vasto alcance. Assim, esse sigilo pode ser limitante para o exercício de alguns direitos fundamentais, dificultando a averiguação

[68] (PONS & OZKAYA, 2019; FDA, 2019)
[69] (DIAKOPOULOS, 2020:210)
[70] (DE LAAT, 2018)
[71] Idem.
[72] (BRUNO; CARDOSO; FALTAY, 2020)
[73] (IGARAPÉ, 2019)

de que não há discriminação, estigmatização e manipulação incorporados ao sistema[74].

O que é importante salientar aqui é que transparência não é questão de tudo ou nada: há várias possibilidades entre a garantia de informações úteis para o interesse público e a prestação de contas, por um lado, e o respeito aos direitos de propriedade industrial e os relativos a segredos comerciais, por outro. Há autores que argumentam que nem sempre transparência completa é necessária - é possível que os sistemas sejam disponibilizados para uma revisão fechada a intermediários, como órgãos de fiscalização ou destinatários específicos que sejam legalmente vinculados e em posição de autoridade para acessar o sistema[75]. Esse pode ser um arranjo benéfico para a regulação brasileira que ora se avizinha.

4. Accountability, governança e fiscalização

4.1. Regimes de responsabilidade civil

Diversos são os temas afetados pelo conjunto do debate da responsabilidade civil tecnológica. A inteligência artificial, fazendo parte deste grupo, é atravessada por considerações já enfrentadas em outros produtos tecnológicos, como a Internet. O tema da responsabilidade civil é fonte frequente de equívocos conceituais, atecnia jurídica e desconsideração de caminhos que escapem à responsabilidade subjetiva ou objetiva. A discussão relaciona-se às circunstâncias, características e agentes aos quais se determina o dever de comprovar culpa e o ônus de responder pelos danos existentes, o que confere importância ímpar ao tema no âmbito de regulação de novas tecnologias.

Em uma legislação que se pretende específica, direcionada e atenta às demandas decorrentes da presença de inteligência artificial nas mais diferentes atividades humanas, é necessário que o tema receba tratamento

[74] (MALONE, 2020)
[75] (DE LAAT, 2018; DIAKOPOULOS, 2020)

adequado. Como a inteligência artificial pode ser avaliada a partir de outras legislações, ao ser classificada como um produto no Código de Defesa do Consumidor, por exemplo, prescrições errôneas e específicas no Marco Normativo de IA podem acabar por enfraquecer diretrizes já existentes.

Por isso, o tema da responsabilidade civil chama atenção na forma como está posta no PL 21/2020. Como ocorre em outras áreas relacionadas à inovação industrial e tecnológica, o "princípio da responsabilidade pela culpa", base da responsabilidade subjetiva atualmente prevista, é incapaz de tutelar corretamente lesões referentes ao terreno das aplicações abarcadas pelo guarda-chuva da inteligência artificial.

Esse regime remete ao consumidor (aquele que imputa o fato) o ônus de comprovar os danos sofridos e estabelece uma visão limitada quanto à responsabilidade tecnológica dos envolvidos no ciclo de vida do produto. É anacrônico, portanto, em relação à compreensão de responsabilidade civil nesse âmbito e expressa baixo comprometimento com os riscos acarretados pela inteligência artificial. Ao retirar a imposição objetal, faz com que eventuais ofensores do campo do desenvolvimento em produtos de IA possam causar dano sem serem devidamente responsabilizados.

Por sua vez, a responsabilidade objetiva tampouco atende idealmente às necessidades do objeto a ser regulado, já que ela é definida a partir de uma pressuposição absoluta que coloca o ônus da prova do lado do objeto, determinando com isso um dado jurídico que independe de exceção ou defesa, quando estas não estão estabelecidas em lei. Essa consideração não deve ser ignorada pela análise proposta pela Comissão e demonstra que o modelo dicotômico é insuficiente para abordar o tema.

Em proposta alternativa, o IP.rec tem sugerido modelos intercalares, especificamente no que se refere a uma responsabilidade civil transubjetiva. Estabelecida por Pontes de Miranda e pouco compreendido pelo pela doutrina, trata-se de um modelo que permite a devida incorporação de um conjunto de responsabilidades intermediárias e mediadas, como as que dizem respeito ao risco que é assumido quando há má escolha ou má vigilância. Dessa forma, ela reconhece tanto a possibilidade de relações causais que se

dão de forma indireta e extrapolam o campo do sujeito para se fundamentarem em atributos inerentes ao objeto, quanto também admite abertura para a contrariedade. Consideramos que o modelo reúne as condições necessárias para garantir a centralidade dos direitos fundamentais na utilização de aplicações de inteligência artificial e para preservar um ambiente propício à inovação.

4.2. Auditoria

A realização de auditorias em inteligência artificial se circunscreve dentro das demandas por fiscalização, ou, na literatura internacional, *accountability*. No contexto brasileiro, a LGPD prevê, em seu artigo 20, § 2º, a possibilidade de "realização de auditoria para verificação de aspectos discriminatórios em tratamento automatizado de dados pessoais". Tanto o pedido de realização de auditorias como de relatórios de impacto ligados à proteção de dados figuram dentre as atribuições do órgão federal da ANPD.

Na literatura específica da área, ainda que referências como Cathy O'Neil[76] e Frank Pasquale[77] venham defendendo auditorias em Inteligência Artificial há anos, o debate está longe de um consenso. Isso porque é preciso considerar uma série de dimensões como: com que periodicidade, através de que instituições e quais indicadores seriam considerados. Por motivos como esses, o termo "auditoria de IA" pode ter muitos significados; sua capacidade em efetivamente detectar e proteger contra vieses não é garantida por si só[78].

Assim, as auditorias mais rigorosas podem ter o escopo limitado; mesmo com acesso irrestrito aos algoritmos, pode ser difícil chegar a uma análise precisa; no que concerne ao setor privado, tem crescido um verdadeiro mercado de regulação[79], com diversas instituições e profissionais que oferecem esses serviços. O principal problema relacionado à contratação

[76] A cientista de dados, aliás, possui uma empresa especializada em auditorias, a O'Neil Risk Consulting and Algorithmic Auditing. A auditoria, em seu livro, aparece como resposta aos vieses algorítmicos (O'NEIL, 2021).
[77] (PASQUALE, 2017)
[78] (SCHELLMANN, 2021)
[79] (SILVEIRA, 2020)

voluntária de um auditor privado é que há o risco de conflitos de interesse, afinal, o auditor pode ser influenciado pelo fato de se tratar de um cliente. Por isso, especialistas em responsabilidade têm pressionado por uma regulamentação mais ampla, assim como de padrões e diretrizes para a auditação[80].

Pensando em termos mais propositivos, Sandvig, junto a alguns autores[81], propõem uma metodologia inspirada nos Estudos de Auditoria para delinear cinco abordagens possíveis para a auditoria de sistemas algorítmicos. Enquanto algumas se assemelham a métodos clássicos da ciência social, outras focam na auditagem de código e na raspagem de dados, havendo ainda a possibilidade de mesclar técnicas. Outros autores[82] reforçam a importância em ter mais clareza sobre a natureza do direito dos auditores e objetivos da auditoria, com o intuito de desenvolver padrões eticamente significativos com relação aos quais diferentes formas de auditoria podem ser avaliadas e comparadas.

4.3. Arranjos institucionais de fiscalização

Um ponto que tem relação íntima com a questão do arranjo institucional de fiscalização é a própria forma de construção e estruturação da lei a ser elaborada. Fiscalização pressupõe obrigações e sanções para o seu descumprimento. Sem isso, qualquer atividade fiscalizatória sequer existe, afinal, não há fiscalização possível se a regulação não prevê obrigações a serem cumpridas.

Então, tudo depende da opção legislativa adotada. Se a opção for por um padrão principiológico de regulação, apenas com diretrizes gerais, mas sem qualquer obrigação específica para os entes envolvidos e, também, sem a atribuição de sanções para o descumprimento, qualquer regime fiscalizatório perde a razão de existir. Tendo em vista o texto atual do PL 21/20, cuja opção legislativa foi exatamente essa, principiológica, sem atribuição de obrigações,

[80] (SCHELLMANN, 2021)
[81] (SANDVIG, et al, 2014; DA SILVEIRA e DA SILVA, 2020)
[82] (LOI e SPIELKAMP, 2021)

não há que se falar em arranjos fiscalizatórios, considerando que a própria lei não atribui o que fiscalizar. Um possível órgão regulador, nesse caso, nascerá esvaziado de atribuições.

Porém, caso a opção legislativa seja pelo modelo oposto, com definição de competências, obrigações específicas, boas práticas metrificadas e as respectivas sanções para o seu descumprimento, os arranjos fiscalizatórios passam a ser não só desejáveis como uma imposição legal, merecendo uma boa reflexão por parte do legislador.

Nesse sentido, é mister ressaltar que a discussão sobre autoridades específicas de fiscalização vem sendo posta em todas as tentativas de regulação de novas tecnologias desde a lei geral de proteção de dados pessoais e a criação da Autoridade Nacional de Proteção de Dados (ANPD), como por exemplo no PL 2630/2020, que pretende definir regras para a regulação da atividade das plataformas de redes sociais, mecanismos de busca e aplicativos de mensagem privada, mas é preciso se questionar se esse movimento de ampliação da máquina pública se mostra eficaz, através da criação de sucessivas autoridades fiscalizatórias para temas diversos que abrangem a adoção de novas tecnologias e seu amplo uso pelos vários setores da sociedade.

Entretanto, com o fim de evitar a multiplicação de autoridades administrativas, entendemos que o Judiciário pode e deve dar conta das demandas advindas do descumprimento legal, seja qual for a matéria tratada. Para que isso seja efetivo, é preciso, novamente, que as obrigações estejam claras e, preferencialmente, que haja sanções aplicáveis ao descumprimento. Mas também, como já dito anteriormente, o marco normativo deve servir para suprir as lacunas e insuficiências de outros regramentos já consolidados e que podem servir de fonte normativa em determinados casos, como o Código de Defesa do Consumidor ou a própria Lei Geral de Proteção de Dados. No mesmo sentido, o regime de responsabilidade civil definido para os casos de dano decorrente do uso de ferramentas de inteligência artificial deve possibilitar ao magistrado a atribuição das devidas consequências, o que não

ocorrerá caso a opção seja pela responsabilidade subjetiva, como se encontra definido no texto do PL 21/20 que se encontra em discussão.

Especificamente neste caso, dentre as hipóteses já levantadas em diversas instâncias de debates, entendemos que, em governos anteriores, houve experiências positivas de conselhos setoriais que tratavam de matérias específicas; assim, talvez o melhor modelo legislativo para o caso da inteligência artificial seja, para além de uma autoridade reguladora ou coordenadora, que por questões orçamentárias, por exemplo, pode ser difícil de ser implementada, a existência de um conselho multissetorial e interdisciplinar de profissionais com notório saber sobre as diversas disciplinas que envolvem a IA. Dentro deste Conselho, entendemos que é preciso que haja paridade nos critérios de representação setorial, entre governo, empresas, sociedade civil e acadêmicos. Além disso, deve ser uma estrutura que tenha segurança legal e que, onde quer que ela esteja inserida, caiba a ela dar concreção às decisões tomadas sobre o tema.

Assim, o IP.rec enxerga que esse arranjo terá mais benefícios em geral para a sociedade e proporcionará também uma maior participação dos diversos setores interessados, como determina a tradição multissetorial.

5. Referências

ALVES, Marcos Antônio Souza; ANDRADE, Otavio Morato de. Da "caixa-preta" à "caixa de vidro": o uso da Explainable Artificial Intelligence (XAI) para reduzir a opacidade e enfrentar o enviesamento em Modelos Algorítmicos. **RDP**, Brasília, v. 18, n. 100, 349-373, out-dez 2021.

AMOROSO, Edward. **Fundamentals of computer security technology**. Englewood Cliffs, N.J: PTR Prentice Hall, 1994.

ARBIX, Glauco. A transparência no centro de uma IA ética. **Novos Estudos CEBRAP**, v. 39, p. 395-413, 2020.

ASGHARI, H.; BIRNER, N.; BURCHARDT, A.; DICKS, D.; FAßBENDER, J.; FELDHUS, N.; HEWETT, F.; HOFMANN, V.; KETTEMANN, MATTHIAS C.; SCHULZ, W.; SIMON, Judith; STOLBERG-LARSEN, J.; and ZÜGER, T. (2021). What to explain when explaining is difficult? An interdisciplinary primer on XAI and meaningful information in automated decision-making. **Alexander von Humboldt Institute for Internet and Society**. Disponível em: https://doi.org/10.5281/zenodo.6375784. Acesso em: 09 de maio de 2022.

AVIZIENIS, Algirdas; LAPRIE, Jean-Claude; RANDELL, Brian; LANDWEHR, Carl. Basic concepts and taxonomy of dependable and secure computing. **IEEE Transactions on Dependable and Secure Computing**, v. 1, n. 1, p. 11–33, 2004. Disponível em: http://ieeexplore.ieee.org/document/1335465/. Acesso em: 11 de maio de 2022.

BRASIL, Lei 13.709, de 14 de agosto de 2018, Art. 5. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm. Acesso em: 8 de jun. 2022.

BRUNO, Fernanda; CARDOSO, Paula; FALTAY, Paulo. Sistema Nacional de Emprego e a gestão automatizada do desemprego. **Derechos Digitales** [online]. Disponível em: https://ia.derechosdigitales.org/wp-content/uploads/2021/04/CPC_informe_BRASIL.pdf. Acesso em 11 de maio de 2022.

BRYSON, Joanna J., Europe Is in Danger of Using the Wrong Definition of AI. **Wired**. 2 de março de 2022. Disponível em: https://www.wired.com/story/artificial-intelligence-regulation-european-union/. Acesso em: 6 maio 2022

CANALES, Maria Paz. What do we talk about when we talk about AI? Algorithmic decision-making in Latin America, *in*: Garat, Vladimir (ed). **Latin America in a glimpse**. Derechos Digitales, p. 3–10, 2020

CROOTOF, Rebecca, A meaningful floor for meaningful human control, **Temp. Int'l & Comp**. LJ, v. 30, p. 53, 2016. Disponível em: https://sites.temple.edu/ticlj/files/2017/02/30.1.Crootof-TICLJ.pdf. Acesso em: 10 maio 2022

CUNHA, W. S. Estudo da Inteligência Artificial aplicada em Internet das Coisas, voltada na Automação Residencial. **Revista Científica Semana Acadêmica**. Fortaleza, ano MMXVIII, n. 000121, 2018.

DA SILVEIRA, Sergio Amadeu. DA SILVA, Tarcizio Roberto. Controvérsias sobre dano algorítmico: discursos corporativos sobre discriminação codificada. **Revista Observatório**. V. 6, n. 4, Julho-Setembro de 2020.

DE LAAT, Paul B. Algorithmic decision-making based on machine learning from big data: can transparency restore accountability?. **Philosophy & technology**, v. 31, n. 4, p. 525-541, 2018.

DIAKOPOULOS, Nicholas. Transparency. In DUBBER; PASQUALE; DAS (org). **The Oxford Handbook of Ethics of AI**. New York: Oxford University Press, 2020.

DÍAZ, Jairo Eduardo Márquez. Inteligencia artificial y Big Data como soluciones frente a la COVID-19. **Revista de bioética y derecho**, n. 50, p. 315-331, 2020.

DÍAZ, Marianne. El Cuerpo como Dato. **Derechos Digitales** [online]. 2018. Disponível em: https://www.derechosdigitales.org/wp-content/uploads/cuerpo_DATO.pdf. Acesso em: 09 jun. 2022.

DONEDA, Danilo Cesar Maganhoto et al. Considerações iniciais sobre inteligência artificial, ética e autonomia pessoal. **Pensar-Revista de Ciências Jurídicas**, v. 23, n. 4, p. 1-17, 2018.

FAISAL, Asif; YIGITCANLAR, Tan; KAMRUZZAMAN, Md.; CURRIE, Graham. Understanding autonomous vehicles: A systematic literature review on capability, impact, planning and policy. **Journal of Transport and Land Use**, v. 12, n. 1, 2019, Disponível em: https://www.jtlu.org/index.php/jtlu/article/view/1405. Acesso em: 9 jun. 2022.

FDA - U.S. Food & Drug Administration. **Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)**. US FDA Artificial Intelligence and Machine Learning Discussion Paper. Disponível em: https://www.fda.gov/files/medical%20devices/published/US-FDA-Artificial-Intelligence-and-Machine-Learning-Discussion-Paper.pdf. Acesso em: 11 de maio de 2022.

FELIPE, Bruno Farage da Costa; PERROTA, Raquel Pinto Coelho. Inteligência artificial no Direito: uma realidade a ser desbravada. 2018.

FIRESMITH, Donald. **System Resilience: What Exactly is it?** SEI Blog. Disponível em: https://insights.sei.cmu.edu/blog/system-resilience-what-exactly-is-it/. Acesso em: 10 de maio de 2022.

FIRESMITH, Donald. **System Resilience Part 2: How System Resilience Relates to Other Quality Attributes**. SEI Blog. Disponível em: https://insights.sei.cmu.edu/blog/system-resilience-part-2-how-system-resilience-relates-to-other-quality-attributes/. Acesso em: 10 de maio de 2022.

FUENTES, Patricio Veloso; VENTURINI, Jamila. **Decisiones automatizadas en la función pública en América Latina - Brasil, Chile, Colombia y Uruguay**. Derechos Digitales, 2021.

GERCINA, Cristiane. Análise automática de benefícios do INSS por robô falha, diz sindicato. **Folha de São Paulo**. 4 de maio de 2022. Disponível em: https://www1.folha.uol.com.br/mercado/2022/05/inss-usa-robos-para-tentar-reduzir-fila-de-beneficios-diz-sindicato.shtml. Acesso em: 7 maio 2022.

GOLDWASSER, Shafi; KIM, Michael Pum-Shin; VAIKUNTANATHAN, Vinod; ZAMIR, Or. **Planting Undetectable Backdoors in Machine Learning Models.** 2022. Disponível em: http://arxiv.org/abs/2204.06974. Acesso em: 11 de maio de 2022.

GOMES, D. dos S. Inteligência Artificial: conceitos e aplicações. Olhar Científico. v1, n. 2, p. 234-246, 2010.

GUNNING, David. Explainable Artificial Intelligence (XAI). **Darpa**, Defense Advanced Research Projects Agency, 2016.

HAO, Karen. This is How AI Bias Really Happen - And Why it's so hard to fix. **MIT Technology Review**. Publicado em 4 de fevereiro de 2019. Acesso em 09 de abril de 2022. Disponível em: https://www.technologyreview.com/2019/02/04/137602/this-is-how-ai-bias-really-happensand-why-its-so-hard-to-fix/

HOLM, Elisabeth A. In defense of the black box: Black Box algorithms can be useful in science and engineering. **Science Magazine**. Vol. 364, issue 6435, 5 de abril de 2019. Disponível em: https://www.science.org/doi/10.1126/science.aax0162. Acesso em: 09 de maio de 2022.

IGNÁCIO, Lucas Vinicio Ribeiro et al. O uso de inteligência artificial para a previsão do preço do petróleo. **Revista ESPACIOS**, vol. 38, nº 24, 2017.

INSTITUTO IGARAPÉ. Desde 2011 vem sendo utilizado o reconhecimento facial no Brasil. 2019. **Instituto Igarapé** [online]. Disponível em: https://bit.ly/2L89rvh. Acesso em: 11 de maio de 2022.

ILYAS, Andrew; ENGSTROM, Logan; ATHALYE, Anish; LIN, Jessy. **Black-box Adversarial Attacks with Limited Queries and Information.** 2018. Disponível em: http://arxiv.org/abs/1804.08598. Acesso em: 11 de maio de 2022.

ISO/IEC JTC 1/SC 7 TECHNICAL COMMITTEE. **ISO/IEC/IEEE 24765:2017**. ISO. Disponível em: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/19/71952.html. Acesso em: 11 de maio de 2022.

JOBIN, Anna; IENCA, Marcello; Effy, Vayena. Artificial Intelligence: the global landscape of Ethics Guidelines. **Nature Machine Intelligence**, nº 9 (September 2019): 389–99, Disponível em: https://doi.org/10.1038/s42256-019-0088-2. Acesso em: 08 de maio de 2022.

LAPRIE, Jean-Claude. DEPENDABLE COMPUTING AND FAULT TOLERANCE : CONCEPTS AND TERMINOLOGY. *In*: **Fifteenth International Symposium on Fault-Tolerant Computing, 1985.** Ann Arbor, MI: IEEE, 1985, p. 2-11. Disponível em: http://www.macedo.ufba.br/conceptsandtermonology.pdf. Acesso em: 11 de maio de 2022.

LOI, Michele; MATZENER, Anna; MULLER, Angela; SPIELKAMP, Matthias. Automated Decision-Making Systems in the Public Sector: An Impact Assessment Tool for Public Authorities. **Algorithm Watch**, 2021.

MALONE, Matt. Trade Secrets, Big Data and The Future of Public Interest Litigation Over Artificial Intelligence In Canada. **Canadian Intellectual Property Review** V. 35, 2020. Disponível em: https://ssrn.com/abstract=3783514, Acesso em 11 de maio de 2022.

MARANHÃO, Juliano. A pesquisa em inteligência artificial e Direito no Brasil. **Conjur**, 2017.

MAZIERO, Carlos. Capítulo 26 - Conceitos básicos de segurança. *In*: **Sistemas Operacionais: Conceitos e Mecanismos**. [s.l.]: Editora da Universidade Federal do Paraná - UFPR, 2020, p. 336–348. Disponível em: https://www.researchgate.net/publication/343921399_Sistemas_Operacionais_Conceitos_e_Mecanismos. Acesso em: 11 de maio de 2022.

MICHELON, Gabriela Karoline. **Aplicação de técnicas de inteligência artificial na agricultura de precisão para estimar a produtividade da soja**. 99 f. Trabalho de Conclusão de Curso (Graduação) - Universidade Tecnológica Federal do Paraná, Medianeira, 2016.

MILLER, Katharine. HAI Fellow Kate Vredenburgh: The Right to an Explanation. **Standford University Human-Centered Artificial Intelligence** [online]. Publicado em 24 de junho de 2020. Acesso em 09 de maio de 2022. Disponível em: https://hai.stanford.edu/news/hai-fellow-kate-vredenburgh-right-explanation

MILLER, Katharine. Shold AI Models be explainable? That depends. **Standford University Human-Centered Artificial Intelligence** [online]. Publicado em 16 de março de 2021. Acesso em 09 de maio de 2022. Disponível em: https://hai.stanford.edu/news/should-ai-models-be-explainable-depends

MITCHELL, Tom M. **Machine Learning**. New York: McGraw-Hill, 1997.

MONTEIRO, Renato. Existe um direito à explicação na Lei Geral de Proteção de Dados do Brasil? **Instituto Igarapé**, Artigo Estratégico 39, Dezembro de 2018.

NADIMPALLI, Meenakshi. Artificial intelligence risks and benefits. **International Journal of Innovative Research in Science, Engineering and Technology**, v. 6, n. 6, 2017.

O'NEIL, Cathy. **Algoritmos de destruição em massa**. Editora Rua do Sabão, 2021.

OECD, **Recommendation of the Council on Artificial Intelligence**. OECD/LEGAL/0449, 2019. Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449. Acesso em: 5 maio 2022

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

PASQUALE, Frank. A esfera pública automatizada. **Líbero**, v. 20 – nº 39 jan./ago. p.16-35, 2017. Disponível em: http://seer.casperlibero.edu.br/index.php/libero/article/view/866. Acesso em: 08 jun. 2022.

PEET, Charlotte. Brazil's embrace of facial recognition worries Black communities – Rest of World. **O Panóptico**. Disponível em: https://opanoptico.com.br/brazils-embrace-of-facial-recognition-worries-black-communities-rest-of-world/. Acesso em: 7 maio 2022.

PONS, Lena; OZKAYA, Ipek. **Priority Quality Attributes for Engineering AI-enabled Systems.** 2019. Disponível em: https://arxiv.org/abs/1911.02912. Acesso em: 11 de maio de 2022.

PORTO, Fábio Ribeiro. O impacto da utilização da Inteligência Artificial no Executivo Fiscal. Estudo de Caso do Tribunal de Justiça do Rio de Janeiro. **Direito em Movimento**, v. 17, n. 1, p. 142-199, 2019.

RIBEIRO, Ana Lídia Lira. **Discriminação em algoritmos de inteligência artificial:** uma análise acerca da LGPD como instrumento normativo mitigador de vieses discriminatórios [monografia]. 61 f. Monografia (Graduação em Direito) - Faculdade de Direito, Universidade Federal do Ceará, Fortaleza, 2021.

SANDVIG, Christian; HAMILTON, Kevin; KARAHALIOS, Karrie; LANGBORT, Cedric. Auditing Algorithms: Research methods for detecting discriminationon internet platforms. **Data and discrimination**: converting critical concerns into productive inquiry, v. 22, 2014.

SCHUETT, Jonas. Defining the scope of AI regulations . **Legal Priorities Project Working Paper Series.** No. 9, August 22, 2021. Disponível em: http://dx.doi.org/10.2139/ssrn.3453632. Acesso em: 6 maio 2022.

SCHELLMANN, Hilke. Auditors are testing hiring algorithms for bias, but there's no easy fix. **MIT Technology Review** [online]. Publicado em 11 de fevereiro de 2021. Acesso em 08 de junho de 2022. Disponível em: https://www.technologyreview.com/2021/02/11/1017955/auditors-testing-ai-hiring-algorithms-bias-big-questions-remain/

SCHWARTZ, Reva; VASSILEV, Apostol; GREENE, Kristen; PERINE, Lori; BURT, Andrew; HALL, Patrick. Towards a standard for identifying and managing bias in Artificial Intelligence. **National Institute of Standards and Technology**: U.S. Department of Commerce. Março de 2022. Acesso em 10 de maio de 2022. Disponível em: https://doi.org/10.6028/NIST.SP.1270

SELLITTO, Miguel Afonso. Inteligência artificial: uma aplicação em uma indústria de processo contínuo. **Gestão & Produção**, v. 9, n. 3, p. 363-376, 2002.

SILVEIRA, Sergio Amadeu da. Discursos sobre regulação e governança algorítmica. **Revista Estudos Sociológicos**. Araraquara, v. 25, n. 48, p. 63-85, jan-jun 2020.

SLOANE, Mona; MOSS, Emanuel, AI's social sciences deficit, **Nature Machine Intelligence**, v. 1, n. 8, p. 330–331, 2019.

SZEGEDY, Christian; ZAREMBA, Wojciech; SUTSKEVER, Ilya; BRUNA, Joan; ERHAN, Dumitru; GOODFELLOW, Ian; FERGUS, Rob. **Intriguing properties of neural networks.** 2014. Disponível em: http://arxiv.org/abs/1312.6199. Acesso em: 10 de maio de 2022.

UNESCO. **Beijing Consensus on Artificial Intelligence and Education**, 2019, disponível em:

https://unesdoc.unesco.org/ark:/48223/pf0000368303. Acesso em: 8 de junho de 2022.

UNESCO. **Recommendation on the Ethics of Artificial Intelligence - UNESCO Digital Library**, disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000380455. Acesso em: 5 maio 2022

VALENTE, Marco Tulio. **Engenharia de Software Moderna**. 1. ed. [s.l.]: Independente, 2022. Disponível em: https://engsoftmoderna.info/. Acesso em: 11 de maio de 2022.

VERHULST, Stefaan G.; ENGIN, Zeynep; CROWCROFT, Jon. Data & Policy: A new venue to study and explore policy–data interaction. **Data & Policy**, Cambridge University Press, v. 1, n. 1, 10 jun. 2019. DOI https://doi.org/10.1017/dap.2019.2. Disponível em: https://www.cambridge.org/core/journals/data-and-policy/article/data-policy-a-new-venue-to-study-and-explore-policydata-interaction/11718D04E3BA94C7EB87891ACF96D519. Acesso em: 8 jun. 2022.

WAGNER, Ben, Liable, but not in control? Ensuring meaningful human agency in automated decision-making systems, **Policy & Internet**, v. 11, n. 1, p. 104–122, 2019. Disponível em: https://onlinelibrary.wiley.com/doi/full/10.1002/poi3.198. Acesso em: 10 maio 2022

WANG, Pei. On Defining Artificial Intelligence. **Journal of Artificial General Intelligence**, v. 10, n. 2, p. 1–37, 2018. Disponível em: https://www.sciendo.com/article/10.2478/jagi-2019-0002. Acesso: 09 mai. 2022

WEBER, Taisy Silva. Um roteiro para exploração dos conceitos básicos de tolerância a falhas. 2022. Disponível em: https://www.inf.ufrgs.br/~taisy/disciplinas/textos/Dependabilidade.pdf. Acesso em: 10 de maio de 2022.

XING, Lei; GIGER, Maryellen L.; MIN, James K. (Orgs.). **Artificial intelligence in medicine: technical basis and clinical applications**. London, United Kingdom ; San Diego, CA: Academic Press, an imprint of Elsevier, 2020.

ZIEMIANIN, Karolina, Civil legal personality of artificial intelligence: Future or utopia?, **Internet Policy Review**, v. 10, n. 2, p. 1–22, 2021.