



**ESTRATÉGIA BRASILEIRA DE INTELIGÊNCIA ARTIFICIAL (EBIA) E
POLÍTICAS PÚBLICAS: PROPOSTAS PARA EFETIVAÇÃO DOS EIXOS
LEGISLAÇÃO, REGULAÇÃO E USO ÉTICO E GOVERNANÇA DE IA**

**BRAZILIAN STRATEGY FOR ARTIFICIAL INTELLIGENCE (BSIA) AND PUBLIC
POLICIES: PROPOSALS FOR THE IMPLEMENTATION OF THE AXES
LEGISLATION, REGULATION, AND ETHICAL USE, AND GOVERNANCE OF AI**

**LA ESTRATEGIA BRASILEÑA DE INTELIGENCIA ARTIFICIAL (EBIA) Y LAS
POLÍTICAS PÚBLICAS: PROPUESTAS PARA LA IMPLEMENTACIÓN DE LOS
EJES LEGISLACIÓN, REGULACIÓN Y USO ÉTICO Y GOBERNANZA DE LA IA**

Sthéfano Bruno Santos Divino¹

Resumo: Este artigo tem como problema de pesquisa o seguinte questionamento: como e quais políticas públicas podem ser adequadas para efetivação e implementação da Estratégia Brasileira de Inteligência Artificial (EBIA) quanto aos eixos Legislação, regulação e uso ético e Governança de Inteligência Artificial (IA)? Objetiva-se apresentar a referida estratégia e seus eixos de implementação para, após, contextualizá-la conforme classificação do Brasil em matéria de Inteligência Artificial no cenário Internacional. Utiliza-se, para tanto, o Índice Global de Inovação elaborado pela Organização Mundial de Propriedade Intelectual e o *Artificial Intelligence Index Report* elaborado pela Universidade de Stanford. A metodologia utilizada é a de pesquisa monográfica. Ao final, conclui-se que a EBIA é um mecanismo capaz de inserir o Brasil na corrida regulamentar da IA, mas necessita de fortes e eficazes políticas públicas destinadas à sua implementação.

Palavras-chave: Estratégia Brasileira de Inteligência Artificial; governança; inteligência artificial; políticas públicas; regulação.

Abstract: This article has as its research problem the following question: how and which public policies can be appropriate for the effectiveness and implementation of the Brazilian Strategy for Artificial Intelligence (EBIA) regarding the axes Legislation, Regulation, and Ethical Use, and Governance of Artificial Intelligence (AI)? The aim is to present the EBIA and its axes of implementation and, afterward, to contextualize it according to Brazil's classification in terms of Artificial Intelligence in the international scenario. Global Innovation Index prepared by the World Intellectual Property Organization and the Artificial Intelligence Index Report prepared by Stanford University are used for this purpose. The methodology used for this is monographic research. In the end, it is concluded that the EBIA is a mechanism capable of inserting Brazil into the AI regulatory race, but it needs strong and effective public policies aimed at its implementation.

Keywords: Brazilian Artificial Intelligence Strategy; governance; artificial intelligence; public policy; regulation.

Resumen: Este artículo tiene como problema de investigación la siguiente pregunta: ¿cómo y qué

¹ Doutorando (2020 - Bolsista do Programa de Excelência Acadêmica - Proex - Capes/Taxa) e Mestre (2019) em Direito Privado pela Pontifícia Universidade Católica de Minas Gerais. Bacharel em Direito pelo Centro Universitário de Lavras (2017). Professor Adjunto do Curso de Direito do Centro Universitário de Lavras (2020 - atual). Professor substituto de Direito Privado da Universidade Federal de Lavras (03/2019 - 03/2021). Advogado. Orcid iD: <https://orcid.org/0000-0002-9037-0405>. Lattes: <http://lattes.cnpq.br/5133514180104561>. E-mail: sthefanoadv@hotmail.com.

políticas públicas pueden ser apropiadas para la efectividad e implementación de la Estrategia Brasileña de Inteligencia Artificial (EBIA) en relación a los ejes Legislación, Regulación y uso ético y Gobernanza de la Inteligencia Artificial (IA)? El objetivo es presentar la referida estrategia y sus ejes de implementación y, posteriormente, contextualizarla según la clasificación de Brasil en materia de Inteligencia Artificial en el escenario internacional. Para ello, se utiliza el Índice Global de Innovación elaborado por la Organización Mundial de la Propiedad Intelectual y el Informe del Índice de Inteligencia Artificial elaborado por la Universidad de Stanford. La metodología utilizada es la de la investigación monográfica. Al final, se concluye que el EBIA es un mecanismo capaz de insertar a Brasil en la carrera regulatoria de la IA, pero necesita de políticas públicas fuertes y efectivas dirigidas a su implementación.

Palabras clave: Estrategia Brasileña de Inteligencia Artificial; gobernanza; inteligencia artificial; políticas públicas; regulación.

1 Introdução: A Estratégia Brasileira de Inteligência Artificial (EBIA) e seus eixos de implementação

A Inteligência Artificial² (IA) tem sido rapidamente evoluída e desenvolvida objetivando oferecer um conjunto de benefícios de ordem econômica e social. Nos últimos anos, surgiu uma necessidade da adoção de medidas regulamentares e políticas públicas³ pelo executivo para implementação da IA tanto no segmento doméstico quanto no segmento industrial. Canadá (2017), China (2017), Dinamarca (2021), Comissão Europeia (2021), Finlândia (2017), França (2018), Índia (2018), Itália (2020), Japão (2018), México (2018), Países Nórdicos (2018), Singapura (2020), Coreia do Sul (2019), Suécia (2018), Taiwan (2018) e Reino Unido (2020) lançaram estratégias para promover o uso e o desenvolvimento da IA (DUTTON, 2018). Embora nenhuma das estratégias seja semelhante, enfoca-se em diferentes aspectos da IA: pesquisa científica; desenvolvimento e captura de talentos; habilidades e educação; adoção nos setores públicos e privados; ética e inclusão; padrões de regulamentação; e infraestrutura digital (DUTTON, 2018).

Ainda em nível internacional, em 22 de novembro de 2021, a cátedra da Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO, 2021) lança o primeiro acordo Internacional sobre Ética nas aplicações de IA. As recomendações também se direcionam à proteção de dados, à vedação da prática social de *score* (CITRON; PASQUALE, 2014; SERASA, 2021) e vigilância em massa, ao monitoramento e avaliação dos sistemas durante sua implementação e execução e à proteção ambiental, tendo em vista a possibilidade de uma IA tornar mais eficiente o uso de energia e outros serviços primários.

Reconhecendo que o grande aumento no poder computacional por meio de avanços práticos na aprendizagem de máquina (*machine learning*⁴) permite sucessos em uma gama de domínios aplicados, chamando a atenção para o desenvolvimento de políticas públicas e de

² Uma das definições mais aceitas no ramo científico – embora também criticada – é a de Russell e Norvig, que definem Inteligência Artificial como “[...] *the study of agents that receive percepts from the environment and perform actions*”. (RUSSELL; NORVIG, 2010, p. VIII).

³ Sobre definições e história das políticas públicas, ver mais em (SOUZA, 2002)

⁴ “*Machine learning is an evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment. They are considered the working horse in the new era of the so-called big data*” (EL NAQA; MURPHY, 2015. p. 3-11).

empresas para se juntar à corrida pela liderança mundial em IA, o Brasil (2021) lança sua Estratégia Brasileira de Inteligência Artificial (EBIA)⁵ objetivando sua atuação em campos como trabalho, educação, tributação, pesquisa, desenvolvimento e inovação e ética.

A EBIA (BRASIL, 2021) pretende nortear o papel das ações estatais para o desenvolvimento de ações, quaisquer que sejam suas vertentes, para estímulo à pesquisa, inovação e desenvolvimento de soluções em IA, bem como garantir que seu uso seja consciente, ético, legal e em benefício de um futuro melhor. A EBIA foi construída em três etapas:

- 1) Contratação de consultoria especializada em IA;
- 2) *Benchmarking* nacional e internacional;
- 3) Processo de consulta pública.

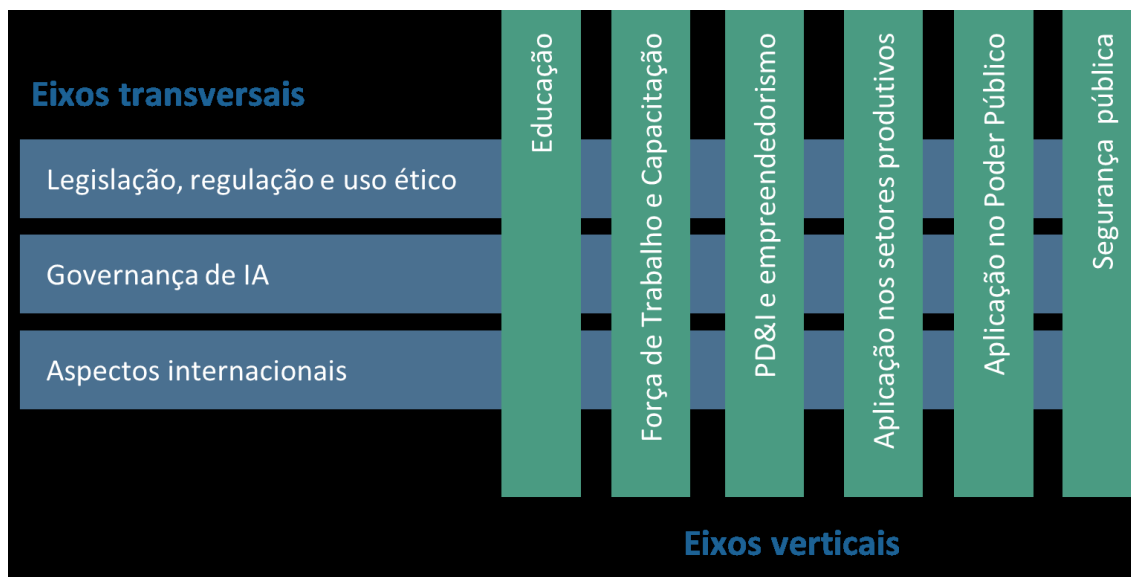
A contratação de consultoria especializada foi realizada pelo Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTI), por meio do Projeto de Cooperação Técnica Internacional (PRODOC) em conjunto à UNESCO. Já a consulta pública foi realizada por meio de plataforma eletrônica do Governo Federal, entre 12 de dezembro de 2019 e 3 de março de 2020, período no qual foram recebidas cerca de mil contribuições que foram utilizadas como base à construção da EBIA (BRASIL, 2021).

A EBIA segue as recomendações da Organização para Cooperação e Desenvolvimento Econômico (OCDE, 2019) sobre IA, destacando-se, mas não apenas: a) seu benefício às pessoas e ao planeta; b) o respeito ao Estado de Direito, aos direitos humanos, aos valores democráticos e à diversidade; c) transparência destinada à compreensão geral sobre os sistemas de IA; e d) funcionamento robusto ao longo de seu ciclo de vida.

A EBIA, portanto, trabalha com nove eixos temáticos, sendo três eixos transversais – 1) Legislação, regulação e uso ético; 2) Governança de IA⁶; 3) Aspectos internacionais –, e seis eixos verticais – 1) Educação; 2) Força de trabalho e capacitação; 3) PDI e empreendedorismo; 4) Ampliação no Poder Público; 5) Aplicação nos setores produtivos; e 6) Segurança pública. Os eixos transversais funcionam como base para a construção dos demais setores verticais. A todos esses devem estar presentes os pressupostos daqueles (ver Figura 1). Embora se aparentem independentes em razão de os assuntos serem diversificados e abrangerem particularmente cada setor, no escopo da criação dos eixos transversais todos eles devem ser abstratos suficientes para abraçar todos os setores de IA.

⁵ Instituída pela Portaria MCTI nº 4.617, de 6 de abril de 2021, alterada pela Portaria MCTI nº 4.979, de 13 de julho de 2021 (Anexo).

⁶ Para Floridi, “a governança digital é a prática de estabelecer e implementar políticas, procedimentos e padrões para o desenvolvimento, uso e gerenciamento adequados da *infosphere*. Através da supervisão humana, pretende-se garantir que um sistema de IA não coloque em causa a autonomia humana e nem produza efeitos negativos. A governança digital pode incluir diretrizes e recomendações que se sobrepõem à regulamentação digital, mas não são idênticas a ela. Essa é apenas outra maneira de falar sobre a legislação relevante, um sistema de leis elaborado e aplicado por meio de instituições sociais ou governamentais para regular o comportamento dos agentes relevantes na *infosphere*” (FLORIDI L. 2018).

Figura 1 – Eixos da EBIA

Fonte: Brasil (2021).

Nesse direcionamento, pode-se afirmar que os algoritmos⁷ devem ser neutros (*agnostics*) e não focarem especificamente em uma modalidade/espécie de IA ou de metodologia de análise de dados. A estratégia deve ser aplicada na constituição, aplicação e no uso de toda e qualquer IA. Em outros termos, a tecnologia é neutra e não deve focar em sistemas, *softwares* ou técnicas específicas, devendo sê-la aplicada independentemente do desenvolvimento da linguagem computacional e das técnicas de armazenamento de dados. A adoção dessas práticas serve como linhas de base para uma série de considerações e medidas para organizações operarem em qualquer setor e adotarem concretamente a estratégia. Setores específicos ou organizações específicas podem escolher adotar considerações adicionais e medidas técnicas para adaptar sua linha de produção de acordo com suas necessidades. Dessa forma, a estratégia não deve focar em organizações públicas ou privadas conforme seu tamanho ou modalidade constitutiva.⁸

A proposta para efetivação dessas pretensões é delineada por meio de objetivos. A EBIA pretende contribuir para elaboração de princípios éticos destinados ao desenvolvimento e uso de IA mais responsáveis. Ao mesmo tempo, pressupõe-se o equilíbrio entre investimentos sustentados em pesquisa e desenvolvimento em IA para remoção de barreiras à inovação em IA. Pretende-se capacitar e formar profissionais para o ecossistema da IA para estimular a inovação e o desenvolvimento da IA brasileira em ambiente internacional. Assim, pode-se promover

⁷ “Algoritmos são a base do processo de desenvolvimento de *software* e fazem parte das ferramentas pelas quais programadores criam estratégias para fracionar problemas em etapas e processos que podem ser traduzidos computacionalmente. Na tecnologia, há exemplos de todos os níveis de complexidade. Uma aplicação mais simples de algoritmo é o processo de inicialização do computador: há um *software* – basicamente a tradução computacional de um algoritmo – encarregado de testar todos os componentes do seu computador para saber se está tudo em ordem e, depois disso, procurar o sistema operacional no disco para carregá-lo” (GARRET, 2020).

⁸ Críticas extraídas de SINGAPORE (2020).

cooperação entre entes nacionais e internacionais, públicos e privados, indústria e centros de pesquisa para o desenvolvimento da IA.

Ocorre que tais objetivos, na conjuntura atual, não são fáceis de serem atingidos. Este artigo tem como **problema de pesquisa** o seguinte questionamento: como e quais políticas públicas podem ser adequadas para efetivação e implementação da EBIA quanto aos eixos Legislação, regulação e uso ético e Governança de IA?

Após contextualizá-la, a primeira seção é responsável pela análise da classificação do Brasil em matéria de IA no cenário internacional. Utiliza-se, para tanto, o Índice Global de Inovação elaborado pela Organização Mundial de Propriedade Intelectual e o *Artificial Intelligence Index Report* elaborado pela Universidade de Stanford. Neste ponto, realiza-se pesquisa quantitativa nos sítios Conecta Startup e Startup Brasil, objetivando verificar a adequação dos índices internacionais ao grau de inovação em território nacional. Como resultado, verifica-se que, de um total de 268 *startups*, predominam os setores: agronegócio (14), industrial (10), saúde e bem-estar (15), educação (23), saúde (17), TI e telecomunicações (19), varejo (13), finanças (11) e mídia e comunicação (12).

Em contribuições originais, constata-se que as políticas públicas destinadas ao avanço científico em tecnologia da informação e, especificamente, em IA não estão conforme as pretensões da EBIA. Vale ressaltar uma limitação deste estudo. Conforme apontado, a pesquisa qualitativa demonstra os setores de atuação como TI e telecom e *hardware*. Neste sentido, torna-se inviável verificar se as *startups* inscritas nesses ramos atuam especificamente no setor de IA. Ou seja, o resultado obtido pode ser mais negativo se verificado concretamente e de forma pormenorizada.

No mais, a segunda seção é responsável pelas contribuições ao eixo Legislação, regulação e uso ético, especialmente na temática responsabilidade civil dos atos práticos por IA. O principal resultado obtido postula que a responsabilidade adequada para os atos ilícitos praticados pelos entes inteligentes artificialmente é a responsabilidade subjetiva, pois ela permite a redução do dever indenizatório com base nos custos marginais de precaução. Para chegar a esse resultado, utiliza-se a Análise Econômica do Direito (AED) e uma breve revisão nas teorias da responsabilidade civil (subjetiva e objetiva) prescritas pelo Código de Defesa do Consumidor (CDC)⁹ e pelo Código Civil (CC).

Por fim, quanto ao eixo Governança de IA, apresentam-se propostas de estruturas e medidas de governança interna, determinação do nível de envolvimento humano na tomada de decisões com base na IA, gestão de operações, interação e comunicação das partes interessadas, tanto em instituições públicas quanto privadas, com fundamento no *Artificial Intelligence Governance Framework* de Singapura. A metodologia utilizada para tanto é a de pesquisa

⁹ Para mais, ver em: (DIVINO, 2021).

monográfica. Ao final, conclui-se que a EBIA é um mecanismo capaz de inserir o Brasil na corrida regulamentar da IA, mas necessita de fortes e eficazes políticas públicas destinadas à sua implementação.

2 Classificação do Brasil em matéria de IA no cenário internacional

A classificação do Brasil conforme indexadores em nível internacional pode trazer considerações sobre os impactos positivos e negativos que a nova economia e a sociedade da informação trazem consigo em matéria de IA. Em outros termos, essa análise permite ampliar a capacidade de postular políticas públicas e delimitar quais setores, pessoas, instituições ou empresas possuem a maior capacidade produtiva ou produzem produtos e serviços ligados a IA. Essas experiências internacionais demonstram sob outro ponto de vista que tais processos podem ou não estar elevando a produtividade em nível local com impacto global.

Como a inovação é ponto basilar e fundamental para o desenvolvimento econômico do país, além de estar atrelada nitidamente ao desenvolvimento da IA por ser técnica dependente de pesquisa, infraestrutura e desenvolvimento, o primeiro passo é recorrer ao Índice Global de Inovação referente ao ano de 2021¹⁰ (WIPO, 2021). Na classificação geral, o Brasil ocupa a posição 57 do ranking global, a melhor desde o ano de 2012, o que representa um avanço superior ao esperado se comparado ao ano de 2019, quando ocupava a 66° posição, e ao ano de 2020, na 62° posição. Relativo à região da América Latina e Caribe, apenas Chile (53°), México (55°) e Costa Rica (56°) estão à frente do Brasil (WIPO, 2021).

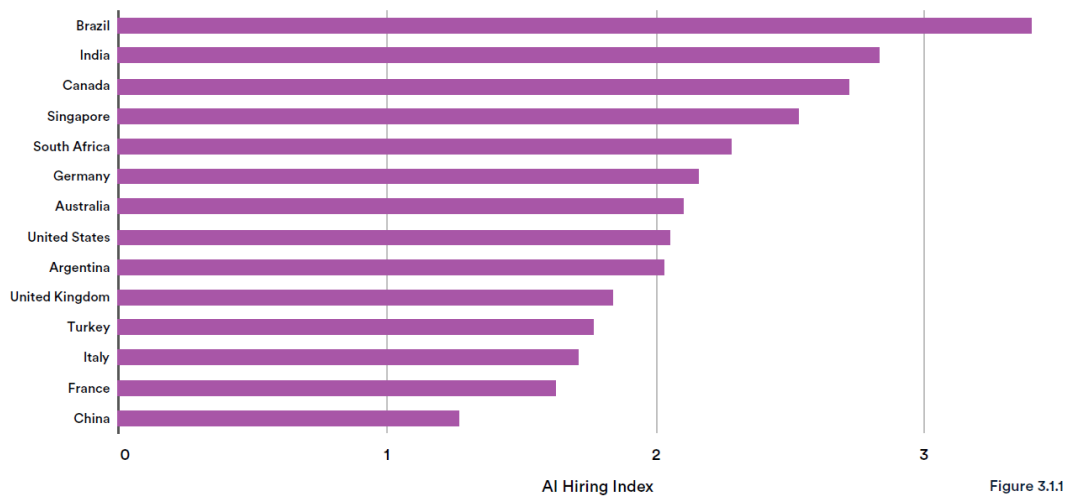
Com relação ao aparato institucional (78°), infraestrutura (69°), produtos criativos (61°) e sofisticação do mercado (75°), são esses os principais pontos a serem desenvolvidos para os próximos anos. Os recursos de capital humano destinados à pesquisa (48°), sofisticação empresarial (34°) e produtos de conhecimento e tecnologia (51°) são pontos que devem ser também aprimorados, mas que apresentam bons posicionamentos no ranking (WIPO, 2021).

Ainda em termos otimistas, conforme o *Artificial Intelligence Index Report* de 2021 elaborado pela Universidade de Stanford (UNITED STATES, 2021), em 2020, o Brasil, em conjunto com a Índia, Canadá, Singapura e África do Sul, esteve entre os países com maior taxa de empregabilidade no ramo da IA, conforme pode ser visto na Figura 2.

¹⁰ Este índice possui 132 países avaliados, entre os quais: Suíça; Suécia; Japão; Grécia; Romênia; Egito; Bangladesh; Mali; Togo; etc.

Figura 2 – Índice de Empregabilidade no setor de IA

AI HIRING INDEX by COUNTRY, 2020
Source: LinkedIn, 2020 | Chart: 2021 AI Index Report

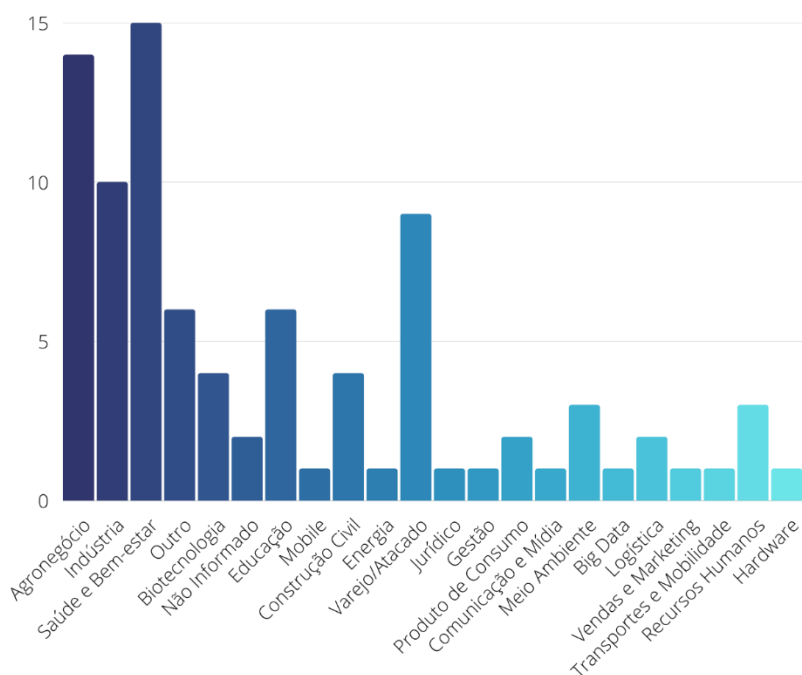


Fonte: United States (2021).

Mas o percurso ainda é longo. Quando analisados os dados trazidos pela EBIA, constata-se que, em 2020, o Brasil (2021) possuía cerca de doze mil *startups* ativas, mas apenas 26 delas eram relativas ao setor de IA. Objetivando a atualização dos dados, utilizando-se dos mesmos programas governamentais de fomento à inovação e empreendedorismo em nível de *startups* trazidos pela EBIA – Conecta Startup (2021) e Programa Startup Brasil (BRASIL, 2021) –, apresentam-se dados atualizados relativo ao ano de 2021.

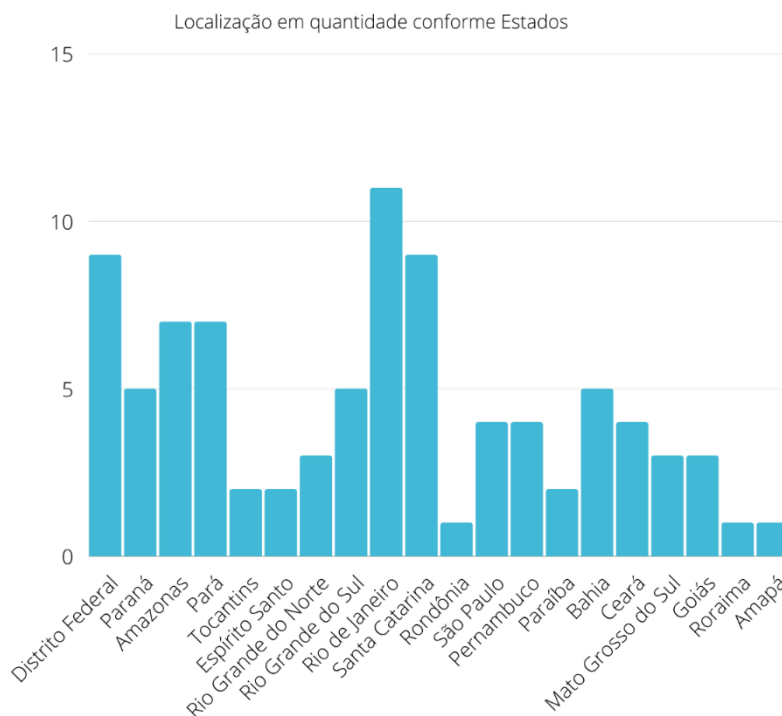
Conforme o Conecta Startup, foram identificadas 94 *startups* cadastradas. Como pode ser visto nas Figuras 3 e 4, predominantemente elas ocupam o setor de agronegócio (14), industrial (10) e saúde e bem-estar (15), sendo a maior parte do total (e não desses setores) localizada no Distrito Federal (9), Rio de Janeiro (11), Santa Catarina (9), Amazonas (7) e Paraná (7).

Figura 3 – Áreas de atuação (em quantidade) das *startups* conforme programa Conecta Startup



Fonte: Elaborado pelo autor com base nos dados fornecidos em Conecta Startup (2021).

Figura 4 – Estados de atuação (em quantidade) das *startups* conforme programa Conecta Startup



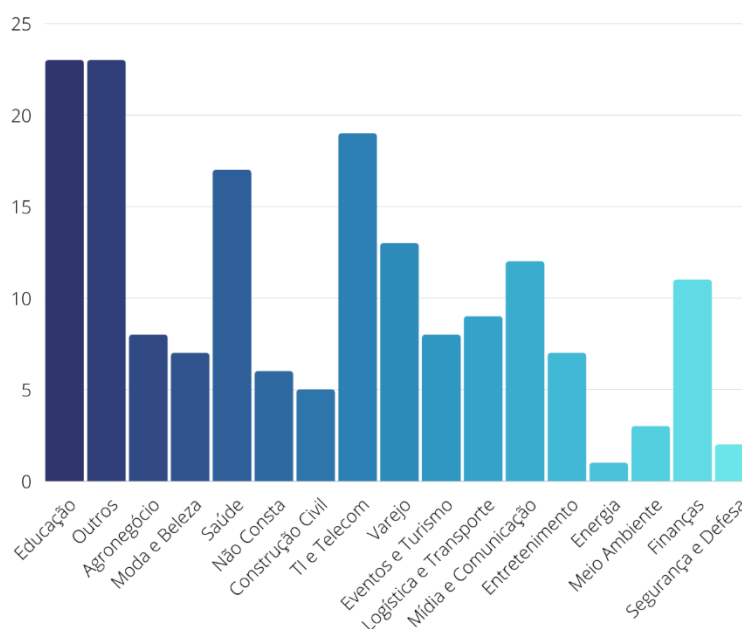
Fonte: Elaborado pelo autor com base nos dados fornecidos em Conecta Startup (2021).

As figuras demonstram uma série de limitações e pontos que necessitam de melhorias, dentre eles o maior esforço no desenvolvimento de políticas públicas voltadas para o fomento de atividades ligadas a IA tendo em vista a presença de uma única *startup* ligada a tecnologia (*hardware*).¹¹

O resultado obtido na plataforma anterior curiosamente se mescla quando se analisa o Programa Startup Brasil. Os dados coletados demonstram a existência de 174 *startups* cadastradas em sua plataforma. Predominantemente elas estão no setor de educação (23), saúde (17), TI e telecomunicações (19), varejo (13), finanças (11) e mídia e comunicação (12), sendo a maior parte delas concentrada em São Paulo (54), Rio de Janeiro (16), Pernambuco (12) e Minas Gerais (10). As informações estão disponíveis nas Figuras 5 e 6.

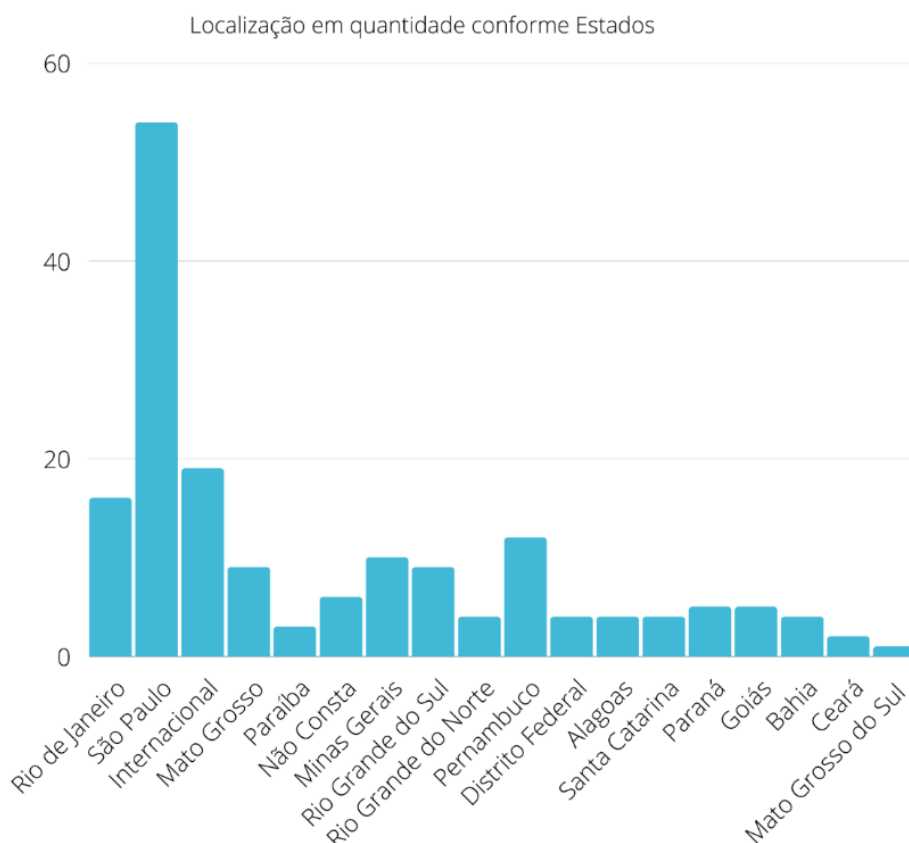
Mesmo a área de TI e telecomunicações sendo consideravelmente maior com relação ao programa anterior (19 para 1), o estudo também é limitado em razão de não conseguir verificar se as referidas *startups* atuam efetivamente no setor de IA. Porém, os resultados parciais que podem ser trazidos levam em consideração que, pela gama do território e da população brasileira, os números são ínfimos se comparados com as pretensões e resultados ao longo prazo.

Figura 5 – Áreas de atuação (em quantidade) das *startups* conforme Programa Startup Brasil



Fonte: Elaborado pelo autor com base nos dados fornecidos em Startup Brasil (BRASIL, 2021).

¹¹ Neste ponto, o estudo é limitado por não definir expressamente se ela atua no ramo da IA.

Figura 6 – Estados de atuação (em quantidade) das *startups* conforme Programa Startup Brasil

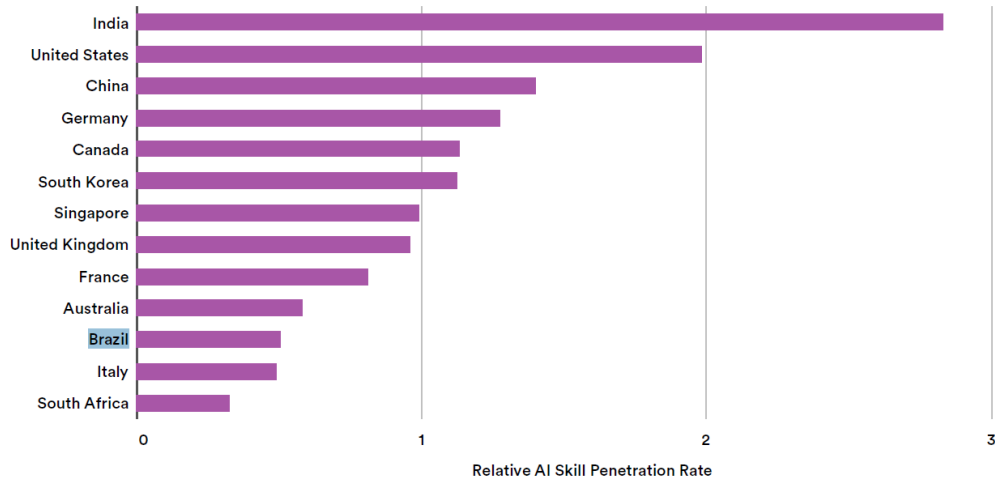
Fonte: Elaborado pelo autor com base nos dados fornecidos em Startup Brasil (BRASIL, 2021).

Um dos importantes resultados obtidos por meio da coleta de dados é a constatação de difusão de *startups* por todo o Brasil, o que demonstra considerável potencial empreendedor na cultura brasileira. Contudo, os principais problemas enfrentados pelas *startups* brasileiras são a escassez de mão de obra qualificada, a alta carga tributária e a burocracia (BRASIL, 2021). O princípio basilar para seu funcionamento é o desenvolvimento de soluções inovadoras e disruptivas para clássicos ou novos problemas com intuito de quebra de padrões em relação a empresas do mesmo segmento. Ocorre que essa característica necessita expressamente de uma relação inventiva e criativa em que o Brasil, apesar de seu contínuo desenvolvimento, ainda é precário.

Conforme o *Artificial Intelligence Index Report* de 2021 elaborado pela Universidade de Stanford (UNITED STATES, 2021), as habilidades em IA dos brasileiros apresentam baixa adesão, desenvolvimento, penetração e destaque, embora ainda estejam no ranking 50 em nível global (ver Figura 7), sendo a Universidade Federal do Rio Grande do Sul referência em produção relativa à ética de IA, *machine learning* e conferências sobre robótica.

Figura 7 – Índice de Habilidade em IA

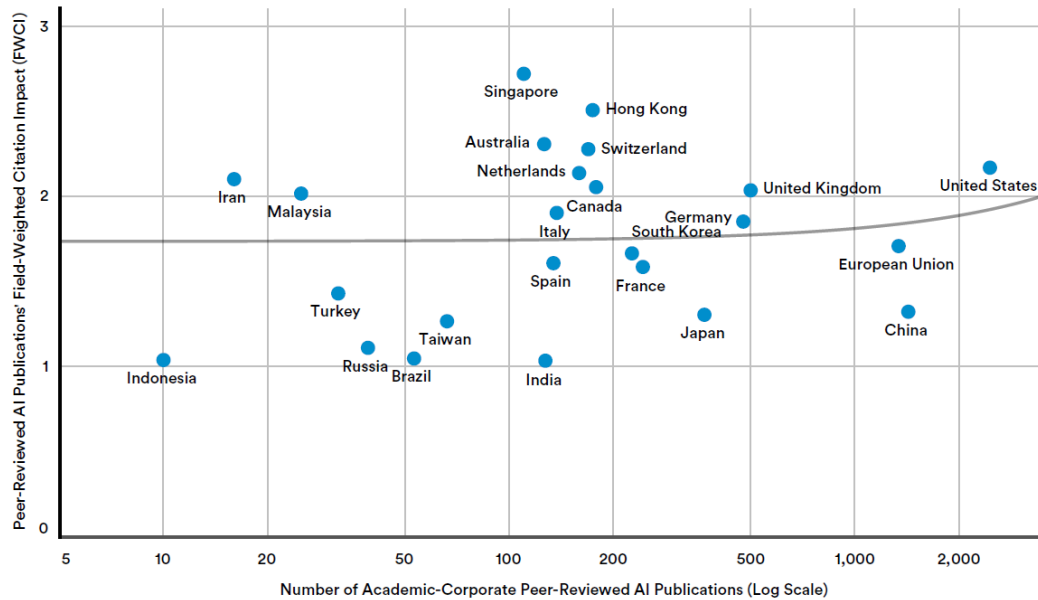
RELATIVE AI SKILL PENETRATION RATE by COUNTRY, 2015-20
Source: LinkedIn, 2020 | Chart: 2021 AI Index Report



Fonte: United States (2021).

Figura 8 – Publicações realizadas pelo sistema *double-blind review* na temática de IA

PEER-REVIEWED AI PUBLICATIONS' FIELD-WEIGHTED CITATION IMPACT and NUMBER of ACADEMIC-CORPORATE PEER-REVIEWED AI PUBLICATIONS, 2019
Source: Elsevier/Scopus, 2020 | Chart: 2021 AI Index Report



Fonte: United States (2021).

Como a EBIA fundamenta-se nos cinco princípios definidos pela OCDE para uma gestão responsável dos sistemas de IA, quais sejam: 1) crescimento inclusivo, o desenvolvimento sustentável e o bem-estar; 2) valores centrados no ser humano e na equidade; 3) transparência e explicabilidade; 4) robustez, segurança e proteção; e 5) a responsabilização ou a prestação de contas (*accountability*), cada um de seus eixos podem e devem ser desenvolvidos por meio de políticas públicas adequadas às suas pretensões. Afinal, os

resultados acima demonstram que o setor brasileiro, por mais promissor e por mais expectativas que existam, necessita de aprimoramentos para possibilitar o alcance de rankings superiores e, primordialmente, melhor satisfação da população.

3 Eixo temático 1: Legislação, regulação e uso ético

Nessa área temática, as maiores preocupações da EBIA são relativas a: 1) proteção de dados pessoais; 2) prevenção de discriminação e viés algorítmico; 3) equilíbrio destinado à preservação de estruturas adequadas de incentivo ao desenvolvimento da IA; 4) criação de parâmetros legais destinados à segurança jurídica quanto à responsabilidade dos diferentes atores integrantes de sua cadeia de produção.

As preocupações 2 e 3 podem ser tratadas inicialmente pela Lei Geral de Proteção de Dados (LGPD) brasileira. Quando o tratamento de dados é realizado em sua forma automatizada, poderá o titular solicitar ao controlador ou ao operador a explicação de como ele foi realizado, bem como suas finalidades.¹² A defesa pela explicabilidade em sistemas de IA advém do pressuposto de que suas decisões podem não ser consideradas necessariamente objetivas, justas ou imparciais (BECKER; FERRARI, 2018; BURELLI, 2016; DIAKOPOULOS, 2013; PASQUALE, 2015). Dessa forma, a criação de parâmetros para efetivar a explicabilidade e a transparência são mecanismos basilares que devem estar fundamentalmente atrelados para evitar vieses em decisões automatizadas e a discriminação na coleta e tratamento de dados (FJELD, 2020).

Sua concretização parte do pressuposto da exposição da lógica fundamental determinante, de forma clara e objetiva, responsável pela elaboração da decisão. Trata-se da apresentação de uma descrição que seja compreensível ao titular para delimitar como o operador ou o controlador utilizou as técnicas de IA para chegar ao resultado obtido (DOSHI-VELEZ; KORTS, 2017). Em outros termos, o que se pretende é a verificação dos parâmetros e critérios que foram utilizados no tratamento automatizado e se foram corretamente utilizados para evitar erros ou discriminações da decisão. Na constituição (*by design*) algorítmica deve-se verificar: 1) quais as principais razões que levaram àquela decisão; 2) se a mudança, substituição, alteração ou modificação dos fatores alteraria a decisão; e 3) se existem casos semelhantes com diferentes decisões etc. (DOSHI-VELEZ; KORTS, 2017)

Uma das formas de se concretizar a explicabilidade na tomada de decisão automatizada é pelo estabelecimento da obrigatoriedade de que os responsáveis pelo desenvolvimento deixem o código de seu software aberto (DIAKOPOULOS, 2013). Caso essa prática não seja possível, deve-se auditar particularmente o algoritmo por meio de uma cooperação entre empresas e

¹² Art. 20. O titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade. (LGPD)

governo, objetivando a comparação do resultado com o comportamento esperado na constituição da IA (SANDVIG, 2014).

A relação entre transparência e *accountability*, portanto, perpassa pela necessidade de adoção de medidas destinadas à compreensão dos processos associados à tomada de decisão automatizada para possibilitar a identificação de vieses envolvidos no processo decisório. Ocorre que essa prática pode ser efetivada e implementada partindo da ideia de que os sistemas de IA devem estar centrados no ser humano (*human-centric AI*), o que poderia facilitar sua auditoria, bem como a criação de sistemas mais confiáveis (*trustworthy AI*).

Um dos maiores desafios destinados à implementação deste objetivo é a falta de conhecimento técnico por parte da população em geral para compreender, interpretar e ler o código responsável pela execução da IA. Uma solução paliativa, que não deve ser tomada como definitiva, é o estabelecimento do dever de informar pelos fornecedores e pelos responsáveis acerca das decisões que estão sendo tomadas, se de forma automatizada, como, onde e quando elas são produzidas (OSOBA; WELSER, 2017).

O empenho da EBIA para alcançar este objetivo alinha o seu quadro regulatório de governança à criação de políticas públicas relacionadas ao tema e apresenta iniciativas consideradas relevantes para sua implementação. A primeira delas é a Estratégia Brasileira para Transformação Digital (e-Digital) (BRASIL, 2018a), a qual busca coordenar as diversas iniciativas governamentais ligadas ao tema em torno de uma singular visão, que seja sinérgica e coerente, para apoiar a digitalização dos processos produtivos e a capacitação para o ambiente digital, promovendo, em conjunto, a geração de valor e crescimento econômico (BRASIL, 2018a). Ao total, estão previstas 100 ações, que vão desde: 1) conectar 22.000 escolas públicas, urbanas e rurais, com acessos banda larga de alta velocidade, em rede terrestre ou via satélite, no âmbito do Programa Educação Conectada; e 2) ampliar o engajamento dos centros de pesquisa e desenvolvimento nas instâncias multilaterais de definição dos padrões internacionais e de faixas de radiofrequência a serem estabelecidos para a quinta geração da telefonia móvel (5G) a otimização de políticas voltadas ao setor de modo a expandir (mais que proporcionalmente) o investimento privado em PD&I em TICs, melhorar a competitividade da economia brasileira, gerar mais empregos com valor agregado e promover maior desenvolvimento social (BRASIL, 2018a).

A EBIA também indica o Programa Startup Brasil, destinado ao apoio de *startups* brasileiras e internacionais para desenvolvimento de *software*, *hardware*, serviços de TI ou que se utilizem dessas tecnologias para inovar. Contudo, conforme visualizado anteriormente, são poucas as empresas ligadas ao programa, e as destinadas ao desenvolvimento da área pretendida ainda são menores.

Nesse contexto, acredita-se que a criação de um Centro Nacional de Inteligência Artificial (CNAI) seja adequado para auxiliar a conexão e o contato entre a sociedade e as

instituições governamentais por meio de experts em IA. No mais, o CNAI poderá atuar especificamente em projetos destinados à criação, ao uso e ao aprimoramento da IA em múltiplos setores da economia. Essa prática facilitaria a colaboração entre os setores industriais e pesquisadores, objetivando a atração de maiores investimentos em razão da existência de mão de obra qualificada.

Outro desafio no setor regulatório relativo à IA são os aspectos éticos e de responsabilidade civil. A EBIA não postula como devemos resolver essa questão, mas serão apresentadas considerações pontuais a seguir.

3.1 Contribuições e propostas em responsabilidade civil da IA

Para tratar da temática, friso que não se aborda sobre a possibilidade de uma IA ser ou não sujeito de direito, mas sobre a responsabilidade ética e jurídica advinda de seus atos. A atenção, neste escopo, será direcionada a um panorama disruptivo em que a IA incidentemente na práxis e pode vir a causar danos à sociedade civil. Mesmo quando tentamos demonstrar os requisitos 1) agente; 2) dano; 3) culpa; e 4) nexos de causalidade, o primeiro é basicamente inverificável. Dessa maneira, entendo que uma IA não pode ser pessoalmente responsável pelos danos causados a terceiros, ao menos neste momento. A pretensão é verificar como as teorias da responsabilidade civil podem ou não ser suficientes na tratativa desse problema posto atualmente

A empresa Microsoft havia inserido um *chat-bot* denominado Tay para interagir autonomamente com os usuários dessa rede social. Seu desenvolvimento foi pretensioso: por meio do ML, Tay deveria ser capaz de compilar informações em seu banco de dados para criar discursos compreensíveis (PEREZ, 2016). Contudo, a empresa desativou a IA após denúncias de que Tay estaria fazendo *tweets* racistas e misóginos (VINCENT, 2016). Sob esse prisma, pouco tempo depois o projeto do Facebook semelhante ao da Microsoft foi desativado, pois percebeu que suas duas IAs (Alice e Bob) estavam interagindo entre si de uma forma ininteligível.¹³

Embora fossem situações controláveis e fáceis de serem verificadas e sanadas, as IAs tinham um potencial ofensivo que poderia se desenvolver em cadeia em duas grandes redes sociais. Esse desenvolvimento poderia chegar a grupos extremistas e servir de fundamento para discriminar outros grupos em rede.

Quando analisamos os casos de aplicabilidade de Responsabilidade pelo Fato do Produto no CDC, normalmente o atrelamos a danos pessoais e patrimoniais de ordem física. Em alguns casos, até mesmo a perda ou a deterioração do produto em razão de seu mau

¹³ Parte do diálogo pode ser expressa: “A few days later, some coverage picked up on the fact that in a few cases the exchanges had become - at first glance - nonsensical: Bob: “I can can I I everything else” Alice: “Balls have zero to me to me to me to me to me to me to me to me to me to” (BARANIUK, 2017).

funcionamento. Esse produto normalmente é físico, embora existam as possibilidades dos produtos digitais. Mas, quando inserimos a IA nesse contexto, vamos para além de danos de ordem física. Uma IA, por exemplo, pode causar um dano à privacidade do titular daquele produto. A Amazon, por exemplo, coleta detalhes íntimos sobre seus usuários por meio da Alexa, sua IA.¹⁴

Outro exemplo que pode ser trazido é o de fechaduras eletrônicas e inteligentes que podem ceifar a autonomia de seus titulares, caso sejam utilizadas por terceiros. Em 2017, um Hotel na Austrália sofreu um ataque de *hackers* que conseguiram invadir o sistema de segurança da empresa, em especial o sistema de fechadura eletrônica. Para entrar novamente em seu quarto e reaver os pertences, foi exigido dos hóspedes uma quantia de \$ 1.800 (BILEFSKY, 2017).

Ocorre que a responsabilidade no CDC possui algumas peculiaridades, dentre elas a objetividade e a fundamentação no risco. Inicia-se pela segunda. A relação entre risco e atividade é sempre uma relação probabilística. Em outros termos, o pilar central da teoria do risco no CDC, conforme entendimento deste autor, é a possibilidade de prever ou não a existência dos danos. No âmbito da IA, essa previsão não é fácil de ser realizada. Vimos há pouco a relação entre Bob e Alice como algo totalmente fora dos padrões pré-estabelecidos. O *Deep Learning* (DL) e o *Machine Learning* (ML) são feitos para, necessariamente, não prever os comportamentos de uma IA. Afinal, se a pretensão é replicar o comportamento do ser humano, nada mais adequado e inspirador que replicar sua aleatoriedade.

De certa forma, não se nega a previsibilidade por meio dos sistemas de ML. Existem regras que podem ser inseridas no sistema de IA para que atuem conforme pré-determinado. Mas essas regras possuem limites semânticos e sintáticos. Além disso, quanto maior o número de regras, maior a complexidade e, conseqüentemente, maior os custos de operação dessa IA. Preencher o conteúdo de uma IA com todas as regras possíveis em conteúdos linguísticos e probabilísticos computacionais parece ir além das pretensões do ML.

Frisa-se que não são todos os atos da IA que necessariamente são imprevisíveis, tal como de um veículo autônomo que para em uma faixa de pedestre para um transeunte passar. A questão em si é que o risco de causar danos não é um risco meramente técnico, mas também um risco de ordem normativa, fugindo ao escopo do próprio CDC quando de sua análise.

¹⁴ “A Amazon coleta dados sobre os consumidores por meio de sua assistente de voz Alexa, as compras em seu marketplace, os leitores eletrônicos Kindle, além da plataforma de música. Assim, a empresa reúne uma vasta gama de informações sobre seus clientes nos Estados Unidos e começou a disponibilizar esses dados para todos mediante solicitação no início do ano passado, depois de tentar e não conseguir derrotar uma medida de 2018 da Califórnia que exigia tais divulgações. [...] O dossiê de um repórter revelou que a Amazon coletou mais de 90 mil gravações Alexa de familiares entre dezembro de 2017 e junho de 2021 - uma média de cerca de 70 por dia. As gravações incluíam detalhes como os nomes dos filhos e suas músicas favoritas. A Amazon capturou as crianças perguntando como elas poderiam convencer seus pais a deixá-las brincar e recebendo instruções detalhadas da Alexa sobre como convencer seus pais a comprar videogames. Algumas gravações envolveram conversas entre membros da família usando dispositivos Alexa para se comunicar em diferentes partes da casa. Várias gravações capturaram crianças de sete a 12 anos fazendo perguntas a Alexa sobre termos como "pansexual". O repórter não percebeu que a Amazon estava armazenando as gravações”. (O GLOBO, 2021).

Acredita-se, com isso, que o risco da atividade em desenvolvimento de IA transcende em muito os riscos tutelados pelo CDC. Mas de que forma isso ocorre? Primeiro, precisamos rever alguns conceitos. A responsabilidade pelo risco do produto tenta resolver alguns problemas, dos quais: 1) defeitos decorrentes de projeto, fabricação, construção, montagem, fórmulas, manipulação, apresentação ou acondicionamento de seus produtos; e 2) informações insuficientes ou inadequadas sobre sua utilização e riscos. Trata-se de responsabilidade objetiva pelos danos, ainda que não intencionais, ao consumidor, sejam eles oriundos do produto em si, enquanto defeituoso, ou pela falha de informação no que tange ao seu uso e risco.

Se uma IA não funcionasse como inicialmente ela deveria, bastaria verificar em sua constituição, projeto ou fabricação, por exemplo, que se trata de um defeito do produto e, conseqüentemente atrairia responsabilidade ao fabricante, ao produtor, ao construtor, nacional ou estrangeiro, e ao importador.¹⁵ O defeito é a chave para concretização da responsabilidade dos sujeitos acima identificados perante o consumidor.

Existem algumas abordagens estabelecidas pelo CDC para verificar a ocorrência de defeitos no produto: apresentação, uso e riscos e época que foi colocado em circulação. Isso significa que o fabricante deverá comprovar a inexistência desse defeito em ao menos três modalidades estabelecidas legalmente. Sob o aspecto econômico, o responsável deve adotar medidas que reduzem o risco imposto ao desenvolvimento do produto por meio de técnicas de precaução e prevenção de danos que possivelmente seriam a chave de um ato ilícito. Trata-se dos custos de precaução defendido pela AED, cuja inobservância atrai a responsabilidade objetiva ao agente.

Ocorre que a IA, ainda assim, pode não apresentar nenhum defeito e, mesmo assim, causar danos ao consumidor. É aqui que o desafio começa. Uma das excludentes da responsabilidade objetiva do art. 12, §3º do CDC é a comprovação de que o produto foi colocado no mercado e o defeito inexistente.

Além disso, existe uma **expectativa de segurança sobre o produto** (art. 12, §1º do CDC), que se torna quase impossível de se verificar na prática em razão da novidade tecnológica. E, dentro desse tema, como traçar informações suficientes e adequadas sobre os possíveis riscos do produto (art. 12, caput, parte final, CDC), sendo que o principal problema da IA é sua imprevisibilidade e, conseqüentemente, a indefinição dos riscos em termos probabilísticos (LEMLEY; CASEY, 2019)? Em outros termos, se a imprevisibilidade é parte do produto, como inserir e aceitar essa imprevisibilidade como parte dos riscos da atividade, sendo que ela é desconhecida? Seria difícil e economicamente inviável incluir na base de banco de dados todos os riscos operacionais dessa IA. Caso contrário, se o fabricante pelos riscos

¹⁵ Nos termos do art. 12, §1º do CDC, “O produto é defeituoso quando não oferece a segurança que dele legitimamente se espera, levando-se em consideração as circunstâncias relevantes, entre as quais: I - sua apresentação; II - o uso e os riscos que razoavelmente dele se esperam; III - a época em que foi colocado em circulação.”

imprevisíveis fossem apenados, estaríamos transformando-os em previsíveis e reconhecendo sua responsabilidade em todos os casos possíveis. Em outros termos, ele nunca conseguiria concretizar a excludente de ilicitude contida no art. 12, §3º, II do CDC, pois a tecnologia seria em si defeituosa. O raciocínio ao contrário pode ser verificável e também inteligível. Rachum-Twaig prescreve que

uma vez que os riscos relacionados à IA são imprevisíveis por natureza e, portanto, não podem ser cobertos pelo defeito de projeto ou pelo dever de advertência e doutrina de instrução, pode haver casos de danos fora do escopo da doutrina de responsabilidade por produtos; mas esses casos podem ser compensados por outras formas de responsabilidade em delitos (RACHUM-TWAIG, 2020, p. 1141).

Outro problema oriundo da análise de risco que visualizo não apenas no CDC é a responsabilidade objetiva. Diferentemente da responsabilidade subjetiva pautada na culpa, a objetividade é essencialmente verificável no risco.¹⁶ O problema geral da ausência de previsibilidade no comportamento da IA é crucial para uma interpretação normativa adequada, vez que sua incidência é indispensável para adequar e realocar os riscos sob o prisma da responsabilidade objetiva. Além disso, existe uma assimetria informacional latente entre fabricantes, desenvolvedores, programadores e todos os envolvidos na cadeia. Em grande parte, os desenvolvedores conhecem e possuem maiores informações sobre os riscos da IA que os fabricantes. Esses também possuem um nível consideravelmente superior de informações sobre a IA que os destinatários finais. Rachum-Twaig (2020, p. 1163) traz exemplos de um robô atuante na medicina que possui mais informações sobre os riscos do procedimento a ser realizado do que o próprio paciente a ser operado. Ressalta o autor que, como a principal característica dos robôs baseados em IA é que eles podem agir de uma forma imprevisível ou inexplicável pelos humanos, em muitos casos nenhum dos interessados estará melhor situado para avaliar os riscos envolvidos em sua operação, e o problema da informação imperfeita se aplicará igualmente a todos os interessados. Nesses casos, os princípios orientadores de aplicabilidade da responsabilidade estrita não podem ser aplicados.

Na cadeia produtiva, a redução dos custos será direcionada exclusivamente aos *designers* e desenvolvedores da tecnologia, que terão que obter todo e qualquer tipo de informação relacionada ao risco para mitigá-lo e erradicá-lo, na medida do possível. Sob o prisma de relação entre custos e benefícios, torna-se inviável a programação dessa forma. A imprevisibilidade está intrínseca na programação eletrônica, e seus riscos devem ser compreendidos para além daqueles estabelecidos no CDC e na responsabilidade objetiva por não serem previsíveis.

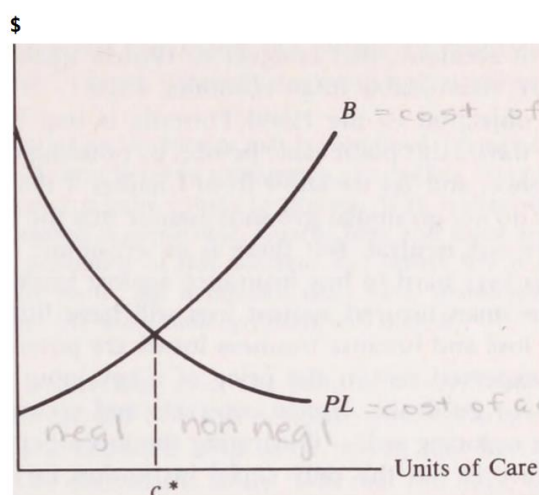
Entendo, portanto, que a adoção do CDC nesses casos somente dificultaria e

¹⁶ “*fault liability makes wrongful agency the fundamental basis of responsibility for harm accidentally done; strict liability makes agency itself the fundamental basis of responsibility*” (KEATING, 2001, p. 1285).

transformaria o ambiente tecnológico em um ambiente normativo dotado de regras e preceitos jurídicos sem aplicabilidade prática e somente com reflexos negativos. Teoricamente, poderíamos concluir que o regime da responsabilidade subjetiva seria ideal, vez que a objetividade deve ser deixada de lado. No prisma da AED, a fórmula de Hand pode ser suscitada como mecanismo direcionado a solucionar essa problemática. Veremos, contudo, que não é tão fácil.

A fórmula de Hand, conforme Posner, impõe às partes da relação jurídica custos marginais de precaução (B) para evitar possíveis danos esperados (PL) advindos de uma multiplicação entre a probabilidade de que esse dano realmente aconteceria (c) e os danos propriamente ditos (d) (POSNER, 2012).¹⁷ Representa-se da seguinte forma:

Figura 9 – Custos de Prevenção



Fonte: Posner (2012).

Pode-se extrair do gráfico acima que, para Hand, a conduta será culposa quando o investimento em precaução (B) seja menor que os possíveis danos esperados (PL). Assim, $B < PL \times c$ (POSNER, 2012, p. 148). A curva decrescente representada por PL representa uma mudança marginal nos custos dos danos esperados em função da precaução adotada. Dito isso, verifica-se que a adoção de práticas de prevenção e precaução tem um efeito decrescente na prevenção de acidentes. Por outro lado, a curva assinalada por B exemplifica o custo marginal dos cuidados e ascende na medida em que as práticas de precaução são escassas e, portanto, eleva-se à medida que mais produtos e serviços são oferecidos no mercado de consumo. Na interseção entre as duas curvas, verifica-se o ideal dever de cuidado onde existe separação entre a negligência e a prudência/precaução (POSNER, 2012).

Quando se aplica essa análise sob a ótica da responsabilidade objetiva, se se determina

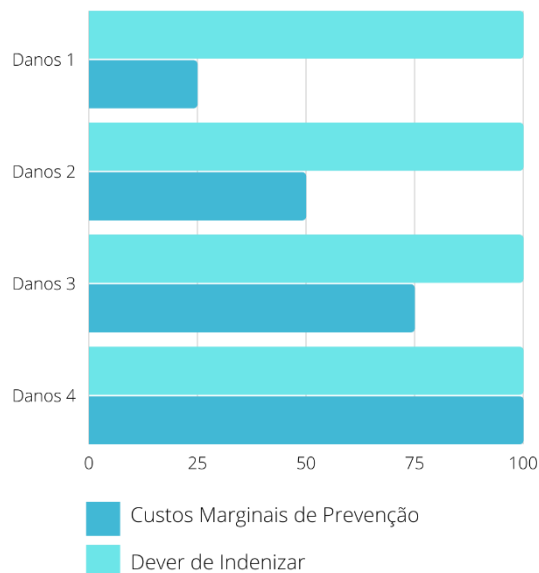
¹⁷ Para mais sobre AED ver em (SHAVELL, 2009)

que a reparação dos danos deve ser feita de forma incondicional e invariavelmente pelo seu desenvolvedor, produtor, fabricante ou responsável pela tecnologia, dispensando-se análise de culpa, negar-se-ia e ausentar-se-ia os interesses para a implementação de novas práticas destinadas à precaução à ocorrência de novos danos.

Portanto, trata-se de um desestímulo e de desincentivo para adoção de custos marginais de prevenção. Caso a sociedade empresária seguisse todos os preceitos recomendados pela legislação e adotasse as práticas de governança e precaução e, ainda assim, fosse responsabilizada por todo e qualquer dano ocorrido em sua álea contratual, os custos nessas ações seriam dispensáveis, tendo em vista que a responsabilização aconteceria com ou sem eles. Dessa forma, adotando-se a responsabilidade objetiva, verifica-se que os custos marginais de precaução (B) para evitar possíveis danos esperados (PL) são indiferentes aos danos propriamente ditos (d).¹⁸ Expresso-os da seguinte forma:

Figura 10 – Custos Marginais x Responsabilidade Civil Objetiva

Relação entre custos marginais de prevenção e responsabilidade civil objetiva



Fonte: Elaborada pelo autor.

Perceba-se que na responsabilidade civil objetiva, independentemente das ações tomadas, o critério indenizatório sempre permeia a totalidade dos danos insurgidos. Dessa forma, caso essa modalidade (objetiva) seja adotada, não faria sentido algum para o desenvolvedor ou para uma sociedade empresária investir e adotar custos de transação

¹⁸ Para melhor entendimento da AED, especialmente em Posner, recomenda-se (GAROUPA; PORTO; FRANCO, 2019).

destinados à prevenção, pois, de qualquer forma, seriam obrigados a indenizar os titulares lesados no futuro. Assim, trata-se de uma relação de perda efetiva e um dispêndio desnecessário na política empresarial.

Sob o contexto econômico e legal, se a fórmula de Hand estiver correta, verifica-se que a possibilidade de adoção da responsabilidade civil em sua modalidade subjetiva parece ser a mais adequada, pois permitiria que os custos destinados à precaução não fossem reduzidos a zero em razão de uma responsabilização incondicional.

Glaubitz e Raymond (2021) reconhecem a possibilidade da fórmula de Hand em solucionar o problema de imputação de responsabilidade pelos atos da IA e inserem uma terceira variável no cálculo probabilístico: o dever de cuidado.

Fórmula de Glaubitz e Raymond (2021) para verificação de responsabilidade de IA:

$$B - \frac{k}{\min\{A\}} < P \cdot L$$

Na fórmula acima, B é o dever e o cuidado em ter a tarefa automatizada completada manualmente por um humano; k é uma constante que determina a influência de A ; A é o custo mínimo de uma solução algorítmica alternativa (requalificação do algoritmo existente, criação de um novo algoritmo ou utilização de um algoritmo existente que atinge o mesmo propósito legítimo com um menor impacto nas classes protegidas); P é a probabilidade de um dever de cuidado pré-existente, determinado por uma avaliação de impacto díspar que avalia a extensão de qualquer impacto adverso desproporcional em uma classe protegida (identificado através da auditoria do algoritmo); e L é a perda ou dano decorrente do viés algorítmico experimentado por uma classe protegida (GLAUBITZ; RAYMONG, 2021, p. 31).

Em termos práticos, A seria como uma sanção de natureza cível. Se o custo marginal de prevenção é alto, assim o agente responsável pelo ilícito seria menos penalizado. Se o custo dispendido no dever de cuidado for baixo, deverá a indenização ser valorada conforme sua culpa. E, segundo Glaubitz e Raymond (2021), a responsabilidade variará conforme os graus de automação do produto e sua supervisão. Quanto mais automatizados, maior dever de cuidado é necessário para mitigar o potencial ofensivo à sociedade. Esse dever de cuidado seria complementar ao custo de prevenção P previamente existente em Hand.

Porém, tanto em Hand quanto em Glaubitz e Raymond (2021) existe um fator indispensável a ser verificado em suas fórmulas: o risco. O risco é, em suma, um elemento que deve ser objetivamente aferível para ocorrer sua redução. Quando falamos na programação de IA, esse risco é difícil – e em alguns casos impossível – de ser delimitado em sua totalidade. Não se consegue verificar ações *ex ante* para prevê-las e evitá-las *ex post*. As avaliações de

impacto ficam prejudicadas enquanto instrumento adequado à aplicação da fórmula de Hand e, conseqüentemente, decaímos em uma atividade completa de riscos e incertezas.

Portanto, em matéria de responsabilidade civil, tem-se que a modalidade subjetiva parece ser mais adequada para implementação dos preceitos da EBIA quanto à manutenção da inovação e ao estímulo ao seu incentivo. Por esse motivo que políticas públicas destinadas ao financiamento de projetos de pesquisa que busquem aplicar soluções não discriminatórias, baseadas na equidade/não discriminação (*fairness*), responsabilidade/prestação de contas (*accountability*) e transparência são as matrizes para o estímulo de parcerias com empresas que estejam pesquisando soluções comerciais e sociais dessas tecnologias.

Em outros termos, ao Estado compete estabelecer padrões e requisitos técnicos que sejam compatíveis com a promoção de uma IA responsável, mas ao mesmo tempo em que atenda às formas de responsabilidade já previamente definidas em padrões legais. Caso seja necessário, o mapeamento das barreiras legais e regulatórias será imprescindível à identificação para atualização destinada à promoção da segurança jurídica. Essa prática somente será possível quando realizados estudos mais aprofundados sobre a temática.

Portanto, conforme estabelecido pela EBIA, a criação de políticas de controle de qualidade de dados para efetivação do ML e do DL sob parâmetros intervencionistas humanos pode auxiliar o encontro de resultados efetivos com baixo grau de riscos para a sociedade e para o indivíduo. E, novamente, a criação de um centro específico para a IA, com experts na temática, é essencial para essa concretização.

À medida que analisamos como a estrutura de responsabilidade pode ser implementada, nos encontramos no setor de governança. Neste ponto, o aspecto fundamental é a criação de práticas de gestão para monitoramento e supervisão dos sistemas de IA.

4 Eixo temático 2: Governança de IA

A EBIA, neste ponto, postula dois princípios basilares e nortes da estrutura de governança: transparência e *accountability* (aqui trazida nos termos da administração pública e compreendida como responsabilidade e prestação de contas). Esses dois parâmetros impõem ao responsável pela aplicação de IA estabelecer estruturas destinadas a assegurar que sua implementação possa ser analisada pelo princípio da precaução, identificando-se as aplicações de alto risco que podem impactar significativamente a sociedade em um contexto de aplicação determinado, tal como saúde ou monitoramento do espaço público.

Segundo a EBIA, a intervenção regulatória quanto à governança deve ser equilibrada entre o grau de risco relacionado à determinada aplicação de IA e as possíveis limitações que possam restringir seus usos. Para tanto, a EBIA propõe a elaboração de relatórios periódicos de impacto de proteção de dados (RIPDs), que podem ser identificados em cada setor de atuação de IA, tais como: Relatório de Impacto de Segurança (RIS); Relatório de Impacto Ambiental

(RIA); ou Relatório de Impacto de Direitos Humanos (RIDH).

Em âmbito legal, o Decreto n. 8.777/2016, responsável pelo Portal Brasileiro de Dados Abertos, bem como o Portal do Software Público Brasileiro, regido pela Portaria n. 46/2016, já são medidas e políticas destinadas à transparência dos projetos. Contudo, o sistema de governança deve partir de um pressuposto de colaboração estrutural do setor público e do setor privado para o desenvolvimento de padrões gerenciais de riscos associados ao uso da IA.

Em outros termos, utilizando-se o *Artificial Intelligence Governance* de Singapura, existem quatro medidas chaves a serem adotadas para efetivar um modelo de governança de IA:

- a. Estruturas e medidas de governança interna destinadas a adaptar as estruturas e medidas já existentes ou estabelecer novas para incorporar valores, riscos e responsabilidades relacionados com a tomada de decisões algorítmicas.
- b. Determinar o nível de envolvimento humano na tomada de decisões com base na IA: Uma metodologia para ajudar as organizações a estabelecer seu apetite de risco para o uso de IA, ou seja, determinar riscos aceitáveis e identificar um nível apropriado de envolvimento humano na tomada de decisão com base em IA.
- c. Gestão de operações: Questões a serem consideradas ao desenvolver, selecionar e manter modelos de IA incluindo gerenciamento de dados.
- d. Interação e comunicação das partes interessadas: Estratégias para a comunicação com as partes interessadas de uma organização, e a gestão das relações com elas (SINGAPORE, 2020).

Trata-se de um modelo flexível em que as organizações públicas ou privadas podem alterá-lo conforme suas necessidades e aspectos culturais. O importante, neste caso, é que as diretrizes podem auxiliar e ajudar as organizações a compreender como implementar cada uma das práticas acima descritas.

4.1 Estruturas e medidas de governança interna

As estruturas e medidas de governança interna têm como objetivo garantir uma supervisão robusta sobre o uso da IA por uma organização, seja ela pública ou privada. As estruturas de governança interna existentes podem ser adaptadas, e/ou novas estruturas podem ser implementadas, se necessário. Por exemplo, os riscos associados ao uso de IA podem ser gerenciados dentro da estrutura de gerenciamento de risco empresarial, enquanto considerações éticas podem ser introduzidas como valores corporativos e gerenciadas através de conselhos de revisão ética ou estruturas similares (SINGAPORE, 2020).

Alguns fatores são relevantes para sua criação e efetivação:

- 1) Funções e responsabilidades claras para o uso ético da IA: como o seu desenvolvimento é realizado por etapas, estágios e atividades, em cada uma delas pode existir a alocação individual de responsabilidade ao seu desenvolvedor. Dessa forma, os departamentos que possuem uma estrutura interna de governança podem conhecer suas funções e estar totalmente alertas sobre suas funções e sobre suas responsabilidades. Dentre as práticas

adotadas estão: a) a manutenção, monitoramento, documentação e revisão dos modelos de IA desenvolvidos; b) revisão dos canais de comunicação e interação com os usuários para prover um efetivo *feedback*; e c) garantir que os empregados/desenvolvedores sejam tecnicamente aptos e propiamente treinados para lidar com os sistemas de IA (SINGAPORE, 2020).

2) Gestão de riscos e controles internos: trata-se de uma opção para identificar, corrigir e lidar com riscos internos quando do desenvolvimento da IA. Pode-se utilizar esforços razoáveis para assegurar que os conjuntos de dados utilizados para o treinamento do modelo de IA sejam adequados para o propósito pretendido, e para avaliar e gerenciar os riscos de imprecisão ou parcialidade, bem como rever as exceções identificadas durante o treinamento do modelo. Praticamente nenhum conjunto de dados é completamente imparcial. As organizações devem se esforçar para compreender as formas pelas quais os conjuntos de dados podem ser tendenciosos e abordar isso em suas medidas de segurança e estratégias de implantação (SINGAPORE, 2020).

Um dos exemplos capazes de ilustrar a implementação de medidas e estruturas de governança interna é a adotada pela empresa Mastercard. Para garantir uma supervisão robusta do uso de IA pela Mastercard, a empresa estabeleceu um Conselho de Governança para revisar e aprovar a implementação de aplicações de IA que são determinadas como de alto risco. O Conselho de Governança é presidido por seu Vice-Presidente Executivo do Centro de Excelência em Inteligência Artificial, e cujos membros incluem o *Chief Data Officer*, o *Chief Privacy Officer*, o *Chief Information Security Officer*, cientistas de dados e representantes de equipes comerciais (SINGAPORE, 2020).

4.2 Determinação do nível de envolvimento humano na tomada de decisões com base na IA

A EBIA propõe uma abordagem *Human-Centric AI*. Em termos singelos, essa abordagem coloca a IA como uma ferramenta em que agentes humanos decidem quando e como usá-la. Segundo Muller, “precisamos de uma abordagem HIC da IA, onde as máquinas permanecem máquinas e as pessoas mantêm o controle sobre essas máquinas o tempo todo” (UNIÃO EUROPEIA, 2018). Em sua concepção, “agentes humanos podem e também devem ter o controle de se, quando e como a IA é usada no cotidiano, bem como quais tarefas transferimos para a IA, quão transparente é, e o respeito aos aspectos éticos” (UNIÃO EUROPEIA, 2018). Pressupõe-se, portanto, o monitoramento e a supervisão do sistema de IA desde a concepção do sistema (*privacy by design, security by design, human rights by design, ethics by design*).

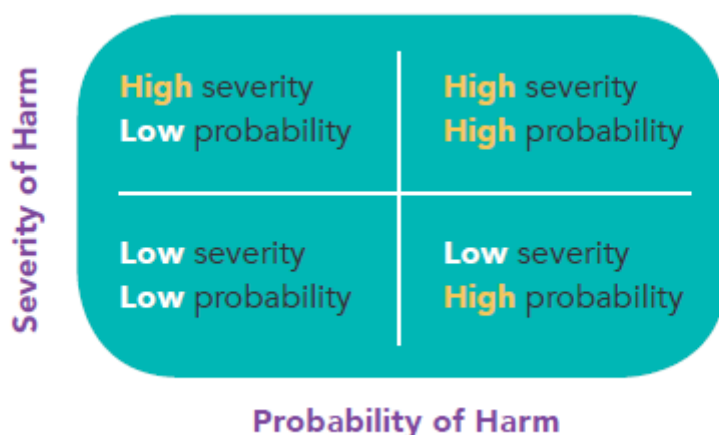
Ocorre que, para atingir esses objetivos, antes de disponibilizar a IA no mercado, as empresas e o governo devem decidir sobre seus objetivos comerciais e verificar quais são os riscos na sua utilização e na tomada de decisões. Quanto a multinacionais, deve-se considerar diferentes fatores culturais e diferentes sistemas normativos ou de valores. No mais, alguns

riscos para determinados sujeitos podem se manifestar apenas quando aplicados em um determinado grupo (tal como a oferta automatizada de produtos que pode causar uma mudança comportamental e, conseqüentemente, compras compulsivas).

No mais, como a identificação dos objetivos e riscos comerciais, bem como a determinação de atuação humana na tomada de decisões automatizadas, são processos interativos e contínuos, é esperado que as organizações continuem identificando-os e revendo-os constantemente para incrementar melhores soluções em suas tecnologias, mitigando, portanto, riscos e mantendo uma efetiva resposta a ações previamente falhas.

Conforme considerações extraídas da EBIA, torna-se interessante pensar na automação não como ausência do envolvimento humano em uma determinada tarefa, mas incluí-la seletivamente na expectativa de que o resultado será um processo efetivo das características da automação inteligente (WANG, 2019).¹⁹ Esse modelo estrutural parte de dois grandes eixos: 1) probabilidade e 2) capacidade de causar danos a um indivíduo (ou organização) como resultado da decisão obtida no processo.

Figura 11 – Relação danos x probabilidade



Fonte: Singapore (2020).

A proposta dos autores estrangeiros é padronizada e semelhante à já existente na teoria geral da administração pública. No setor público, da mesma forma que no setor privado, o gestor deve utilizar ferramentas que possam auxiliar na tomada de decisões. As ferramentas são importantes ainda mais quando se trata de riscos ambientais, de saúde, econômicos, ergonômicos e outros tantos que precisam ser gerenciados e controlados nas instituições de todas as áreas da administração pública, uma vez que o bem-estar da sociedade é a essência do serviço público. Torna-se indispensável tomar decisões e usar medidas corretas com relação às políticas e programas públicos, sendo fundamental adotar estratégias efetivas de gestão de

¹⁹ Para mais (DIVINO; MAGALHAES, 2020).

riscos.

Uma das principais regulamentações para auxiliar o gestor público no exercício de seu cargo e de sua função enquanto pessoa do povo é a Instrução Normativa Conjunta CGU/MP nº 1 de 2016. Esse normativo tenta assegurar que os responsáveis pela tomada de decisão, em todos os níveis do órgão ou entidade, tenham acesso tempestivo a informações suficientes quanto aos riscos aos quais está exposta a organização. Para isso, deve-se compreender que a postura do chefe do executivo disseminada em suas redes sociais ou em seu ambiente “privado” não reflete apenas em sua esfera subjetiva, mas também na daqueles que adotam o mesmo posicionamento externado por ele. Por esse motivo, as condutas praticadas, ainda que sob o espectro da liberdade de expressão, devem levar em consideração o impacto na sociedade, principalmente no âmbito crítico da saúde.

A gestão de riscos parte do pressuposto impacto x probabilidade. Quando externada uma posição negativa da vacinação, deve-se verificar qual o impacto que ela surtirá na sociedade, bem como na probabilidade de que isso aconteça. De acordo com o Tribunal de Contas da União, pode-se exemplificar da seguinte forma:

Figura 12 – Método de avaliação de riscos

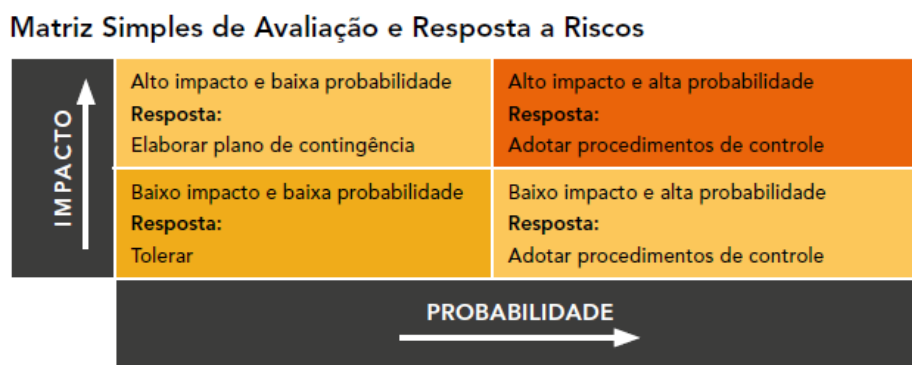


Figura 3: Matriz 2 x 2 de resposta a risco (INTOSAI GOV 9130, traduzido e adaptado)

Fonte: Brasil (2018b).

Quanto maior o impacto e maior sua probabilidade, maior será o esforço necessário para evitá-lo na medida do possível. Trata-se, portanto, de uma relação diretamente proporcional destinada à mitigação de riscos.

Figura 13 – Modelo de gerenciamento de Risco

Modelo de Gerenciamento de Risco

		AÇÕES DE GERENCIAMENTO DE RISCO		
		IMPACTO	Alto	6 Considerável esforço de gerenciamento é necessário
Médio	3 Riscos podem ser aceitos, com monitoramento		5 Esforço de gerenciamento é necessário	7 Esforço de gerenciamento exigido
Baixo	1 Aceitar Riscos		2 Aceitar, mas monitorar riscos	4 Gerenciar e monitorar riscos
		Baixa	Média	Alta
		PROBABILIDADE		

Figura 4: Matriz 3 x 3 de gerenciamento de risco (Secretaria do Tesouro do Canadá)

Fonte: Brasil (2018b).

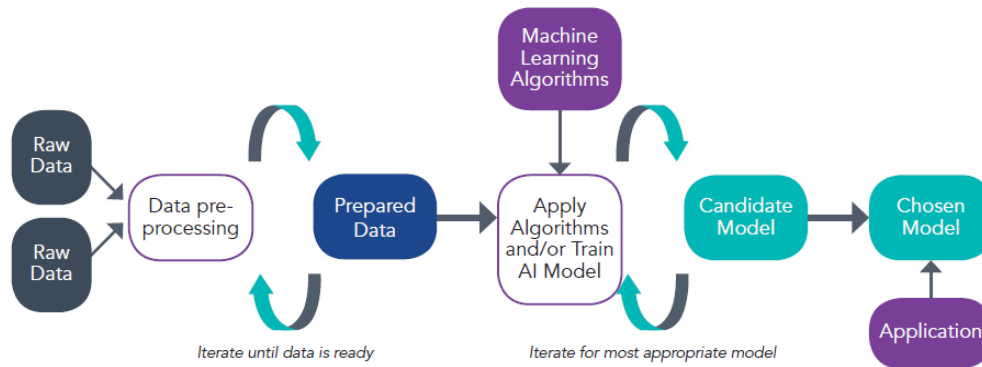
Com essas práticas em ação, há considerável possibilidade da melhoria da qualidade dos sistemas de IA em níveis governamentais e privados. No tocante às políticas públicas do setor estatal, os relatórios de impacto de proteção de dados podem adquirir novas feições e incrementar resultados concretos destinados à promoção da IA em conjunto com sua responsabilidade e eticidade. Trata-se de um diálogo social e multissetorial indispensável a alavancar as práticas de *accountability* relacionadas à IA nas organizações.

4.3 Gestão de operações

Para que as práticas do tópico anterior sejam implementadas, uma série de processos e operações devem estar intrinsecamente ligadas à organização de dados, algorítmico e da própria constituição da IA. A EBIA, contudo, não estabelece tais critérios. Porém, o *Artificial Intelligence Governance Framework* de Singapura propõe a organização em três passos: 1) preparação dos dados; 2) análise do algoritmo; e 3) escolha do modelo adequado ao proposto.

O primeiro estágio, a coleta de dados é realizado, formatado e tratado para conclusões mais assertivas serem obtidas. Neste caso, a precisão e os *insights* aumentam conforme a quantidade de dados na base de treinamento da IA. Posteriormente, esses modelos são treinados em banco de dados e uma análise algorítmica será realizada. Aqui se pode incluir análises estatísticas ou modelos de aprendizado de máquina extraído de redes neurais. Os resultados são examinados e inseridos nos modelos mais apropriados. Para tanto, a análise será probabilística e incorporada em aplicações para oferecer predições, tomar decisões e resolver problemas.

Figura 14 – Modelo de Gestão Operacional em IA



Fonte: Singapore (2020).

Todo o processo passa pela compreensão linear de todos os dados, desde como, quando e de onde eles chegaram, bem como foi coletado, tratado e transferido, e para onde, quando e como eles irão. A chave para um procedimento transparente perante os demais órgãos será a existência de registros que permitem à organização, seja ela pública ou privada, garantir a qualidade dos dados conforme sua origem. Neste meio termo, práticas destinadas a minimizar vieses discriminatórios na tomada de decisões também serão plenamente adotadas. Um dos vieses mais comuns a serem tratados refere-se à questão racial e estereótipa e à omissão de determinadas pessoas conforme sua ascendência em um grupo econômico ou social. Por essa razão, a relação entre diferentes bancos de dados para treinamento, teste e validação da IA em instituições públicas e privadas é essencial para garantir uma revisão periódica e atualizar constantemente os resultados obtidos.

A proposta regulamentar de Singapura direciona às organizações uma aplicação modular algorítmica baseada em riscos representada em duas etapas. A primeira deve identificar o subconjunto de características ou funcionalidades que têm o maior impacto sobre as partes interessadas para as quais tais medidas são relevantes. Em segundo lugar, identificar quais dessas medidas serão mais eficazes para criar confiança com suas partes interessadas. Algumas dessas medidas, como explicabilidade (ou repetibilidade, quando se usa modelos que não são facilmente explicados), robustez e sintonia regular, são suficientemente essenciais para que possam, em diferentes graus, ser incorporadas como parte do processo de implementação de IA da organização. Outras medidas, tais como reprodutibilidade, rastreabilidade e auditoria, são mais intensivas em recursos e podem ser relevantes para características específicas ou em cenários específicos (SINGAPORE, 2020).

Neste mesmo sentido, a explicabilidade pode ser atingida quando da delimitação de como os algoritmos funcionam ou como o processo de tomada de decisão incorpora modelos de

predição. Trata-se de uma correlação entre compreensão e confiança nos sistemas que estão sendo desenvolvidos e implementados. Mas ainda existem limitações tecnológicas e sociais quanto à explicabilidade, conforme já mencionado anteriormente. Onde a explicabilidade não pode ser alcançada, as organizações podem considerar documentar a repetibilidade dos resultados produzidos pelo modelo de IA. A repetibilidade refere-se à capacidade de executar uma ação ou tomar uma decisão de forma consistente, dado o mesmo cenário. Embora a repetibilidade (dos resultados) não seja equivalente à explicabilidade (do algoritmo), algum grau de garantia de consistência no desempenho poderia proporcionar aos usuários de IA um grau maior de confiança (SINGAPORE, 2020).

Outros fatores, como robustez²⁰, melhoria contínua²¹, rastreabilidade²², padronização e possibilidade de reprodução²³ e auditoria²⁴ da IA, são indispensáveis a serem implementados e destinados à compreensão processual linear e de segurança.

Por fim, políticas públicas destinadas ao desenvolvimento da IA somente são possíveis mediante interações e comunicações das partes interessadas que, neste caso, é a sociedade civil e as instituições públicas e privadas.

4.4 Interação e comunicação das partes interessadas

Como a incorporação da IA no poder público tem como fundamento a referência da potencialização das atividades humanas, o poder público, neste caso, não possui interesse na satisfação de interesses particulares, o que relegaria o interesse público. Em outros termos, a implementação da IA nesse setor deve ser realizada pelas vias de equidade e inclusão social.

²⁰ “Robustness refers to the ability of a computer system to cope with errors during execution and erroneous input and is assessed by the degree to which a system or component can function correctly in the presence of invalid input or stressful environmental conditions. Ensuring that deployed models are sufficiently robust will contribute towards building trust in the AI system” (SINGAPORE, 2020).

²¹ “Establishing an internal policy and process to perform regular model tuning is effective for ensuring that deployed models cater for changes to customer behaviour over time. This allows organisations to refresh models based on updated training datasets that incorporate new input data. Model tuning may also be necessary when commercial objectives, risks, or corporate values change” (SINGAPORE, 2020).

²² “An AI model is considered to be traceable if (a) its decisions, and (b) the datasets and processes that yield the AI model’s decision (including those of data gathering, data labelling and the algorithms used), are documented in an easily understandable way. The former refers to traceability of AI-augmented decisions, while the latter refers to traceability in model training. Traceability facilitates transparency and explainability and is also helpful for other reasons. First, the information might also be useful for troubleshooting, or for an investigation into how the model was functioning or why a particular prediction was made. Second, the traceability record (in the form of an audit log) can be a source of input data that can be used as a training dataset in the future” (SINGAPORE, 2020).

²³ “While repeatability refers to the internal repetition of results within one’s organisation, reproducibility refers to the ability of an independent verification team to produce the same results using the same AI method based on the documentation made by the organisation. Reproducibility can influence the trustworthiness of the AI product and the organisation deploying the AI model. As implementing reproducibility entails the involvement of external parties, organisations can take a risk-based approach towards identifying the subset of AI-powered features in their products or services that requires external reproducibility testing” (SINGAPORE, 2020).

²⁴ “Auditability refers to the readiness of an AI system to undergo an assessment of its algorithms, data and design processes. The evaluation of the AI system by internal or external auditors (and the availability of evaluation reports) can contribute to the trustworthiness of the AI system as it demonstrates the responsibility of design and practices and the justifiability of outcomes. It should, however, be noted that auditability does not necessarily entail making information about business models or intellectual property related to the AI system publicly available” (SINGAPORE, 2020).

Para tanto, as organizações governamentais devem disponibilizar informações gerais sobre quais setores, produtos e serviços aplicações de IA serão efetivamente utilizadas. Neste ponto, políticas e termos de uso são fortemente recomendados para serem desenvolvidos. Eles devem conter explicações para auxiliar seus usuários a se comunicar com a IA, bem como a solicitar informações ao ente estatal responsável pelo seu desenvolvimento e uso. No mais, a política também pode conter suas diferentes funções e a responsabilidade individualizada de cada membro responsável pela sua elaboração.

Perceba-se que a explicação detalhada pode auxiliar a comunicação entre o setor público e aumentar consideravelmente a transparência na tomada de decisões. Assim, tanto a explicabilidade quanto a transparência são mecanismos apropriados e indispensáveis para interação e comunicação entre os setores e, conseqüentemente, aumento da confiança nesses tipos de aplicações. Como o setor governamental é amplo, o primeiro passo para delimitar essa política e os termos de serviço é identificar o público-alvo da prestação de serviços e, após, inseri-los em um contexto em que suas pretensões e propostas sejam compatíveis com as aplicações de IA.

Informações mais gerais podem ser direcionadas para potenciais usuários decidirem se aderem ou não ao serviço executado pela IA. Ocorre que, caso seja aceito, informações mais específicas devem ser direcionadas para demonstrar como a aplicação funciona. Por fim, caso seja necessária a intervenção humana, um passo-a-passo deve ser indicado para seus usuários, bem como, se eles solicitarem informações sobre o funcionamento da tomada de decisões automatizadas, elas precisam estar incluídas em um parâmetro de *feedback*.

Em outros termos, percebam que a criação de canais de retorno (*feedback*) destinados à avaliação do usuário é indispensável para a organização, principalmente para o *Data Protection Officer* tomar suas decisões e verificar a adequação das práticas à legislação. Assim o deve ser, pois, conforme a Comissão Europeia, todos os governos em democracias constitucionais devem ser limitados pela lei, inclusive aqueles que utilizam de sistemas de IA para manter ou expandir seus processos democráticos.

Nota-se que os parâmetros estabelecidos para implementação de políticas públicas, embora individualizados, constituem um conjunto processual indissociável para garantir a inclusão apropriada do sistema na sociedade. Segurança e transparência, conforme já evidenciado, são instrumentos basilares para maximizar as inovações entre regulação e economia.

Diante o exposto, verifica-se que a EBIA possui grandes expectativas em nível internacional de implementação, mas necessita de um elevado e contínuo trabalho em nível de políticas públicas para sua efetivação.

5 Conclusão

O problema de pesquisa inicialmente proposto foi como e quais políticas públicas podem ser adequadas para efetivação e implementação da EBIA quanto aos eixos Legislação, regulação e uso ético e Governança de IA. Constatou-se na primeira seção que a EBIA, em nível nacional, possui grandes desafios a serem cumpridos, dentre eles o maior incentivo a pesquisa, tecnologia, inovação e desenvolvimento em nível nacional e para *startups*. Nesse caso, os programas direcionados ao fomento estão residualmente aplicados em setores de TI. Conforme visualizado, uma das limitações do estudo com os dados obtidos é traçar o âmbito concreto de atuação, sendo que Tecnologia da Informação é um setor muito amplo. Assim, os resultados na prática podem ainda ser piores quando concretamente localizados nos setores relativos à IA.

As principais contribuições deste trabalho, portanto, residem na constatação de que a responsabilidade civil dos atos ilícitos praticados por IA ou durante o curso de sua atuação, execução e implementação deve ser a modalidade subjetiva. Nessa forma, permite-se a mitigação dos danos conforme custos de precaução efetivamente alocados para configuração de uma IA ética, sólida e robusta. Quanto aos aspectos de governança, verifica-se que políticas públicas devem ser desenhadas para criação de relatórios de impacto de segurança e relatórios contínuos do uso de IA nos sistemas governamentais. Medidas de governança interna que incorporam fases de implementação são essenciais para sua concretização. Portanto, devem ser estipuladas funções e responsabilidades claras para o uso ético da IA. Além disso, trata-se de uma opção para identificar, corrigir e lidar com riscos internos quando do desenvolvimento da IA.

No mais, a EBIA propõe uma abordagem *Human-Centric AI*. Em termos singelos, essa abordagem coloca a IA como uma ferramenta em que agentes humanos decidem quando e como usá-la. Nesse ponto, a implementação de políticas públicas pode seguir as tradicionais teorias da administração com relação à gestão de riscos, ou seja, a relação entre impacto x probabilidade. No tocante às políticas públicas do setor estatal, os relatórios de impacto de proteção de dados podem adquirir novas feições e incrementar resultados concretos destinados à promoção da IA em conjunto com sua responsabilidade e eticidade. Trata-se de um diálogo social e multissetorial indispensável a alavancar as práticas de *accountability* relacionadas à IA nas organizações.

Essa abordagem pode ser efetivada mediante gestões operacionais em que todo o procedimento é acompanhado do início ao fim, delimitando-se sua: origem, finalidade, coleta e tratamento de dados, aplicação, decisão, modelos regulamentares e modelos estruturais para aplicação. É indispensável ao gestor público conhecer seu público-alvo para elaborar termos de serviços adequados e suficientemente explicáveis aos usuários. Em outros termos, todos os processos de gestão pública devem estar pautados na explicabilidade para atingir uma

delimitação de como os algoritmos funcionam ou mesmo como o processo de tomada de decisão foi realizado, bem como na transparência, para que se possa obter informações de todos os tipos de forma inteligível e direcionada ao desenvolvimento contínuo, rastreabilidade, padronização e auditoria das práticas de IA. Por essa razão, é importante a comunicação e a interação entre as partes envolvidas. Informações mais gerais podem ser elaboradas, desde que direcionadas a um primeiro contato. À medida que o cidadão se depara com o efetivo uso da IA, documentos específicos devem ser elaborados contendo um passo-a-passo para sua correta compreensão, informando, também, os responsáveis pela execução.

As considerações finais aqui apresentadas de forma alguma são conclusões. São apenas indicações exemplificativas para a criação de políticas públicas e documentos direcionados especialmente ao *feedback* do usuário final. As tecnologias que envolvem IA estão constantemente evoluindo e atreladas ao caráter ético, sendo que sua governança não deve e nem pode possuir parâmetros fixos para sua aplicação. Caso isso aconteça, existiria em si uma contradição na própria estrutura de governança, que requer uma revisão periódica para adequar os sistemas ao pretendido pelo usuário, pelo responsável e pelo público. Portanto, são caminhos iniciais a serem traçados e complementados por uma estratégia que ainda terá um longo e **real** caminho a ser percorrido.

Referências

- BARANIUK, Chris. The 'creepy Facebook AI' story that captivated the media. **BBC**. 2017. Disponível em: <https://www.bbc.com/news/technology-40790258>. Acesso em: 20 dez. 2021.
- BECKER, Daniel; FERRARI, Isabela. O direito à explicação sobre decisões automatizadas: uma análise comparativa entre a União Europeia e o Brasil. **Revista de Direito e as Novas Tecnologias**, São Paulo, vol. 1, n. 1, out./dez. 2018.
- BILEFSKY, Dan. Hackers Use New Tactic at Austrian Hotel: Locking the Doors. **The New York Times**. 2017. Disponível em: <https://www.nytimes.com/2017/01/30/world/europe/hotel-austria-bitcoin-ransom.html>. Acesso em: 20 dez. 2021.
- BURRELL, Jenna. How the machine ‘thinks’: Understanding opacity in machine learning algorithms. **Big Data & Society**, [s.l.], v. 3, n. 1, p.1-12, jan. 2016.
- BRASIL. Estratégia Brasileira de Inteligência Artificial – EBIA. **MCTI**. 2021. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial>>. Acesso em: 30 jan. 2022.
- BRASIL. **Estratégia Brasileira para Transformação Digital (e-Digital)**. 2018a. Disponível em: <https://www.gov.br/mcti/pt-br/centrais-de-conteudo/comunicados-mcti/estrategia-digital-brasileira/estrategiadigital.pdf>. Acesso em: 31 jan. 2022.
- BRASIL. Manual de Gestão de Riscos. **Tribunal de Contas da União**. 2018b. Disponível em: <https://portal.tcu.gov.br/planejamento-governanca-e-gestao/gestao-de-riscos/manual-de-gestao-de-riscos/>. Acesso em: 30 jan. 2022.
- BRASIL. **Startup Brasil**. 2022. Disponível em: <https://www.startupbrasil.org.br/startups/conheca-as-startups-do-programa/>. Acesso em: 30 jan. 2022.

- CANADÁ. **Pan-Canadian AI Strategy**. 2017. Disponível em: <https://cifar.ca/ai/>. Acesso em: 30 jan. 2022.
- CHINA. **A Next Generation Artificial Intelligence Development Plan**. 2017. Disponível em: <https://na-production.s3.amazonaws.com/documents/translation-fulltext-8.1.17.pdf>. Acesso em: 30 jan. 2022.
- CITRON, D. K.; PASQUALE, F. A. The Scored Society: Due Process for Automated Predictions. **Washington Law Review**, Washington, v. 89, n. 1, p. 2-27, 2014.
- CONECTA STARTUP BRASIL. **Startups**. 2021. Disponível em: <https://conectastartupbrasil.org.br/startups>. Acesso em: 30 jan. 2022.
- COMISSÃO EUROPEIA. **Artificial Intelligence Act**. 2021. Disponível em: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>. Acesso em: 30 jan. 2022.
- DIAKOPOULOS, Nicholas. Algorithmic Accountability Reporting: On the Investigation of Black Boxes. **Tow Center for Digital Journalism**, 2013.
- DINAMARCA. **Strategy for Denmark's Digital Growth**. 2021. Disponível em: https://eng.em.dk/media/10566/digital-growth-strategy-report_uk_web-2.pdf. Acesso em: 30 jan. 2022.
- DIVINO, S. B. S. Desafios e benefícios da inteligência artificial para o Direito do Consumidor. **Revista Brasileira de Políticas Públicas**, Brasília, v. 11, n. 1, p. 655-689, 2021.
- DIVINO, S. B. S.; MAGALHAES, R. A. Inteligência Artificial e Direito Empresarial: Mecanismos de Governança Digital para Implementação e Confiabilidade. **Economic Analysis of Law Review**, Brasília, v. 11, p. 72-89, 2020.
- DOSHI-VELEZ, Finale; KORTZ, Mason. Accountability of AI under the law: the role of explanation. **Berkman Klein Center Working Group on Explanation and the Law**, 2017.
- DUTTON, Tim. An Overview of National AI Strategies. **Medium**. 2018. Disponível em: <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>. Acesso em: 30 jan. 2022.
- EL NAQA, Issam; MURPHY, Martin J. What is machine learning?. In: **machine learning in radiation oncology**. Springer: Cham, 2015. p. 3-11.
- FINLAND. **Finland's Age of Artificial Intelligence**. 2017. Disponível em: https://knowledge4policy.ec.europa.eu/ai-watch/finland-ai-strategy-report_en. Acesso em: 30 jan. 2022.
- FJELD, Jessica et al. Principled artificial intelligence: Mapping consensus in ethical and rights-based approaches to principles for AI. **Berkman Klein Center Research Publication**, Cambridge, n. 2020-1, 2020.
- FLORIDI L. Soft ethics, the governance of the digital and the General Data Protection Regulation. **Philosophical Transactions of the Royal Society**, London, v. 376, n. 2133, 2018. Disponível em: <http://dx.doi.org/10.1098/rsta.2018.0081>. Acesso em:
- FRANCE. **AI for Humanity: French Strategy for Artificial Intelligence**. 2018. Disponível em: <https://super-ai.diascreative.net/ai-for-humanity-french-strategy-for-artificial-intelligence>. Acesso em: 30 jan. 2022.
- GAROUPA, Nuno; PORTO, Antonio José Maristrello; FRANCO, Paulo Fernando De Mello. As indenizações pela perda do tempo útil do consumidor. Espera e custos de oportunidade. **Revista de Direito do Consumidor**, Brasília, v. 124, p. 263-293, jul./ago. 2019
- GARRET, Filipe. O que é algoritmo? Entenda como funciona em apps e sites da Internet. **TechTudo**. 2020. Disponível em: <https://www.techtudo.com.br/listas/2020/05/o-que-e-algoritmo-entenda-como-funciona-em-apps-e-sites-da-internet.ghtml>. Acesso em: 31 jan. 2022.

- GLAUBITZ, Alina; RAYMOND, Nathaniel. **How should liability be attributed for harms caused by biases in Artificial Intelligence?**. 2021. Disponível em: https://politicalscience.yale.edu/sites/default/files/glaubitzi_alina.pdf. Acesso em: 20 dez. 2021.
- INDIA. **National Strategy For Artificial Intelligence**. 2018. Disponível em: <https://indiaai.gov.in/research-reports/national-strategy-for-artificial-intelligence>. Acesso em: 30 jan. 2022.
- ITALY. **Strategia Nazionale per l'Intelligenza Artificiale**. 2020. Disponível em: https://knowledge4policy.ec.europa.eu/ai-watch/italy-ai-strategy-report_en. Acesso em: 30 jan. 2022.
- JAPAN. **Artificial Intelligence Technology Strategy**. 2017. Disponível em: https://ai-japan.s3-ap-northeast-1.amazonaws.com/7116/0377/5269/Artificial_Intelligence_Technology_StrategyMarch2017.pdf. Acesso em: 30 jan. 2022.
- KEATING, Gregory C. The theory of enterprise liability and common law strict liability. **Vanderbilt Law Review**, Nashville, v. 54, p. 1285, 2001.
- KOREA. **National Strategy for Artificial Intelligence**. 2019. Disponível em: <https://www.msit.go.kr/bbs/view.do?sCode=eng&mId=10&mPid=9&bbsSeqNo=46&nttSeqNo=9>. Acesso em: 30 jan. 2022.
- LEMLEY, Mark A.; CASEY, Bryan. Remedies for robots. **The University of Chicago Law Review**, Chicago, v. 86, n. 5, p. 1311-1396, 2019.
- MEXICO. **Artificial Intelligence Agenda MX**. 2018. Disponível em: <https://oecd.ai/en/dashboards/countries/Mexico>. Acesso em: 30 jan. 2022.
- O GLOBO. Amazon coleta detalhes íntimos sobre seus usuários através da Alexa; veja quais. **IG Tecnologia**. 2021. Disponível em: <https://tecnologia.ig.com.br/2021-11-22/amazon-alexa-coleta-dados-intimos.html>. Acesso em: 20 dez. 2021.
- OCDE. **Recommendation of the Council on Artificial intelligence**. 2019. Disponível em: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>. Acesso em: 30 jan. 2022.
- OSOBA, Osonde; WELSER IV, William. An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence. **RAND Corporation**, Santa Mônica, 2017.
- PAÍSES NÓRDICOS. **Declaration on AI in the Nordic-Baltic Region**. 2018. Disponível em: https://www.regeringen.se/49a602/globalassets/regeringen/dokument/naringsdepartementet/20180514_nmr_deklaration-slutlig-webb.pdf. Acesso em: 30 jan. 2022.
- PASQUALE, Frank. **The Black Box Society: The Secret Algorithms that Control Money and Information**. Cambridge: Harvard University Press, 2015. 320 p.
- PEREZ, Sarah. Microsoft Silences Its New A.I. Bot Tay, After Twitter Users Teach It Racism, **TECHCRUNCH**. 2016. Disponível em: <https://techcrunch.com/2016/03/24/microsoft-silences-its-new-a-i-bot-tay-after-twitter-users-teach-it-racism/>. Acesso em: 20 dez. 2021.
- POSNER, Richard. **Economic Analysis of Law**. 3. ed. Alphen aan den Rijn: Wolters Kluwer, 2012. 1026 p.
- RACHUM-TWAIG, Omri. Whose Robot Is It Anyway? Liability for Artificial-Intelligence-Based Robots. **University of Illinois Law Review**, Illinois, v. 2020, p. 1141, 2020.
- RUSSELL, Stuart. J.; NORVIG, Peter. **Artificial intelligence: a modern approach**. 3. ed. New Jersey: Pearson Education, 2010. 1115 p.
- SANDVIG, Christian et al. Auditing algorithms: Research methods for detecting discrimination on internet platforms. **Data and discrimination: converting critical concerns into productive inquiry**, Seattle, v. 22, p. 4349-4357, 2014.

SERASA. **O que é Serasa Score 2.0?** Disponível em: <https://www.serasa.com.br/score/score-2-0/>. Acesso em: 03 set. 2021.

SHAVELL, Steven. **Economic analysis of accident law**. Cambridge: Harvard University Press, 2009. 352 p.

SINGAPORE. **Artificial Intelligence Governance Framework**. 2020. Disponível em: <https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/smodelaigovframework2.pdf>. Acesso em: 30 jan. 2022.

SOUZA, Celina. **Políticas públicas: conceitos, tipologias e subáreas**. Trabalho elaborado para a Fundação Luís Eduardo Magalhães. São Paulo, 2002.

SWEDEN. **Nationell inriktning för artificiell intelligens**. 2018. Disponível em: <https://www.regeringen.se/informationsmaterial/2018/05/nationell-inriktning-for-artificiell-intelligens/>. Acesso em: 30 jan. 2022.

TAIWAN. **AI Taiwan Action Plan**. 2019. Disponível em: <https://english.ey.gov.tw/News3/9E5540D592A5FECD/1dec0902-e02a-49c6-870d-e77208481667#:~:text=Capitalizing%20on%20this%20wave%2C%20the,greater%20momentum%20into%20Taiwan's%20industries>. Acesso em: 30 jan. 2022.

UNESCO. **UNESCO member states adopt the first ever global agreement on the Ethics of Artificial Intelligence**. 2021. Disponível em: <https://en.unesco.org/news/unesco-member-states-adopt-first-ever-global-agreement-ethics-artificial-intelligence>. Acesso em: 30 jan. 2022.

UNIÃO EUROPEIA. **Artificial Intelligence: Europe needs to take a human-in-command approach, says EESC**. 2017. Disponível em: <https://www.eesc.europa.eu/en/news-media/press-releases/artificial-intelligence-europe-needs-take-human-command-approach-says-eesc>. Acesso em: 11 abr. 2020.

UNITED STATES. **Artificial Intelligence Index Report 2021**. Stanford University. 2021. Disponível em: https://aiindex.stanford.edu/wp-content/uploads/2021/11/2021-AI-Index-Report_Master.pdf. Acesso em: 30 jan. 2022.

UNITED KINGDOM. **AI Sector Deal**. 2019. Disponível em: <https://www.gov.uk/government/publications/artificial-intelligence-sector-deal/ai-sector-deal>. Acesso em: 30 jan. 2022.

VINCENT, J. Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day. **The Verge**. 2016. Disponível em: <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>. Acesso em: 20 dez. 2021.

WANG, Ge. **Humans in the Loop: The Design of Interactive AI Systems**. Human-centered Artificial Intelligence. Stanford University. 2019.

WIPO. **Índice Global de Inovação**. 2021. Disponível em: https://www.wipo.int/edocs/pubdocs/pt/wipo_pub_gii_2021_exec.pdf. Acesso em: 30 jan. 2022

Artigo recebido em: 2022-01-31

Artigo aceito para publicação em: 2022-03-28