XIV ENCONTRO INTERNACIONAL DO CONPEDI BARCELOS - PORTUGAL

GOVERNO DIGITAL, DIREITO E NOVAS TECNOLOGIAS

G326

Governo digital, direito e novas tecnologias [Recurso eletrônico on-line] organização CONPEDI

Coordenadores: Ana Catarina Almeida Loureiro; Danielle Jacon Ayres Pinto; José Renato Gaziero Cella. – Barcelos, CONPEDI, 2025.

Inclui bibliografia
ISBN: 978-65-5274-207-0
Modo de acesso: www.conpedi.org.br em publicações
Tema: Direito 3D Law

1. Direito – Estudo e ensino (Pós-graduação) – Encontros Internacionais. 2. Governo digital. 3. Direito e novas tecnologias. XIV Encontro Internacional do CONPEDI (3; 2025; Barcelos, Portugal).

CDU: 34

CONPEDI

Conselho Nacional de Pesquisa
e Pós-Graduação em Direito Florianópolis
Santa Catarina – Brasil
www.conpedi.org.br

XIV ENCONTRO INTERNACIONAL DO CONPEDI BARCELOS - PORTUGAL

GOVERNO DIGITAL, DIREITO E NOVAS TECNOLOGIAS

Apresentação

No XIV Encontro Internacional do CONPEDI, realizado nos dias 10, 11 e 12 de setembro de 2025, o Grupo de Trabalho - GT “Governo Digital, Direito e Novas Tecnologias”, que teve lugar na tarde de 12 de setembro de 2025, destacou-se no evento não apenas pela qualidade dos trabalhos apresentados, mas pelos autores dos artigos, que são professores pesquisadores acompanhados de seus alunos pós-graduandos. Foram apresentados artigos objeto de um intenso debate presidido pelos coordenadores e acompanhado pela participação instigante do público presente no Instituto Politécnico do Cávado e do Ave - IPCA, em Barcelos, Portugal.

Esse fato demonstra a inquietude que os temas debatidos despertam na seara jurídica. Cientes desse fato, os programas de pós-graduação em direito empreendem um diálogo que suscita a interdisciplinaridade na pesquisa e se propõe a enfrentar os desafios que as novas tecnologias impõem ao direito. Para apresentar e discutir os trabalhos produzidos sob essa perspectiva.

Os artigos que ora são apresentados ao público têm a finalidade de fomentar a pesquisa e fortalecer o diálogo interdisciplinar em torno do tema “Governo Digital, Direito e Novas Tecnologias”. Trazem consigo, ainda, a expectativa de contribuir para os avanços do estudo desse tema no âmbito da pós-graduação em direito, apresentando respostas para uma realidade que se mostra em constante transformação.

Os Coordenadores

Prof. Dr. José Renato Gaziero Cella

MODELOS DE REGULAÇÃO DE SISTEMAS DE INTELIGÊNCIA ARTIFICIAL NA ERA DE NÃO-COISAS

REGULATION MODELS FOR ARTIFICIAL INTELLIGENCE SYSTEMS IN THE ERA OF NON-THINGS

Cinthia Obladen de Almendra Freitas [1]

Resumo

O estudo sobre a regulação de sistemas de Inteligência Artificial (IA) tem ganhado destaque nas discussões acadêmicas e legislativas. As principais questões levantadas são: o que exatamente se busca regular? A ciência da IA, os sistemas desenvolvidos ou os riscos decorrentes de seu uso? Diante disso, o artigo propõe uma reflexão sobre os modelos mentais de regulação aplicáveis aos sistemas de IA, especialmente à luz de propostas como o AI Act, da União Europeia, e o Projeto de Lei nº 2.338/2023, no Brasil, que adotam modelos de regulação baseados em risco, buscando mitigar os impactos de tais sistemas na sociedade. Com método dedutivo e revisão bibliográfica especializada, o texto parte do entendimento de que os sistemas de IA são compostos essencialmente por algoritmos que operam sobre dados, portanto, não-coisa a partir da filosofia de Byung-Chul Han. Assim, propõe-se uma concepção tecno-jurídica da regulação, adequada a uma realidade digital que é imaterial e intangível, na qual os algoritmos permeiam as relações sociais. O artigo contribui para a compreensão dos fundamentos e direções possíveis para regulamentar sistemas de IA, respeitando a complexidade tecnológica e os impactos sociais envolvidos.

Palavras-chave: Direito e tecnologia, Sociedades, Inteligência artificial, Modelos de regulação, Não-coisas

Abstract/Resumen/Résumé

The study of the regulation of Artificial Intelligence (AI) systems has gained prominence in academic and legislative discussions. The main questions raised are: what exactly is to be regulated? The science of AI, the developed systems, or the risks arising from their use? In

possible directions for regulating AI systems, taking into account their technological complexity and the social impacts they entail.

Keywords/Palabras-claves/Mots-clés: Law and technology, Societies, Artificial intelligence, Regulation models, Non-things

1 INTRODUÇÃO

Estudar, compreender e refletir sobre regulação de sistemas de Inteligência Artificial (IA) tem ocupado o espaço de discussões. Então, pergunta-se: O que se deseja regular? A IA como ciência? Os sistemas de IA? Os riscos decorrentes da aplicação da IA? Como construir regulamentos ou legislações que norteiem o desenvolvimento e o uso de sistemas de IA? São muitas as perguntas e busca-se contribuir com explicações sobre os modelos mentais de regulação que podem ser aplicados no contexto dos sistemas de IA, bem como com alguns pontos de atenção.

Primeiramente, antes de discutir sobre modelos de regulação de sistemas de IA, a exemplo do *AI Act*, na União Europeia, e o Projeto de Lei No. 2.338 de 2023, em trâmite no Congresso Nacional brasileiro, é necessário compreender que existem diversos modelos mentais para se buscar uma regulação, quando o tema envolve algoritmos, métodos e técnicas dos ramos da IA, sem esquecer os impactos sobre aqueles que poderão ser beneficiados ou não por estes sistemas. Portanto, tem-se por objetivo contribuir para a compreensão dos motivos pelos quais as regulações de sistemas de IA tem adotado os modelos de regulamentação com base em riscos (*Risk Regulation*) visando, assim, mitigar os impactos da aplicação das tecnologias baseadas em IA.

Para realização da pesquisa foi aplicado o método dedutivo. O artigo visa esclarecer o tema de modelos de regulação de sistemas de Inteligência Artificial trazendo conceitos e fundamentos necessários ao tema. Partindo-se da premissa de que a IA é composta por algoritmos que atuam sobre dados, assume-se a IA como não-coisa e discute-se sobre os modelos mentais de regulação de sistemas de IA, diferenciando-os e buscando compreender, particularmente, os modelos de regulamentação com base em riscos. O artigo possui uma concepção tecno-jurídica a partir de revisão bibliográfica especializada. Assim, para que o Direito possa usufruir das regulamentações já existentes, bem como propor novas regulamentações, é necessário tratar o tema diante de uma realidade que não mais se apresenta material e tangível nem pode ser arraigada na posse e na coisa física, visto que se vive uma sociedade permeada por algoritmos na Era de Não-Coisas (Han, 2022). O artigo é resultado de projeto de pesquisa aprovado na Chamada Universal 10/2023 do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

2 A ERA DE NÃO-COISAS

A discussão sobre não-coisas foi apresentada por Freitas (2024) e encontra suporte na filosofia de Byung-Chul Han (2022). Byung-Chul Han (2022) analisa não-coisas sob os

seguintes elementos, a saber: posse, *smartphone*, *selfies* e Inteligência Artificial (IA), passando pelo estudo da coisa à não-coisa, agora centrada na informação obtida a partir do processamento de dados digitais e nos atores que processam dados e informações, vigiando e controlando cada um dos seres humanos da infoesfera (Floridi, 2014).

Byung-Chul Han (2022, p. 9) preconize que a "ordem terrena está a ser substituída pela ordem digital. Esta desreifica o mundo, ao mesmo tempo que o informatiza." Para o autor, as não-coisas "penetram de todos os lados no meio que nos rodeia e ocupam o lugar das coisas". De modo que "nos encontramos na transição da era das coisas para a era das não-coisas." E não-coisas se tornaram sinônimo dos dados e informações que determinam a vida no mundo digital ou a *onlife* na infosfera (Floridi, 2014, p. 40-41).

Yuval Noah Harari (2016, p. 370-399), em seu livro "Homo Deus: uma breve história do amanhã", discute a religião dos dados, ou dataísmo (tradução para o português de *dataism*, sendo o termo *data* mantido em inglês para o termo que se refere aos dados), pelo qual "o Universo consiste num fluxo de dados e o valor de qualquer fenômeno ou entidade é determinado por uma contribuição ao processamento de dados." E, o autor aponta que "o supremo valor dessa religião é o fluxo de informação". E, Harari, ousa deixar os leitores com três questionamentos: (i) Será que organismos são apenas algoritmos e a vida apenas processamento de dados?; (ii) O que é mais valioso – inteligência ou consciência?; (iii) O que vai acontecer à sociedade, aos políticos e à vida cotidiana quando algoritmos não conscientes mas altamente inteligentes nos conhecerem melhor do que nós nos conhecemos? São muitas as perguntas que precisam de respostas.

Todo esse cenário conta ainda com os estudos de Shoshana Zuboff (2021) sobre o que ela nomeia como capitalismo de vigilância, tendo como base uma arquitetura global de modificação comportamental que altera e desfigura o mundo físico, visto que a nova configuração global tem por fundamento uma estruturação digital por meio dos dados que os usuários "deixam" na Internet e são tratados sem o devido consentimento. A autora exemplifica e discute fortemente o tratamento de dados frente a ausência ou fraco regramento de proteção de dados pessoais e ao desconhecimento tanto da coleta e tratamento de dados quanto dos reflexos na ordem social e futuro digital da sociedade contemporânea.

E Han (2022, p. 45-50) estabelece um paralelo entre o pensamento humano e a Inteligência Artificial para apresentar a IA como não-coisa. Tal paralelo parte da premissa que o pensamento é composto por uma "totalidade, que precede os conceitos, as ideias e a informação", de modo a se encontrar "numa disposição afetiva básica" (Han, 2022, p. 45), a qual se coloca para fora do ser humano. Ou seja, o pensamento pode ser externado, para fora

de si. Já a IA, ao realizar cálculos, "nunca está fora de si mesma" (HAN, 2022, p. 46). Os cálculos não envolvem emoção e estar fora de si mesma é emoção. Falta à IA o espírito. Portanto, entende-se que IA é existência em bits, mas não é essência.

Para Han (2022, p. 47) "o pensamento ouve, ou melhor, escuta e presta toda a atenção. A Inteligência Artificial é surda." Tudo isso compõe a dimensão analógica do pensamento humano, a qual não se pode reproduzir por meio da IA (Han, 2022, p. 47). E, Han (2022, p. 48) corrobora essa explicação, afirmando que "A Inteligência Artificial é apática, ..., sem paixão. Só calcula." Para os seres humanos, existem fatos, contextos e mudanças. Para a IA existem dados, que podem ser revistos e modificados para efetuar novos cálculos e obter novos resultados, até mesmo retroalimentar os dados já existentes em uma base de dados. Mas tudo isto depende de um algoritmo que precisará ser planejado para tal tarefa. O ser humano não necessita de algoritmo, ele mesmo lhe dá novos fatos e contextos, gerando mudanças e entendimentos. Han (2022, p. 49) explica que o contexto é compreendido pelo ser humano, por meio do pensamento. Para a IA, a relação entre A e B, expressa por C, somente é compreendida se o algoritmo assim estiver previsto para relacionar A e B e encontrar C. Por exemplo, se A representa o valor 10 e B representa o valor 20, C poderá ser positivo ou negativo a depender da relação matemática que se deseja analisar, ou seja, se A é maior, menor ou igual a B. Portanto, o contexto para os seres humanos é relacional e para a IA é um o resultado de uma análise matemática lógica.

Outra característica dos seres humanos e a capacidade, por meio da inteligência, de realizar escolhas, sendo que a IA "só opta por uma escolha entre opções previamente dadas", sendo o resultado positivo (1) ou negativo (0) ou, ainda, uma probabilidade (um valor numérico entre 0 e 1 ou 0 e 100%). O pensamento é não determinístico e infinito, enquanto um algoritmo que opera com dados é determinístico e finito. Assim IA não é baseada em pensamentos, mas em algoritmos que representam métodos e técnicas de diferentes ramos de aplicação, por exemplo, visão computacional (*computer vision*), processamento de linguagem natural (*natural language processing*) ou robótica (*robotics*).

De um modo geral, os métodos e técnicas baseados em IA, funcionam tendo como *input* uma base de dados que reflete situações e cenários do passado, de modo que, por exemplo, tal qual como explicado por Freitas e Barddal (2019, p. 110) "a análise preditiva é uma abordagem popular para obter informações e padrões sobre os dados e criar modelos preditivos. A análise preditiva visa aproveitar os dados do passado para obter informações em tempo real e prever eventos futuros". Continuam os autores explicando que "Na prática, a análise preditiva está na interseção entre a estatística, matemática e ciência da computação, que, em sua influência, pode

ser aplicada para obter *insights* e ganhos em diferentes aplicações." Han (2022, p. 50) contradiz a questão da IA prever eventos futuros, visto que "o futuro que calcula não é um futuro no sentido próprio do termo." Entende-se que para Han o futuro calculado pela IA não tem contexto, semântica ou significado, sendo somente uma representação numérica (matemática, lógica, estatística ou probabilística). O contexto, a semântica ou o significado serão postos pelos seres humanos.

Por isso, Han (2022, p. 51) afirma que "A informação e os dados não tem profundidade", visto que contexto, semântica ou significado surgem do pensamento humano que "é mais do que cálculos e resolução de problemas". "Os dados e a informação não seduzem" (Han, 2022, p. 51). O que seduz é o pensamento do ser humano ao envolver-se com o resultado (*output*) fornecido pela IA. Não seduzem, uma vez que para a IA são dados, *input* para os algoritmos, portanto, representação matemática em bits de coisas. Só o pensamento vê, escuta ou entende a coisa representada. A coisa torna-se não-coisa por meio dos dados e algoritmos. E são não-coisas que passaram a influenciar a sociedade contemporânea e estabelecer a sociedade de algoritmos na Era de Não-Coisas de Han.

3 CONTEXTUALIZANDO A REGULAÇÃO DE SISTEMAS DE IA

Inicialmente, deve-se ter por premissa que a regulação, de modo geral, se refere aos sistemas de IA e não à IA como ciência e área do conhecimento humano. Justifica-se essa diferença diante das tentativas de parar a IA, a exemplo da Carta Aberta[1] que conclamou a todos os laboratórios de IA que pausassem imediatamente, por pelo menos 6 meses, a partir de 22 de março de 2023, interrompendo o treinamento de sistemas de IA mais poderosos do que o GPT-4. A carta previa uma pausa pública e verificável, de modo a incluir todos os principais atores do desenvolvimento da IA em nível mundial. E, se tal pausa não pudesse ser promulgada rapidamente, os governos deveriam intervir e instituir uma moratória. Isso não ocorreu, nem a paralisação e nem a moratória. Mas é importante observar os *Asimolar AI Principles*[2], desenvolvidos em 2017 a partir da premissa de que o desenvolvimento contínuo, mas guiado por tais princípios, oferecerá oportunidades para auxiliar e capacitar as pessoas para o uso de sistemas de IA. Dentre os princípios encontra-se a indicação de aspectos éticos e valores, tais como: segurança, transparência, responsabilidade, valores humanos, privacidade, liberdade, prosperidade, controle humano, não subversão e não à armas autônomas letais. A atenção é

[1] Disponível em: https://futureoflife.org/open-letter/pause-giant-ai-experiments/
[2] Disponível em: https://futureoflife.org/open-letter/ai-principles/

menor aos princípios, visto que até o momento (abril/2025) havia 5.720 assinaturas, enquanto a Carta Aberta possui um total de 33.705 assinaturas, o que demonstra sua importância.

A referida Carta aberta questiona: "*Should we let machines flood our information channels with propaganda and untruth? Should we automate away all the jobs, including the fulfilling ones? Should we develop nonhuman minds that might eventually outnumber, outsmart, obsolete and replace us? Should we risk loss of control of our civilization?*"[3]. São perguntas relevantes e a resposta, de modo geral, está no próprio texto:

> Such decisions must not be delegated to unelected tech leaders. Powerful AI systems should be developed only once we are confident that their effects will be positive and their risks will be manageable. This confidence must be well justified and increase with the magnitude of a system's potential effects.[4]

A Carta considera que há um momento, um ponto específico na linha do desenvolvimento, que é necessário olhar para uma revisão independente antes de começar a treinar sistemas para uso no futuro, de modo que a pausa proposta poderia ser utilizada para desenvolver e implementar conjuntamente um arcabouço de protocolos de segurança compartilhados tanto para projeto (*design*) quanto para o próprio desenvolvimento de sistemas de IA, os quais necessitam ser auditáveis e supervisionados por especialistas externos independentes.

A Carta explica que a pausa não se refere ao desenvolvimento de sistemas IA em geral, mas "*merely a stepping back from the dangerous race to ever-larger unpredictable black-box models with emergent capabilities*."[5]

Além disto, qualquer que seja a regulação voltada aos sistemas de IA, concorda-se que tal regulação deve ser sustentada por princípios advindos de valores éticos. Para além, deve-se também considerar o que os técnicos e especialistas em Ciência da Computação entendem e preconizam sobre regulação de sistemas de IA, visto que Direito e Tecnologia precisam andar

[3] Tradução livre: Deveríamos deixar que as máquinas inundassem nossos canais de informação com propaganda e inverdades? Deveríamos automatizar todos os empregos, inclusive os gratificantes? Deveríamos desenvolver mentes não humanas que pudessem acabar nos substituindo, nos tornando obsoletos e nos superando em números e inteligência? Deveríamos arriscar a perda do controle da nossa civilização?

[4] Tradução livre: Tais decisões não devem ser delegadas a líderes tecnológicos não eleitos. Sistemas de IA poderosos devem ser desenvolvidos somente quando estivermos confiantes que seus efeitos serão positivos e seus riscos serão administráveis. Essa confiança deve ser bem justificada e aumentar com a grandeza dos efeitos potenciais de um sistema.

[5] Tradução livre: Apenas um passo a trás na corrida perigosa por modelos de caixa-preta cada vez maiores e imprevisíveis com capacidades emergentes.

de mãos dadas neste terreno que, aparentemente, é pantanoso e desconhecido por muitos especialistas fora da área da Ciência da Computação.

No Brasil, o destaque é para o "Plano de Inteligência Artificial" da Sociedade Brasileira de Computação (SBC) que apresenta um levantamento do cenário atual e um plano estratégico sobre a implementação e o uso de sistemas de IA, contendo ações com metas de curto, médio e longo prazo, e sua adoção requer uma diretriz pública incluindo um projeto para implantação e monitoramento contínuo (SBC, 2024). O objetivo do Plano é propor um "conjunto de ações para que a população brasileira desenvolva as competências e obtenha o conhecimento e informações essenciais para utilizar e desenvolver eficazmente as tecnologias de IA." (SBC, 2024, p. 6). O documento foi elaborado por uma Comissão composta por um grupo de especialistas altamente qualificados e experientes na área de IA, com competências que abrangem tanto conhecimento teórico quanto prático em diferentes contextos de aplicação.

O "Plano de Inteligência Artificial" da SBC (2024, p. 11) é composto pelos seguintes tópicos, todos apresentando um diagnóstico, uma proposta e um planejamento, a saber: a) formação de recursos humanos; b) pesquisa, desenvolvimento tecnológico e inovação; c) empresas brasileiras; d) IA ética e socialmente responsável; e) ecossistema de IA na Computação. Particularmente, o tópico sobre "IA ética e socialmente responsável" demonstra que o diagnóstico reconhece que "As soluções de IA oferecem grandes oportunidades e benefícios, mas também apresentam possíveis consequências negativas, exacerbando frequentemente a desigualdade econômica e social, e impactando de forma desproporcional diferentes comunidades". E os cientistas da computação apontam como importante a discussão sobre questões relacionadas à "justiça, transparência, responsabilização, diversidade, privacidade de dados e proteção da propriedade intelectual." Especificamente, destacando "a necessidade de garantir a adequação dos dados usados para treinamento e desenvolvimento das soluções de IA à realidade brasileira, e a transparência dessas soluções, principalmente em aplicações críticas." A preocupação remete ainda à "segurança e proteção aos dados" com base na Lei Geral de Proteção de Dados Pessoais (LGPD) (Brasil, 2018). Em termos de planejamento a SBC (2024, p. 13) aponta:

> Curto prazo: Aprovação de uma regulamentação mínima da IA, onde seja possível inovar com segurança jurídica. Construção dos vários instrumentos de letramento e formação.
>
> Médio prazo: Desenvolvimento de um marco regulatório robusto. Aperfeiçoamento dos arcabouços técnicos para construção responsável de sistemas. Validação dos instrumentos de letramento e formação.

> Longo prazo: Letramento da população nos benefícios e riscos da IA. Mudança de paradigma em termos de produtos e serviços.

E, ainda, recomenda como indicadores: "Produtos e serviços éticos, socialmente responsáveis e em conformidade com a regulamentação. Pessoas formadas com uma perspectiva ética e socialmente responsável. Processos, mecanismos e tecnologias para construção responsável de sistemas de IA." (SBC, 2024, p. 13).

Seguindo pelo caminho dos aspectos tecnológicos da IA, a Associação Brasileira de Normas Técnicas (ABNT) vem trabalhando para trazer ao Brasil o que há de mais recente em normatização de sistemas de IA. Confirma-se um movimento mundial em torno do tema de Inteligência Artificial e, com certeza, esta movimentação não será somente mais uma bolha de sabão que em breve será estourada. IA veio para ficar e modificar relações entre humanos, entre humanos e máquinas e, exercitando o futuro, entre máquinas. Na Era de Não-Coisas, a Inteligência Artificial é o ponto central das discussões, o ponto de inflexão que permitirá muito aprendizagem seja por humanos ou algoritmos.

4 MODELOS MENTAIS PARA REGULAÇÃO DE SISTEMAS DE IA

Não há como adentrar ao tema de regulação sem antes compreender os modelos possíveis de aplicação. Petit e De Cooman (2020) apresentam que existem quatro modelos mentais para se pensar a regulação de sistemas de IA, a saber: (i) *Black Letter Law*; (ii) *Emergent*; (iii) *Ethical* e (iv) *Risk Regulation*.

Petit e De Cooman (2020, p. 2) explicam que as regulações que utilizam o modelo *Black Letter Law* fazem uso de leis rígidas ou regras jurídicas bem estabelecidas ou jurisprudência consolidada ou pacificada. Busca-se encontrar como a legislação existente se aplica aos sistemas de IA, mas levando em consideração as restrições impulsionadas por estruturas legais compostas por normas vinculativas. Assim as regulações podem tanto ser elaboradas a partir de questões específicas, por exemplo: segurança cibernética e responsabilidade, proteção ao consumidor, propriedade intelectual, privacidade, responsabilidade civil, responsabilidade criminal, personalidade jurídica, entre outros. Mas podem seguir uma abordagem mais ampla baseada em direitos humanos ou direitos fundamentais, direitos estes que estão se tornando o foco principal das discussões sobre regulamentação de sistemas de IA.

O modelo *Emergent* preocupa-se em verificar se os sistemas de IA levantam novas questões que exigem a criação de "um novo ramo do direito" (Petit; De Cooman, 2020, p. 3-4), partindo da premissa que tais sistemas produzem fenômenos emergentes, questionando

aspectos econômicos, éticos e científicos, de modo que exijam proibições legais específicas ou não. Alguns exemplos podem ser: "A lei dos carros autônomos", "A lei dos drones" ou "A lei dos robôs" (Petit; De Cooman, 2020, p. 4). Esse tipo de modelo pode ser enviesado a favor ou contra sistemas de IA, por exemplo, antropomórficos (robô Asimo[6][7] - fabricado pela Honda) ou simbólicos (Siri ou Alexa), uma vez que os autores afirmam que "*Studies have shown that individuals treat computers like they behave with other human beings*."[8] (Petit; De Cooman, 2020, p. 5).

Os modelos baseados em aspectos éticos (*Ethical*) para sistemas de IA, de acordo com Petit e De Cooman (2020, p. 5-7) afloram um campo da ética conhecido como Ética Normativa, estabelecendo normas morais que distingam o certo do errado, o bom do mau. Os modelos podem lançar mão também da Ética Aplicada, a qual analisa problemas morais específico, por exemplo, aborto, eutanásia e aplicações específicas de IA a exemplo do *social score* ou do reconhecimento facial. No que concerne ao reconhecimento facial, ressalta-se que o tema é preocupação alinhada com as prioridades do G20 (*Sherpa Track's Digital Economy Working Group Agenda*), de modo que Caparroz *et al.* (2024) apresentam e discutem as seguintes preocupações sobre o uso de sistemas de reconhecimento facial em espaços públicos: discriminação e preconceito, falta de transparência e responsabilização, violações de privacidade, proteção de dados e segurança cibernética. Os autores justificam que o G20 procura estabelecer discussões para "investigar as implicações éticas da IA e suas tecnologias associadas de aprendizagem de máquina e aprendizagem profunda, desenvolver uma estrutura modelo para o uso do reconhecimento facial em espaços públicos, delinear princípios comuns e padrões mínimos para orientar a legislação nacional e lançar uma iniciativa de governança de dados para promover a harmonização dos padrões de proteção de dados." Além disto, os autores alertam que o desenvolvimento responsável de sistemas desta natureza demanda uma abordagem holística priorizando os direitos humanos e aspectos éticos e democráticos.

Interessante observar que já existem estudos que buscam compreender como as pessoas confiam em sistemas de IA e, ao mesmo tempo, atribuem responsabilidades por danos causados por estes sistemas por meio da análise de aspectos morais a exemplo de: agir certo ou errado – *moral agency* e ser tratado corretamente ou não – *moral patiency* (Ladak *et al.*, 2025).

[6] Disponível em: https://robotsguide.com/robots/asimo
[7] Disponível em: https://www.repicture.com/project/asimo-hondas-autonomous-humanoid-robot
[8] Tradução livre: Estudos mostraram que indivíduos tratam computadores da mesma maneira que tratam outros seres humanos. Recomenda-se: https://scitechdaily.com/hey-alexa-are-you-trustworthy-human-like-social-behaviors-improve-trust-in-digital-assistants/

Visando contribuir com o entendimento sobre a identificação e a análise de implicações éticas em algoritmos, Tsamados *et al.* (2022, p. 226) realizaram estudo para fornecer uma análise atualizada das preocupações epistêmicas e normativas, de modo a oferecer orientações práticas para a governança do *design*, desenvolvimento e implantação de algoritmos. Os autores concluem que o debate crescente sobre princípios e critérios éticos deve nortear o projeto e a governança de algoritmos e, também, das tecnologias digitais, com o propósito explícito de fomentar o bem social. Apontam que "Análises éticas são necessárias para mitigar os riscos e, ao mesmo tempo, aproveitar o potencial benéfico destas tecnologias, na medida em que atendem a dois objetivos: esclarecer a natureza dos riscos éticos e do potencial benéfico dos algoritmos e das tecnologias digitais, e traduzir este entendimento em orientações sólidas e acionáveis para a governança do *design* e do uso de artefatos digitais."

Ao se considerar modelos baseados em aspectos éticos há que se mencionar os valores e princípios que nortearam o trabalho desenvolvido pela UNESCO (2022) com o objetivo de apontar recomendações ao desenvolvimento de sistemas de IA. Os valores incluem: respeito, proteção e promoção dos direitos humanos e das liberdades fundamentais e da dignidade humana; o surgimento de um ecossistema de IA, diversidade e inclusão; paz, justiça e sociedades interconectadas. Como princípios são listados: proporcionalidade e não geração de danos; segurança e proteção; justiça e não discriminação; sustentabilidade; direito à privacidade e proteção de dados; supervisão humana; autodeterminação; transparência e explicabilidade; responsabilidade e prestação de contas; conscientização e letramento digital; colaboração e governaça adaptativa e multissetorial.

Atualmente, menciona-se a *Algorethics*, ou seja, a ética dos algoritmos, sendo que já é entendida como um ramo da ética que estuda o *design*, implementação e uso de algoritmos (Wagle, 2024). Para Benanti (2023) o termo vem sendo desenvolvido desde 2018 para denotar a necessidade de um estudo dedicado à avaliação das implicações éticas das tecnologias, particularmente de sistemas de IA, discutindo questões como transparência e responsabilização de algoritmos, vieses e responsabilização. A *Algorethics* busca a relação entre tecnologia, sociedade e indivíduos, orientando cientistas de dados e pesquisadores a construir de maneira ética os sistemas de IA para benefício de toda a sociedade. Mökander e Floridi (2022) discutem a denominada *ethics-based auditing* como um mecanismo de governança que pode preencher a lacuna entre os princípios e a prática ética em sistemas de IA.

E, assim, chega-se aos modelos *Risk Regulation* ou modelos de regulamentação com base em riscos que possuem por objetivo reduzir a probabilidade de ocorrência ou os níveis de danos decorrentes de eventos inerentes à tecnologia (Petit; De Cooman, 2020, p. 7-8). Os

autores mencionam o "*White Paper on Artificial Intelligence: a European approach to excellence and trust*", da Comissão Europeia que considera que "O quadro regulamentar deverá incidir na forma de minimizar os vários riscos de potenciais danos, em especial os mais significativos." (Comissão Europeia, 2020, p.11). E, vai além apontando sete princípios norteadores: iniciativa e controlo por humanos; robustez e segurança; privacidade e governança de dados; transparência, diversidade, não discriminação e equidade; bem-estar social e ambiental; responsabilização. O objetivo de uma regulamentação ou lei que faz uso do modelo baseado em riscos visa prevenir e não corrigir. Além disto, cabe parte relevante à avaliação de riscos, a qual necessitará de trabalhos estatísticos ou aplicação de técnicas e métodos específicos.

E, ainda, há que se compreender que modelos de regulamentação baseados em riscos usam de uma estrutura hierárquica para distinguir diferentes níveis de riscos entre os sistemas de IA, aproximando-se do consequencialismo e da análise de custo-benefício, ou seja, quanto maior o risco mais forte será a resposta regulatória (Petit; De Cooman, 2020, p. 8).

E, por outro lado, há que se considerar como premissa que é impossível reduzir riscos à zero, sempre haverá o risco de um sistema de IA cometer erros, independentemente da área de aplicação, portanto, há que se mitigar os riscos jurídicos e tecnológicos decorrentes do meio ambiente digital (Cavedon et al., 2015) para se alcançar a proteção de direitos fundamentais e garantir liberdades aos indivíduos. Eis aqui a função da análise de risco, também chamada, às vezes, de gerenciamento de risco. Espelhando a dimensão dupla do risco, a análise de risco é composta de duas etapas (Gellert, 2017, p. 2): (i) avaliação de risco: medir o nível de risco em termos de probabilidade e gravidade; (ii) gerenciamento de risco: decidir se deve ou não assumir o risco (tomada de decisão). E, a decisão em nível de gerenciamento de risco é geralmente acompanhada de medidas que visam reduzir o nível de risco (Gellert, 2017, p. 2). E existem muitos riscos jurídicos e tecnológicos relacionados aos sistemas de IA. Gellert (2017, p. 2) alerta que as medidas aplicadas na redução de riscos podem ser referidas por vezes como: redução de risco, controle de risco, resposta a risco ou, mais genericamente, como medidas de mitigação de risco, termo comumente utilizado na área jurídica. E estar em conformidade com as legislações ou regramentos nacionais ou internacionais, é mitigar riscos. Mas há um ponto crucial nessa discussão que é determinar se o nível de risco é suficientemente baixo para que possa ser tomado (aceito).

Desta forma, sobre avaliação de riscos, deve-se ter em mente que ela é composta por etapas, a saber (Gellert, 2017, p. 3): (i) critérios de risco, (ii) identificação dos riscos e (iii) a avaliação de risco propriamente dita (ISO, 2009). Por isso, Petit e De Cooman (2020, p. 7-8)

explicam que existe dependência entre a regulação de risco e a sua medição, estimativa, valoração propriamente dita. Haverá situações em que a determinação de um valor será impossível devido à incerteza científica e, assim, o princípio da precaução deverá prevalecer: "*Absence of evidence does not mean evidence of absence*."[9]

Diante de sistemas de IA, o princípio da precaução leva os legisladores a estabelecer linhas vermelhas, proibições ou moratórias em aplicações como armas autônomas letais, *social score* ou reconhecimento facial. "*Science is not irrelevant in the precautionary approach*."[10] (Petit; De Cooman, 2020, p. 8).

Entende-se que ao mencionar que a Ciência não é irrelevante na abordagem de precaução, Petit e De Cooman (2020) reconhecem que, embora esta abordagem priorize a prevenção de danos em situações de incerteza, as evidências científicas ainda desempenham um papel fundamental frente à regulação de sistemas de IA. A precaução se aplica em contextos para os quais os riscos potenciais são desconhecidos ou incertos, tal qual os riscos ora enfrentados em sistemas de IA. Todavia, a Ciência continua sendo importante para identificar possíveis riscos, avaliar probabilidades e orientar a tomada de decisões informadas sobre que riscos deseja-se aceitar ou mitigar. A diferença é que, mesmo diante da ausência de certezas científicas plenas, as medidas de precaução podem ser justificadas para evitar consequências graves ou irreversíveis. Portanto, a Ciência fornece a base para entender os riscos, mas a abordagem de precaução complementa isto com uma postura de responsabilidade diante da incerteza.

Petit e De Cooman (2020) também relacionam a análise de riscos, o nexo causal e a abordagem de precaução para apontar maneiras de gerenciar riscos, seja em termos de custos e benefícios ou de conformidade com legislações ou regramentos de uso de sistemas de IA, como já mencionado anteriormente. Há que se ter em mente que qualquer que seja a decisão tomada ela envolverá incertezas. Isto devido ao entendimento de que o princípio da precaução é também um dever moral de garantir que tudo seja feito para evitar danos catastróficos. Assim, se a análise de riscos é método que permite estimar a probabilidade de um risco se materializar e causar danos, há que se estabelecer uma estrutura baseada na gravidade dos impactos e, para tal, necessita-se compreender a relação entre causa e efeito dos eventos estatísticos, ou seja, conhecer o nexo causal. Porém, essa relação não é de simples constatação, especialmente ao tratar de riscos relacionados a eventos complexos ou incertezas estatísticas. E, considerando-se que os sistemas de IA são compostos tanto por eventos complexos quanto por incertezas

[9] Tradução livre: Ausência de evidências não significa evidência de ausência.
[10] Tradução livre: A ciência não é irrelevante na abordagem de precaução - preventiva.

estatísticas, tem-se uma justificativa forte à explicabilidade da IA e o princípio da precaução será o fundamento dos modelos de regulamentação baseados em riscos, para que sistemas de IA previnam efeitos danosos graves ou irreversíveis.

Mas o que torna um risco aceitável ou inaceitável em sistemas de IA? Koessler *et al.* (2024) contribuem com esta discussão ao buscar explicação sobre abordagens que definem limites de risco para, então, declarar quanto risco seria aceitável ou não. Tais abordagens podem, por exemplo, observar os limites das capacidades dos métodos e técnicas de IA estabelecendo limiares, do inglês *risk thresholds*. Por exemplo, em um sistema de verificação para busca de indivíduos procurados pela polícia em locais públicos, sendo tal sistema baseado em reconhecimento facial, se aceita como "suspeito localizado" todas as faces que estiverem acima do limiar (*thresholds*) de X% de similaridade entre a face "procurada" e as faces de quem está naquele local num determinado momento, fazendo com que todos os indivíduos apontados pelo sistema sejam, por exemplo, abordados pela polícia para identificação (por meio de documentação de identificação). Os autores explicam que, por um lado, a principal vantagem dos limites de risco (*risk thresholds*) é que eles são baseados em princípios, contrariamente aos limites baseados em capacidades (*capability thresholds*). Por outro lado, a principal desvantagem é que os limites de risco (*risk thresholds*) são mais difíceis de avaliar de forma confiável.

Koessler *et al.* (2024, p. 6-10) apontam a existência de três maneiras para observar e definir limites (*thresholds*) para riscos em sistemas de IA:

> a) *Compute thresholds*: são definidos em termos de recursos computacionais usados para treinar um modelo, visto que as taxas de acerto e erro serão decorrentes tanto da técnica empregada para o treinamento quanto da base de dados utilizada no treinamento do modelo. Esse tipo de limiar pode ser interessante para observar o procedimento de treinamento do modelo, visto que pode ser facilmente medido e previsto no início do processo de desenvolvimento. Os autores afirmam que "*Compute thresholds should thus be used as an initial filter to identify models that warrant further scrutiny, oversight, and precautionary safety measures*."[11];
>
> b) *Capability thresholds*: Definir limiares com base em capacidades do modelo é uma boa estratégia, visto que servem como um gatilho-chave (*key trigger*) para saber se

[11] Tradução livre: *Compute thresholds* devem, portanto, ser usados como um filtro inicial para identificar modelos que garantam maior precisão nos cálculos, supervisão e medidas de segurança preventivas.

medidas de segurança adicionais devem ser implementadas antes que uma atividade de alto risco possa prosseguir;

c) *Risk thresholds*: As estimativas de risco tentam medir o nível de risco diretamente, mas ainda são altamente não confiáveis, por diversos motivos que fogem ao escopo deste artigo. Para os autores, em teoria, os limites de risco seriam o determinante ideal para quando medidas de segurança adicionais são necessárias. Porém, na prática, os limites de risco ainda não podem ser confiáveis para a tomada de decisões.

Assim, no que se refere aos limites de risco (*risk thresholds*), Koessler *et al.* (2024, p. 10-12) explicam que há duas maneiras pelas quais tais limites podem ser aplicados: (i) *Directly feeding into decisions* e (ii) *Indirectly feeding into decisions*. Usar limites de risco para alimentar diretamente a tomada de decisões de alto risco significa comparar as estimativas de risco com limites de risco predefinidos. Se o nível estimado de risco estiver acima dos limites de risco (*thresholds*), recomenda-se a implantação de medidas de segurança adicionais, mantendo-se o ciclo entre avaliar e gerir riscos. Já alimentar indiretamente as decisões, significa aplicar limites de risco para manter o risco abaixo de determinados valores ou níveis em uma matriz de risco. Quando o risco excede os limites de risco (*thresholds*) entram em ação as medidas de segurança estabelecidas para manter o risco abaixo dos limites de risco. A diferença está no momento do agir. Na primeira opção o risco estimado contribui para a definição dos limites e indica o momento da implantação de medidas de segurança. Enquanto que na segunda opção, o limite de risco previamente definido é confrontado a cada momento com o risco calculado e ao exceder os limites aplica-se alguma medida de segurança.

Koessler *et al.* (2024, p. 12-15) apresentam como argumentos para aplicar modelos baseados em *risk thresholds*: a) auxiliar no alinhamento da conduta empresarial com a preocupação social; b) possibilitar a alocação consistente de recursos de segurança; c) garantir que os resultados da estimativa de risco sejam realmente aplicados; d) evitar raciocínio motivado sobre qual nível de risco é aceitável; e) evitar o bloqueio prematuro de medidas de segurança. Por outro lado, listam alguns pontos contrários: a) *risk thresholds* dependem de estimativas de risco e estimar riscos em sistemas de IA é extremamente difícil; b) *risk thresholds* podem criar um incentivo para produzir estimativas de risco artificialmente baixas; c) definir limites de risco para sistemas de IA envolve lidar com escolhas normativas complexas.

Como fechamento do trabalho, Koessler *et al.* (2024, p. 12-15) explicam como aplicar *risk thresholds* para alimentar indiretamente as decisões, ajudando a definir limites de capacidade. No entanto, limites de risco devem ser aplicados para informar sobre aplicabilidade

de sistemas, bem como projetar, adequar e manter medidas de segurança. Não devem ser a única base de uma regra de decisão, levando-se em conta outras considerações. Por isso, Koessler *et al.* (2024, p. 16) relacionam *risk thresholds* com *type of risk*, *area of risk* ou *category of risk*. E, mesmo que estes termos não tenham um conceito padrão, os autores apontam que cada *type of risk* refere-se a um grupo de cenários de risco que têm impacto, origem ou outras características semelhantes. Assim, cada *risk threshold* estará ligado a um tipo de risco específico, por exemplo, riscos financeiros, legais, de reputação, operacionais ou estratégicos. O que acontece na prática em relação aos riscos de sistemas de IA é que os tipos de riscos acabam por ser distinguidos pelo tipo de dano e, ainda, pelo domínio ou modalidade de ocorrência do dano. É compreensível, portanto, o fato do *AI Act* (União Europeia, 2024) e o projeto de Lei No. 2.338/2023 (Brasil, 2023) categorizarem os riscos com base no dano advindo do uso pretendido ou esperado dos sistemas de IA.

5. CONSIDERAÇÕES FINAIS

Decorrente do que foi apresentado e discutido, conclui-se que adotar modelos de regulação com base em riscos (*Risk Regulation*) pode ser interessante do ponto de vista do impacto ou dano advindo de um sistema de IA. Finalmente, alguns aspectos são relevantes ao se analisar regulamentações que seguem um modelo baseado em riscos, a saber: (i) visa garantir direitos fundamentais, como privacidade, não discriminação e dignidade humana, mas sem deixar de incentivar o uso responsável e ético de sistemas de IA; (ii) visa promover a inovação, o desenvolvimento e a sustentabilidade, sem esquecer do crescimento econômico, da competitividade e da livre concorrência; (iii) pode-se considerar diferentes níveis de risco dependendo do impacto do sistema de IA, regulando com rigor práticas de alto risco ou até mesmo considerando inaceitáveis práticas que envolvem dados pessoais sensíveis, exploração de grupos sensíveis ou vulneráveis, uso de técnicas subliminares para influenciar ou modificar comportamentos, entre outros; (iv) tem por interesse promover a confiança dos usuários, visto que os sistemas de IA terão por base a transparência e a responsabilização por meio da respectiva regulação; (v) promover o equilíbrio do ônus regulatório diante de danos potenciais, de modo a permitir que desenvolvedores de aplicações baseadas em IA mantenham o interesse pela área. Tecnicamente, regulamentações de sistemas de IA que seguem modelos baseado em riscos fornecem flexibilidade para enfrentar a diversidade e complexidade inerentes aos sistemas de IA.

REFERÊNCIAS

BENANTI, P.. The urgency of an algorethics. Discovery Artificial Intelligence, vol. 3, n. 11, 2023. Disponível em: https://link.springer.com/article/10.1007/s44163-023-00056-6 Acesso em: 05 jun. 2025.

BRASIL. **Projeto de Lei *n° 2.338, de* 2023**. Dispõe sobre o uso da Inteligência Artificial. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/157233 Acesso em: 05 jun. 2025.

BRASIL. **Lei Geral de Proteção de Dados Pessoais** (LGPD), Lei No. 13.709, de 14 de agosto de 2018, Lei Geral de Proteção de Dados - LGPD, 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm Acesso em: 05 jun. 2025.

CAPARROZ, R.; TONIN, C.B.; FERREIRA, B.S.; CRUVINEL, A. MATOS, D.; BATISTA, V.; BALBY, L.. **Dataset discrimination in government surveillance**: a threat to equality and justice. Task Force 05 – Inclusive Digital Transformation, T20 Policy Brief, G20 Brasil, 2024.

CAVEDON, R.; FERREIRA, H.S.; FREITAS, C.O.A.. O Meio ambiente digital sob a ótica da teoria da sociedade de risco: os avanços da informática em debate. **Revista Direito Ambiental e Sociedade**, v. 5, p. 194-223, 2015.

COMISSÃO EUROPEIA. **White Paper on Artificial Intelligence: a European approach to excellence and trust** (Livro Branco sobre a inteligência artificial: uma abordagem europeia virada para a excelência e a confiança). Brussels, 19.2.2020, COM(2020), 65 final, 2020. Disponível em: https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en Acesso em: 05 jun. 2025.

FLORIDI, L.. **The 4th Revolution**: How the Infosphere is Reshaping Human Reality. New York: Oxford University Press, 2014.

FREITAS, C.O.A.. O Direito e a inteligência Artificial como não-coisa. **CONPEDI LAW REVIEW**, XIII ENCONTRO INTERNACIONAL DO CONPEDI URUGUAI – MONTEVIDÉU, v. 10, n. 1, p. 88 – 109, JUL – DEZ, 2024.

FREITAS, C.O.A.; BARDDAL, J.P.. Análise preditiva e decisões judiciais: controvérsia ou realidade?. **DEMOCRACIA DIGITAL E GOVERNO ELETRÔNICO**, 2019, v. 1, p. 107-126.

GELLERT, R.. Understanding the notion of risk in the General Data Protection Regulation. Computer Law & Security Review: **The International Journal of Technology Law and Practice**, p. 1-10, 2017.

HAN, Byung-Chul. **Não-coisas**: transformações no mundo em que vivemos. Trad. Ana Falcão Bastos. Lisboa: Relógio D'Água Editores, 2022.

HARARI, Y. N.. **Homo Deus**: uma breve história do amanhã. Trad. Paulo Geiger. 1ª. ed., São Paulo: Companhia das Letras, 2016.

ISO. **Risk management** - principles and guidelines. International Organization for Standardization, Geneva, Switzerland, 2009.

KOESSLER, L.; SCHUETT, J.; ANDERLJUNG, M.. Risk thresholds for frontier AI. Centre for the Governance of AI, 2024. Disponível em: https://arxiv.org/pdf/2406.14713 Acesso em: 05 jun. 2025.

LADAK, A.; WILKS, M.; LOUGHNAN, S.; ANTHIS, J.R.. **Robots, chatbots, self-driving cars**: perceptions of mind and morality across artificial intelligences. In CHI Conference on Human Factors in ComputingSystems (CHI '25), April 26-May 1, 2025, Yokohama, Japan. ACM, New York,NY, USA, p.1-19 pages. https://doi.org/10.1145/3706598.3713130

MÖKANDER, J.; FLORIDI, L.. Operationalising AI governance through ethics-based auditing: an industry case study. **AI Ethics**, 2022. https://doi.org/10.1007/s43681-022-00171-7 Acesso em: 05 jun. 2025.

PETIT, N.; DE COOMAN, J.. **Models of law and regulation for AI**. European University Institute, RSCAS 2020/63 - Robert Schuman Centre for Advanced Studies, 2020. Disponível em: https://hdl.handle.net/1814/68536 Acesso em: 05 jun. 2025.

SBC - SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. **Plano de Inteligência Artificial da Sociedade Brasileira de Computação**. Porto Alegre: Sociedade Brasileira de Computação (SBC), jul., 2024. DOI 10.5753/sbc.rt.2024.141.

UNESCO. **Recommendation on the ethics of Artificial Intelligence**. United Nations Educational, Scientific and Cultural Organization, 2022. Disponível em: https://www.unesco.org/en/artificial-intelligence/recommendation-ethics Acesso em: 05 jun. 2025.

UNIÃO EUROPEIA. **Artificial Intelligence Act**. Bruxelas, 21.4.2021 COM (2021) 206, final 2021/0106 (COD). Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:52021PC0206 Acesso em: 05 jun. 2025.

TSAMADOS, A.; AGGARWAL, N.; COWLS, J.; MORLEY, J.; ROBERTS, H.; TADDEO, M.; FLORIDI. L.. The ethics of algorithms: key problems and solutions. **AI & SOCIETY**, v. 37, p. 215-230, 2022. Disponível em: https://link.springer.com/article/10.1007/s00146-021-01154-8 Acesso em: 05 jun. 2025.

WAGLE, P.. **What Is algorethics?** 2024. Disponível em: https://paulwagle.com/what-is-algorethics/ Acesso em: 28 abr. 2025.

ZUBOFF, S.. **A era do capitalismo de vigilância**. Trad. George Schlesinger. Rio de Janeiro: Editora Intrínseca Ltda., 2021.