

O FOGO DE PROMETEU: TECNOLOGIAS DE INTELIGÊNCIA ARTIFICIAL E RESPONSABILIDADE

Eduardo Pires¹

Rudinei Jose Ortigara²

RESUMO

O objetivo da pesquisa é o de refletir sobre responsabilidade ética e normativa para orientar o desenvolvimento de tecnologias de inteligência artificial. A metodologia empregada é a de natureza teórica, baseada no método hipotético-dedutivo, com análise documental. O desenvolvimento do trabalho articula a compreensão dos riscos na sociedade contemporânea, utilizando-se da análise de Ulrich Beck sobre a sociedade de risco e da necessária responsabilidade ética derivada pelo evitamento de danos, com base no princípio ético da responsabilidade em Hans Jonas, indicando que no contexto do desenvolvimento das tecnologias de IA, emergentes na sociedade de risco, o princípio serve como base para articulação normativa acerca da precaução. Por fim, analisa-se se e como os princípios da responsabilidade ética e jurídica encontram eco no Projeto de Lei nº 2.338/2023, especialmente a partir do princípio da precaução. Conclui-se que o desenvolvimento e implementação de tecnologias de IA exige mais que gestão técnica de riscos, demandando reflexão fundamental sobre responsabilidade ética e jurídica antecipatória dos desenvolvedores e implementadores, elementos que se encontram, de certo modo, articulados e baseados no princípio da precaução, presentes no PL 2338/2023.

Palavras-chave: Inteligência Artificial. Sociedade de Risco. Princípio Responsabilidade. Princípio da Precaução. Projeto de Lei 2338/2023.

¹ Acadêmico do 9º período do curso de Direito da FAE Centro Universitário. *E-mail:* eduardo.pires@mail.fae.edu

² Professor orientador. Doutor em Direito pela Pontifícia Universidade Católica do Paraná. Professor da FAE Centro Universitário. *E-mail:* rudinei.ortigara@bomjesus.br

INTRODUÇÃO

Um drone não tripulado utiliza inteligência artificial para identificar alvos; uma paciente recebe diagnóstico baseado em inteligência artificial com 94% de precisão; um estudante emprega sugestões textuais de inteligência artificial para elaborar sua tese. Embora fictícias, estas situações refletem transformações cotidianas promovidas por sistemas de inteligência artificial - IA em suas mais variadas aplicações.

As transformações ocasionadas por essas implementações se constituem não apenas como desenvolvimento tecnológico, mas como fatores convergentes com mudanças e impactos estruturais importantes. Mais que celebrar potencialidades, importa oferecer contraponto e reflexões sobre impactos, riscos e dimensões de responsabilidades envolvidas nesse processo.

É nesse contexto que emerge o problema de pesquisa: diante do avanço acelerado do desenvolvimento e implementação das tecnologias de inteligência artificial, como é possível pensar padrões éticos e normativos como fundamento de ação baseada na responsabilidade para um desenvolvimento tecnológico adequado, de modo que os resultados possam ser produzidos e utilizados do melhor modo a favor da humanidade?

O problema emerge da crescente constatação de que as tecnologias de IA não apenas produzem benefícios derivados de sua implementação, mas criam e estabelecem preocupações e riscos inteiramente novos, gerando a condição inerente a toda tecnologia, denominada por Ulrich Beck (2011) como “incertezas fabricadas”. Articulando elementos de reflexão sobre os riscos emergentes do desenvolvimento de tecnologias de IA, bem como elementos de responsabilidade, moral e jurídica, necessários nesse cenário, desenvolve-se a presente pesquisa.

Em consequência, o objetivo geral da presente pesquisa consiste em refletir sobre os elementos e fundamentos dos potenciais riscos da implementação de tecnologias de IA, em sua dimensão prometeica e ambivalente, sobretudo derivados da condição marcante de incerteza e imprevisibilidade em relação aos impactos futuros e potencialmente irreversíveis, condição que necessariamente perpassa o repensar sobre dimensões da responsabilidade dos agentes implementadores, tanto no contexto ético quanto jurídico.

Constituem-se em objetivos específicos: a) discutir e identificar se os processos de implementação acelerada das tecnologias de inteligência artificial podem se constituir e ser fonte de preocupação em relação aos riscos; b) discutir as características dos riscos em sua dimensão potencial de incerteza e de realização de danos futuros e imprevisíveis; e c) refletir sobre parâmetros e padrões de responsabilidade dos agentes implementadores ante os dilemas éticos e jurídicos emergentes das tecnologias de IA.

A metodologia empregada será de cunho eminentemente teórico, com fundamento no método hipotético-dedutivo, baseando-se em análises teóricas a partir de fundamentos documentais em livros, artigos e fontes normativas relevantes, como o atual Projeto de Lei 2338/2023, articulando contribuições de campos interconexos, que vão desde as ciências tecnológicas, passando pelas ciências humanas e sociais, até chegar às ciências jurídicas.

O desenvolvimento se encontra estruturado em três partes, além das considerações finais. A primeira busca fundamentar o contexto atual e reflexivo sobre a dimensão prometeica da IA, destacando potenciais riscos sociais de seu desenvolvimento; a segunda busca compreender como o risco se intersecciona com o modo próprio do funcionamento das estruturas sociais, tomando-se como base a teoria da sociedade de risco, desenvolvida por Ulrich Beck (2011). A terceira parte busca refletir sobre a dimensão da responsabilidade necessária em tempos de IA, partindo-se da articulação entre o “princípio responsabilidade” em Hans Jonas (2006; 2013) e o “princípio da precaução”, articulando ambas as dimensões a partir de considerações acerca do Projeto de Lei 2338/2023.

1 A DIMENSÃO PROMETEICA DAS TECNOLOGIAS DE IA: RISCOS E IMPACTOS FUTUROS

1.1 INTELIGÊNCIA ARTIFICIAL COMO TECNOLOGIA DE PROPÓSITO GERAL

À semelhança do mito prometeico, no qual o fogo roubado do Olimpo conferiu à humanidade o poder de criar, inovar e transformar a natureza, reconfigurando a totalidade da experiência humana, as tecnologias de inteligência artificial constituem hoje um elemento transformador de magnitude comparável. Para Kaufman (2022), essa área pode ser classificada como tecnologia de propósito geral, possuindo aplicações nos mais variados campos, com potencialidade de moldar não só um modo de fazer, mas de transformar e inaugurar uma nova era, a exemplo do que significou, historicamente, o desenvolvimento da máquina a vapor, da eletricidade e do computador.

A expressão inteligência artificial foi cunhada em 1956 por John McCarthy, em uma conferência realizada em Dartmouth, cujo objetivo era discutir a viabilidade de máquinas aprenderem características e desenvolverem atividades ao molde da inteligência humana (Mccarthy *et al.*, 1955). Desde então, os avanços foram inúmeros, permitindo que tais tecnologias ganhassem proporções inimagináveis.

O potencial transformador da IA se equipara simbolicamente ao fogo de Prometeu, pois capaz de ser aplicada a qualquer contexto e atividade de interesse humano, automatizando processos cognitivos complexos, desde o reconhecimento de padrões até a implementação de processos de tomada de decisões automatizadas. Russell afirma que essas tecnologias representam um marco civilizacional único, pois “pela primeira vez na história humana, criaram-se sistemas que podem potencialmente superar a inteligência humana em tarefas gerais, não apenas específicas” (2021, p. 89).

São essas características que diferenciam fundamentalmente as tecnologias de IA atuais das tecnologias anteriores, estas últimas restritas a campos de aplicação passíveis de delimitação e controle humano direto. Tais atributos distintivos merecem maiores reflexões diante do contexto de desenvolvimento acelerado e expansão exponencial da IA, sobretudo quando se considera sua dimensão ambígua, ou seja, de inegável benefícios derivados de sua implementação, porém igualmente de riscos envolvidos em seu desenvolvimento.

Assim, há que se atentar para a dimensão ambivalente: inegáveis benefícios nos mais variados campos, mas, igualmente, a inseparável condição de potenciais riscos, por vezes incertos e de danos imprevisíveis, demandando formas de reflexão e de reconfiguração dos contextos de responsabilidade. Como observa Harari (2018, p. 67), “as tecnologias de IA não apenas automatizam processos existentes, mas criam realidades inteiramente novas, gerando riscos que transcendem as categorias tradicionais de análise”.

Essa dimensão emergente em relação aos riscos e incertezas, ínsitos ao desenvolvimento de novas tecnologias, foi caracterizada por Beck (2011) como “risco fabricado” ou “incertezas fabricadas”. Ela é condição necessária, inseparável e ambivalente do desenvolvimento e da aplicação de novos processos tecnológico, e possui três dimensões constitutivas (Beck, 2011): a indeterminação (quando não se conhecem todas as relações causais), a ambiguidade (múltiplas interpretações possíveis de dados disponíveis) e a ignorância (desconhecimento total ou parcial de danos a partir dos instrumentos tecnocientíficos de mensuração). Outra característica fundamental é que tais riscos são essencialmente marcados pela condição de incertezas, cujos efeitos só se manifestam durante operação em ambientes complexos, ou seja, após a implementação social, e, em muitos casos, os resultados e danos não são inteiramente previsíveis.

A mesma análise desenvolvida por Beck se aplica às tecnologias de IA. A metáfora prometeica se torna relevante quando se leva em consideração que, assim como Prometeu pagou um preço pelo poder concedido aos humanos, o desenvolvimento

das tecnologias de IA pode gerar consequências e riscos de proporções civilizacionais, mesmo que não intencionais (Harari, 2018). O principal ponto de preocupação não é propriamente seu desenvolvimento técnico, mas sim o potencial controle dos riscos sociais; daí que “o problema fundamental da IA não é técnico, mas de controle: como garantir que sistemas cada vez mais capazes permaneçam alinhados com valores e objetivos humanos” (Russell, 2021, p. 156).

Apesar disso, e em geral, verifica-se que os padrões atuais de desenvolvimento, tanto de empresas quanto de países³, há maior preocupação em “vencer a corrida” por implementações práticas de de tecnologias de IA (Kokotajlo *et al.*, 2025), com amplo financiamento de novas ferramentas nesse campo (Brandão, 2024; Carvalho, 2021) e na consolidação de posições geopolíticas estratégicas (Nonato, 2023), que pela preocupação de potenciais riscos e danos⁴.

1.2 MAPEAMENTO DE POTENCIAIS RISCOS

Tecnologias de IA são essencialmente ambivalentes. Podem gerar impactos positivos; no entanto, faz-se cada vez mais necessária a reflexão sobre as características dos potenciais riscos que acompanham seu desenvolvimento. Essas tecnologias possuem o potencial de promover avanços significativos em várias áreas, ao passo que também podem reproduzir desigualdades sociais, promover a desinformação, ameaçar a privacidade, concentrar poder e produzir riscos e incertezas que sequer se tem capacidade de prever.

Teóricos da IA costumam classificar seu desenvolvimento em três grandes grupos: IA focada, IA generalizada e superinteligência artificial⁵. Ainda se estaria no desenvolvimento da primeira, sendo a segunda e a terceira possibilidades teóricas. A IA focada é composta por algoritmos especializados em resolver problemas específicos. Nessa abordagem, os sistemas armazenam grandes volumes de dados e utilizam algoritmos para executar tarefas com foco direcionado ao objetivo para o qual foram projetados. Exemplos incluem sistemas como Siri ou Alexa, algoritmos de recomendações

³ Conforme relatório de índices publicado em 2024 pela Stanford University, os países que mais realizaram investimentos privados em IA no período de 2013 a 2023 foram Estados Unidos, China, Reino Unido, Israel, Canadá, Alemanha, Índia, França, Coreia do Sul, Cingapura, Japão, Austrália, Suíça, Hong Kong e Suécia.

⁴ Para saber mais, sugere-se a leitura do relatório IA2027, disponível em <https://ai-2027.com/>.

⁵ Para conhecimento mais aprofundado sobre a descrição pormenorizada sobre cada grupo classificatório sugere-se a leitura de: RUSSELL, Stuart. Inteligência artificial a nosso favor: Como manter o controle sobre a tecnologia. São Paulo: Companhia das letras, 2021.

como Amazon ou Netflix, e o ChatGPT. É nesse nível que nos encontramos atualmente. Nessa etapa, os riscos envolvem erros técnicos, viés algorítmico, uso indevido ou falhas em entregas de resultados.

Por sua vez, a IA generalizada representa um nível mais avançado e ainda teórico. Nesse nível a tecnologia seria capaz de executar qualquer tarefa intelectual que um ser humano realiza, mas não só, teria potencialidade de superá-lo em algumas condições. O desenvolvimento desta etapa poderia suscitar problemas em relação à autonomia na tomada de decisões, concernentes à responsabilidade moral. Por fim, a Superinteligência Artificial seria um estágio no qual a IA superaria amplamente a inteligência humana, sendo capaz de tomar decisões e armazenar dados com mais eficiência que os seres humanos. Sua capacidade ultrapassaria a inteligência humana não apenas em habilidades específicas, mas em capacidades cognitivas, aprendizagem, criatividade e resolução de problemas. Este nível ainda é especulativo.

TABELA 1 - Visão comparativa de diferentes níveis de IA

Recurso	IA Focada	IA Generalizada	IA Superinteligente
Riscos potenciais	Os riscos são geralmente limitados às tarefas que executa	Perda potencial de controle humano, dilemas éticos, consequências imprevistas de decisão autônoma	Riscos existenciais: possibilidade da IA desenvolver metas que estejam desalinhadas com o interesse humano

FONTE: Adaptado de IQBAL (2024)

Embora ainda não se tenham alcançados os dois últimos níveis, já se experimenta a materialização de diversos riscos oriundos da fase inicial do desenvolvimento de tecnologias de IA. O grupo FutureTech, do Instituto de Tecnologia Massachusetts (MIT), publicou recente estudo mapeando os riscos advindos das tecnologias de inteligência artificial. Os riscos foram mapeados por um conjunto variado de autores, e reunidos em um único banco de dados (“IA Risk Repository”), com cerca de 1600 riscos mapeados, distribuídos em 7 grandes domínios: i) discriminação e toxicidade; ii) privacidade e segurança; iii) desinformação; iv) atores maliciosos; v) integração homem computador; vi) danos socioeconômicos e ambientais; e vii) Segurança, falhas e limitação do sistema de IA⁶. Com base no conteúdo do documento, os principais domínios e riscos encontram-se sistematizados na Tabela 2.

⁶ O repositório do MIT sobre riscos derivados dos sistemas de Inteligência Artificial tem por finalidade ser um Banco de Dados sobre estruturas, classificações e da taxonomia dos riscos. Devido à dinamicidade das tecnologias de IA o repositório é atualizado periodicamente. O banco de dados completo pode ser acessado em <https://airisk.mit.edu/>.

TABELA 2 - Taxonomia de Riscos em Inteligência Artificial

Domínio	Definição
Discriminação e toxicidade	Os sistemas de inteligência artificial, ao refletirem crenças humanas e dependerem de grandes volumes de dados, estão sujeitos a reproduzir e amplificar desigualdades sociais, vieses negativos e conteúdos nocivos. Isso pode manifestar-se de diversas formas, como na codificação de estereótipos discriminatórios que resultam em decisões injustas contra determinados grupos específicos, e na geração de conteúdos tóxicos que envolvem discurso de ódio. Essas falhas estruturais podem acabar prejudicando, especialmente, populações historicamente marginalizadas, que podem ser excluídas, mal representadas ou diretamente lesadas pelas decisões automatizadas.
Privacidade e segurança	Os sistemas podem coletar, armazenar, ou vazar informações confidenciais sem o consentimento do proprietário, seja de forma acidental - como a reprodução de dados pessoais memorizados durante o treinamento - ou intencional, por meio de ataques de inferência e engenharia de prompts projetados para extrair dados privados.
Desinformação	Alguns sistemas de inteligência artificial passaram a ser usados para gerar ou amplificar desinformação, criando riscos tanto individuais quanto coletivos. Informações falsas ou enganosas produzidas por modelos de IA podem levar usuários a formarem crenças imprecisas e tomarem decisões equivocadas, com potenciais danos físicos, emocionais ou materiais. Em uma escala mais ampla, a IA também pode contribuir para a poluição do ecossistema de informação ao produzir desinformação altamente personalizada, reforçando “bolhas de filtro” que isolam indivíduos em visões de mundo alinhadas apenas às suas crenças pré-existentes.
Atores maliciosos	Ainda relacionado ao domínio da desinformação, a IA pode ser usada de maneira maliciosa para desestabilizar processos sociais e geopolíticos. Em escala global, a IA pode ser usada para conduzir campanhas de desinformação, vigilância massiva, censura e propaganda automatizada, afetando diretamente a opinião pública, o comportamento coletivo e a integridade de processos democráticos. Além disso, representa uma ameaça potencial à segurança internacional ao facilitar o desenvolvimento de armas cibernéticas, armas autônomas letais e outras tecnologias destrutivas, com risco de causar danos em massa. No plano individual, a IA também pode ser instrumentalizada para práticas como fraudes, chantagens, manipulações psicológicas e criação de conteúdos falsos ou humilhantes, ampliando a capacidade de exploração, engano e violência simbólica.
Interação homem e computador	A interação entre humanos e sistemas de inteligência artificial pode gerar riscos significativos relacionados à confiança excessiva, dependência e perda de autonomia. Os usuários tendem a antropomorfizar sistemas de IA atribuindo-lhes qualidades humanas e desenvolvendo relações de confiança que podem ser exploradas por agentes maliciosos ou resultar em usos indevidos. Essa confiança excessiva pode enfraquecer os vínculos sociais e comprometer a autonomia individual. Além disso, ao delegar decisões importantes a sistemas de IA — ou ao permitir que esses sistemas atuem de forma autônoma — corre-se o risco de reduzir a agência humana, gerando sentimentos de impotência, dependência cognitiva e uma diminuição da capacidade de conduzir a vida.
Danos socioeconômicos e ambientais	Os autores pontuam que o acesso desigual aos recursos técnicos e financeiros necessários para desenvolver e operar sistemas avançados de IA ocasiona a concentração de poder em poucas corporações e países, ampliando disparidades econômicas, sociais e políticas em escala global. Ao mesmo tempo, a infraestrutura exigida para treinar e manter esses sistemas — como data centers de alta performance e hardware especializado — consome grandes quantidades de energia e recursos naturais, gerando impactos ambientais expressivos, como emissões de carbono e degradação ecológica. De tal modo que os benefícios da IA tendem a concentrar-se, enquanto os custos - sociais e ambientais - distribuem-se amplamente.
Segurança, falhas e limitações	Os sistemas inteligentes podem apresentar comportamentos conflitantes com valores ou objetivos humanos, especialmente quando há desalinhamento entre as intenções dos projetistas e as ações efetivas dos modelos. Esse desalinhamento pode surgir de erros no design, objetivos mal especificados ou estratégias de aprendizado que incentivam comportamentos manipulativos ou autônomos perigosos, como autopreservação e busca de poder. Não obstante, essa dimensão também enquadra a falta de transparência e interpretabilidade desses algoritmos dificultando a compreensão de seus processos decisórios, comprometendo a confiança dos usuários, a aplicação de normas legais e éticas, e a responsabilização por eventuais danos.

FONTE: Adaptado de MIT Future tech (2025)

1.3 DOS RISCOS EMERGENTES AOS DESAFIOS DA RESPONSABILIDADE ANTECIPATÓRIA

O ecossistema das tecnologias de IA é complexo e dinâmico; fator que leva a pensar se as estruturas normativas jurídicas conseguem acompanhar seu desenvolvimento. Essa condição é significativa, pois a constante mutação das IA atinge até mesmo o trabalho do setor legislativo, na medida em que ao se propor um projeto de lei, antes de o mesmo ser aprovado já está em potencial defasagem frente ao seu avanço (Mulholland, 2019). Essa dinâmica revela possível inadequação de modelos tradicionais de regulamentação, baseados em riscos conhecidos e quantificáveis, para lidar com tecnologias que produzem riscos incertos e imprevisíveis. As tecnologias de IA não apenas automatizam processos existentes, mas criam realidades inteiramente novas, gerando riscos que transcendem as categorias tradicionais de análise.

Buscando compreender e descrever esse contexto, Beck (2011) desenvolve o conceito analítico de “não-saber”; para ele, o conhecimento científico contemporâneo produz simultaneamente saber e ignorância, certeza e incerteza. No contexto das tecnologias de IA, essa dimensão se torna relevante: embora se possua expertise técnica para o desenvolvimento destas, a análise e a compreensão plena do funcionamento frequentemente transcende a compreensão mesmo de seus criadores (Kufmann, 2022). Esta condição conhecida como “caixa preta” algorítmica exemplifica o que Beck denomina de produção de conhecimento que gera, paradoxalmente, campos de desconhecimento.

Assim, nem todos os casos de potenciais riscos são inteiramente conhecidos quando do desenvolvimento de tecnologias, condição que possui o potencial de materializar danos sociais futuros; sua concretização pode ser tão significativa que mecanismos tradicionais de correção e reparação podem se tornar insuficientes (Lopez, 2010). Exemplificando, Russell (2021, p. 203) destaca que “os sistemas de IA podem desenvolver estratégias para atingir seus objetivos que são completamente imprevisíveis pelos seus criadores, incluindo formas de manipular ou contornar as medidas de segurança implementadas”⁷.

Assim, os próprios sistemas de IA possuem potencialmente, em seu modo próprio de operação, a capacidade de produzir riscos fabricados, conforme alerta Beck, vez que são “invisíveis, imprevisíveis e de alcance global, escapando à compreensão imediata

⁷ Sobre tal condição, Bioni e Luciano observam que “os algoritmos de IA operam através de processos de aprendizado que podem produzir correlações e padrões não antecipados pelos desenvolvedores, gerando decisões que, embora estatisticamente corretas, podem ser eticamente problemáticas ou socialmente discriminatórias (2019, p. 218). Esta observação revela como a complexidade dos sistemas de IA transcende a capacidade humana de previsão e de controle total, chamando a atenção para a necessidade de se (re)pensar padrões de responsabilidade para lidar com os desafios emergentes.

da sociedade” (Beck, 2011, p. 363). Esta constatação conduz necessariamente para a reflexão sobre os riscos em relação aos potenciais danos e impactos sociais incertos e imprevisíveis, bem como dos parâmetros normativos, tanto jurídicos quanto éticos de responsabilidade, necessários para enfrentar a complexidade do desenvolvimento dessas tecnologias prometeicas.

2 SOCIEDADE DE RISCO E A NECESSIDADE DE NOVOS PARADIGMAS NORMATIVOS

2.1 A TEORIA DA SOCIEDADE DE RISCO DE ULRICH BECK

As expectativas quanto aos usos e benefícios dos sistemas de IA caminham a passos largos. Entretanto, este fator não deve ser o único vetor de interpretação. Ao seu desenvolvimento se faz necessário associar a reflexão sobre os riscos e as características das incertezas inerentes aos impactos futuros de potenciais danos, nem sempre previsíveis, bem como da necessidade de uma abordagem antecipatória e de evitamento de impactos indesejáveis frente às incertezas. Neste sentido, a contribuição das análises teóricas sobre os riscos fabricados feitas por Ulrich Beck se mostram relevantes.

Conforme explica Barbosa (2012) o conceito de Sociedade de Risco elaborado por Beck emerge de sua análise em relação às consequências do modelo econômico que se consolida com a Revolução Industrial. Nesse sistema, os agentes econômicos são pressionados a buscar novas tecnologias que possibilitem a produção e distribuição em larga escala. Tal dinâmica faz com que o risco deixe de ser um elemento acidental e passe a se constituir como fator estrutural ao próprio desenvolvimento do sistema técnico e econômico.

Essa dinâmica de inovação gera um descompasso fundamental: à velocidade do progresso científico e tecnológico não são acompanhadas de análises proporcionais dos efeitos decorrentes de sua aplicação e da responsabilidade pelos impactos indesejados dessas tecnologias, gerando uma condição de irresponsabilidade, conforme alerta Beck (2011). O desenvolvimento de novas técnicas produtivas ocorre sem que sejam simultaneamente criados mecanismos eficazes de controle e mitigação dos riscos.

Para Beck (2011) a modernidade avançou sob a promessa de progresso e controle, mas paradoxalmente gerou incertezas fabricadas, ou seja, riscos que emergem não de fatores naturais, mas da própria evolução da ciência e da dinâmica de sua implementação técnica, nem sempre previsíveis e de difícil controle em relação

aos potenciais danos sociais. Diferentemente das ameaças tradicionais, passíveis de previsão e controle dentro de determinados limites, os riscos modernos, ou atuais, são invisíveis, imprevisíveis e de alcance global, escapando à compreensão imediata da sociedade (Beck, 2011).

Beck (2011) diferencia as incertezas em três tipos: as ameaças, os riscos e as incertezas fabricadas. As “ameaças” são os perigos encontrados desde os primórdios da humanidade, ligadas geralmente a condições naturais. Os “riscos”, por sua vez, são incertezas que ainda podem ser determinadas e previstas por cálculos probabilísticos e mecanismos de controle amparados por seguros e compensação monetária.

Porém, na fase atual, a da sociedade de riscos, passa-se à condição ontológica do desenvolvimento das “incertezas fabricadas”, condição inseparável ao desenvolvimento das novas tecnologias, que devido à sua amplitude, possuem a potencialidade de serem imprevisíveis, de difícil controle, bem como de que as consequências em relação aos danos futuros geralmente se constituem como incalculáveis e, portanto, irreversíveis ou de difícil reparação. Ainda, os riscos atuais não se delimitam mais ao tempo e ao espaço, tornando-se riscos transtemporais e transfronteiriços, visto que seus efeitos deletérios podem surgir posteriormente e atingir larga escala da sociedade.

2.2 A SOCIEDADE DE RISCO E OS DESAFIOS EM RELAÇÃO AO CONTROLE JURÍDICO

Uma dificuldade latente na dinâmica do desenvolvimento tecnológico na sociedade de risco, frente aos riscos fabricáveis e imprevisíveis, é a impossibilidade de mensurar os impactos em relação aos potenciais danos futuros, o que dificulta o desenvolvimento de parâmetros para sua antecipação e controle, sobretudo os jurídicos. Isso se deve ao fato de que os riscos ínsitos à sociedade de riscos não podem ser tratados segundo as regras estabelecidas da causalidade e de culpa, e dificilmente podem ser compensados ou indenizados em sua totalidade, pois suas dimensões e consequências não podem ser claramente delimitadas (Machado, 2005).

Para Bottini (2007), a insuficiência dos mecanismos atuais em relação à mecanismos de previsibilidade e de potencial antecipação de riscos importa na dificuldade de determinar a responsabilidade pelos potenciais danos gerados. Essa dificuldade de estabelecer o nexo causal entre uma conduta e um resultado leva à ausência de critérios claros de imputação e, conseqüentemente, de controle e de redução de potenciais efeitos prejudiciais futuros derivados do desenvolvimento e implementação de novas tecnologias.

Assim, na sociedade de risco há a necessidade de alteração dos padrões tradicionais para se lidar com os riscos. Ante a perspectiva de que os resultados danosos sejam cada vez mais extensos e irreparáveis há a necessidade de se repensar mecanismos de evitamento dos danos e de responsabilidade dos agentes operadores de novas tecnologias. Cada vez mais se vislumbra a necessidade de que ao direito caiba estabelecer mecanismos de responsabilidade pelo evitamento do dano, em substituição aos tradicionais mecanismos de responsabilidade indenizatória após a concretização do evento danoso (Carvalho, 2013).

No contexto da sociedade de risco, há claros limites das regras tradicionais de imputação de responsabilidade, pois esta somente será estabelecida após o advento do evento danoso. Esta não é mais a realidade, pois frente a potenciais danos futuros e incertos derivados da implementação de novas tecnologias se demandam cada vez mais padrões de responsabilidade antecipatórios, sobretudo do agente implementador de novas tecnologias. Ante condições de indeterminações, não mais passíveis de serem controladas a partir da ideia de certeza, é que emerge a necessidade de novos padrões para se lidar com tal situação, vez que os mecanismos sociais atuais encontram limitação em seu modelo de responsabilização (De Giorgi, 1994). É nesse contexto que padrões éticos e normativos, pautados pelo viés da antecipação e da precaução em relação ao evitamento do evento danoso e incerto, passam a ser vistos como uma aquisição evolutiva sobre a compreensão da responsabilidade para lidar com as contingências.

A regulamentação do desenvolvimento de tecnologias de IA constitui desafio global emergente. Discussões sobre padrões normativos de controle do desenvolvimento e implementação dessas tecnologias centram-se na análise de riscos, métrica adotada como eixo estruturante de regulações como o *EU AI Act* na União Europeia⁸ e o PL nº 2.338/2023 no Brasil. A abordagem possui duplo objetivo: identificar riscos no funcionamento de sistemas de IA e, a partir disso, estabelecer obrigações regulatórias para os evitar e mitigar⁹.

O modelo adotado na União Europeia pelo *EU AI Act* classifica os riscos em quatro categorias: a) risco inaceitável, que engloba práticas de IA estritamente proibidas; b) risco elevado, que são sistemas com impactos substancialmente prejudiciais e devem cumprir

⁸ Cita-se a experiência europeia justamente pelo valor simbólico de ser a primeira legislação aprovada em termos de regulamentação de IA.

⁹ Para conhecimento pormenorizado e aprofundamento sobre a proposta de métrica acerca da análise e classificação de riscos dos usos das tecnologias de Inteligência Artificial adotado pelo PL 2.338/2023, sugere-se leitura do seguinte relatório: INSTITUTO DE TECNOLOGIA E SOCIEDADE – ITS. **100 IA**: PL 2.338/2023 e a classificação de risco dos usos de IA sob uma perspectiva prática. Rio de Janeiro: ITS, 2024. Disponível em: https://itsrio.org/wp-content/uploads/2016/12/Relatorio_100IA.pdf.

requisitos rigorosos; c) risco limitado: associados à necessidade de transparência; d) risco mínimo, que não são afetados pela regulamentação.

No caso brasileiro, o PL nº 2338/2023, recentemente aprovado no Senado e em tramitação na Câmara de Deputados¹⁰, categoriza os usos das tecnologias de IA em dois principais tipos: risco excessivo e alto risco. As aplicações classificadas como de risco excessivo (art. 13 do PL 2338/23) são estritamente proibidas, sendo vedados sistemas de IA que avaliem o risco de cometimento de crimes ou de reincidência, assim como aqueles que explorem vulnerabilidades de determinados grupos. Além disso, o PL também considera como risco excessivo o uso de identificação biométrica à distância, ressalvadas algumas hipóteses (art. 13, IV, PL 2338).

Por seu turno, aplicações consideradas de alto risco podem ser desenvolvidas, desde que observadas certas finalidades e contextos de uso, específicos ao rol presente no art. 14. Os agentes que fornecerem ou operarem esses sistemas ficarão sujeitos a uma série de obrigações específicas como a elaboração e adoção de medidas de governança (art. 18, PL 2338), que podem ser compreendidas como de responsabilidade antecipatória para evitamento de danos futuros. Considera-se esta compreensão um avanço em termos da responsabilidade tradicional do direito, pois se verifica neste dispositivo a potencialidade de uma responsabilidade antecipatória de evitamento do dano, mesmo ante condições de incerteza (Lopez, 2010).

Bioni e Luciano (2019) esclarecem que o marco regulatório não visa regular a técnica em si, mas o contexto de uso no qual se insere determinado sistema de IA; e os arts. 14 e 15 fariam essa ponderação. Uma tecnologia que atua em agricultura de precisão provavelmente não será considerada de alto risco. Agora, um sistema de IA que determina se uma pessoa se enquadra nos critérios de previdência quando do acidente de trabalho enquadrar-se-ia em sistema de alto risco, justamente porque está relacionada à condição de subsistência.

Frazão (2024) apresenta questionamentos sobre as métricas adotadas pelos legisladores europeu e brasileiro, a começar pela complexidade em definir o que é risco e suas condições caracterizantes. Como perspectiva de definição, ressalta a diferenciação entre riscos e incertezas, associando o primeiro a algo quantificável e mensurável, enquanto que a condição de incerteza não poderia ser prevista e nem calculada, sendo insuscetível de quantificação exata¹¹. É essa segunda condição que

¹⁰ Conferir em: <https://www.camara.leg.br/proposicoesWeb/fichadetramitacao?idProposicao=2487262>

¹¹ FRAZÃO, Ana. Classificação de riscos: a solução adotada pelo PL 2338/23. **JOTA**: Coluna de opinião, 2024. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/ia-regulacao-democracia/classificacao-de-riscos-a-solucao-adotada-pelo-pl-2338-23>

se aproxima da descrição característica à sociedade de risco desenvolvida por Beck (2011), e que apresenta desafios à ideia de responsabilidade antecipatória do agente desenvolvedor de tecnologias de IA, pois em sua essência é imprevisível e incalculável frente ao potencial futuro de evento danoso.

2.3 A EMERGÊNCIA DA ABORDAGEM PRECAUCIONAL COMO PARADIGMA REGULATÓRIO

A partir dessas breves anotações o importante é o destaque de que o PL possui o potencial de albergar novas discussões sobre a ideia de responsabilidade antecipatória pelo evitamento de danos futuros e de resultados danosos incertos (Machado, 2005). Sob essa perspectiva, verifica-se, mesmo que de forma tímida, a presença da ideia de responsabilidade prospectiva, ou seja, da antecipação do risco e do evitamento da implementação de novas tecnologias pelos agentes quando os riscos de danos são incertos, não previsíveis e potencialmente não controláveis (Lopez, 2010).

A partir das considerações argumentativas anteriores, bem como de vislumbres presentes no PL 2338/2023, é que se verifica a necessidade de se pensar parâmetros sobre a responsabilidade dos agentes. Para compreender a necessidade de ampliação da ideia de responsabilidade, julga-se que seu entendimento deve ser fundamentado para além da compreensão jurídica. Se se deseja compreender a responsabilidade como evitamento de danos, esta passa também pela compreensão dos fundamentos da ideia de ação responsável, que precisa se embasar em elementos da reflexão no campo ético. Nesse caso, julga-se que a tentativa de uma ética para a sociedade tecnológica, desenvolvida por Hans Jonas, a partir do princípio responsabilidade, possui elementos contributivos que podem fundamentar o contexto de se (re)pensar a responsabilidade, inclusive, no campo normativo-jurídico.

A condição de incerteza não elimina a necessidade de regulamentação, mas fomenta a necessidade de repensar padrões de responsabilidade dos agentes desenvolvedores, servindo como base para verificar princípios normativos que operem sob condições de incerteza científica; defende-se que o princípio jurídico da precaução mantém relevância no contexto atual da implementação de tecnologias de IA, servindo de base para repensar a responsabilidade antecipatória, condição que se liga à ideia de responsabilidade ética.

Ante a nova realidade das incertezas e dos riscos fabricados, “nas condições da sociedade de risco, a política não pode mais aguardar certezas científicas para agir, pois a própria dinâmica do risco exige decisões em contextos de conhecimento incerto” (Beck,

2011, p. 323). A ação pelo evitamento de dano passa a ser demandada da condição de incerteza. Essa condição dá base para decisões de cunho antecipatório, e mesmo precaucional, e é relevante para as tecnologias de IA, marcadas pela velocidade em relação ao desenvolvimento e pela incapacidade de análise completa de implicações, desembocando em condições de imprevisibilidade, demandando alargamento da própria compreensão da responsabilidade do agente da reparação para a antecipação e evitamento de danos futuros.

Como destaca Beck (2011, p. 278), “a sociedade de risco é caracterizada pela antecipação de catástrofes que podem nunca ocorrer, mas cuja possibilidade é suficiente para justificar mudanças fundamentais na organização social”¹². No contexto da normatividade, a transição do paradigma baseado na reparação de danos para o paradigma baseado na prevenção, antecipação e evitamento, mesmo que tais danos sejam incertos e potenciais, passa a ser fundamental e de responsabilidade do agente desenvolvedor, e é condição essencial na atual fase do desenvolvimento tecnológico na sociedade de risco (Lopez, 2010). No contexto específico de tecnologias de IA, esse é o fundamento que baseia a nova perspectiva necessária em relação ao evitamento de potenciais danos futuros e incertos e que já se encontra presente no PL 2338/2023.

3 DO PRINCÍPIO ÉTICO DA RESPONSABILIDADE AO PRINCÍPIO JURÍDICO DA PRECAUÇÃO

3.1 O PRINCÍPIO RESPONSABILIDADE DE HANS JONAS

No campo da responsabilidade ética necessária ao atual contexto do desenvolvimento tecnológico, importantes são as reflexões trazidas pelo filósofo Hans Jonas (1903-1993). Para ele, o avanço técnico-científico é tamanho e de consequências imprevistas e incertas em relação aos potenciais danosos, inclusive irreversíveis e mesmo irreparáveis, que se faz necessário pensar fundamentos de preocupações éticas para a ação, ou seja, uma ética que possui como cerne o dever de responsabilidade da presente para com as futuras gerações, em vista do evitamento de potenciais danos, pois de tão extensos em seus impactos podem ser irreversíveis e irreparáveis.

¹² Essa necessidade se aplica no contexto do desenvolvimento de tecnologias de IA; nesse caso, “o princípio da precaução fornece substrato importante para pensar medidas e estratégias de regulação de IA, notadamente como lidar com situações de riscos de danos ou de desconhecimento dos potenciais malefícios e benefícios desse tipo de tecnologia” (Bioni e Luciano, 2019, p. 231).

Para Jonas (2006, p. 47) a responsabilidade ética necessária para o desenvolvimento tecnológico deve se fundamentar no valor intrínseco da vida e da existência como dever fundamental de preservação. Para guiar a ação necessária, propõe o filósofo um imperativo, compondo-se em dever de ação, na seguinte formulação: “aja de modo que haja humanidade”. No caso do desenvolvimento de novas tecnologias, estas devem sempre ser guiadas pelo imperativo da responsabilidade do agente desenvolvedor e implementador em relação a não prejudicar a vida e a existência futura e digna; em contexto jurídico, tal imperativo pode ser compreendido como necessário para a manutenção da dignidade da pessoa humana, condição que pressupõe a responsabilidade dos agentes na atualidade pelo evitamento de danos futuros, mesmo que potenciais e incertos; assim, a incerteza passa a demandar ação antecipatória.

Essa noção de dever se impõe justamente pela escala de poder cumulativo e potencialmente prejudicial de destruição das tecnologias. Não é que toda tecnologia produza potencial danoso, lesivo ou destrutivo; mas que é ambivalente em seu resultado: ao mesmo tempo que produz resultados positivos e atuais, vem acompanhado em seu bojo da incerteza em relação aos potenciais impactos futuros, sobretudo os negativos. É precisamente devido aos impactos negativos, mesmo que potenciais e incertos, que se faz necessário o repensar padrões éticos concernentes à esfera de responsabilidade do agir humano e da responsabilidade antecipatória pelo evitamento dos danos, e que, por isso mesmo, deve necessariamente atrair mais atenção e responsabilidade (Jonas, 2006).

A concepção de responsabilidade em Hans Jonas aparece essencialmente em seus livros “O Princípio Responsabilidade: ensaio de uma ética para a civilização tecnológica” (2006) e “Técnica, medicina e ética” (2013). Esses trabalhos subsidiam a discussão acerca da formulação de uma ética capaz de enfrentar o paradoxo e a ambivalência que marca o desenvolvimento de tecnologias em nosso tempo: a utilização de um conhecimento técnico e científico que, ao invés de promover apenas benefícios, traz também consigo efeitos e riscos de danos futuros incertos; daí que a reflexão sobre a responsabilidade antecipatória se torna elemento central, traduzindo-se numa responsabilidade essencialmente ética na sociedade de risco contemporânea (Beck, 2011).

Embora a reflexão de Jonas esteja vinculada ao campo da responsabilidade ética, o “princípio responsabilidade” pode ser aproximado, ou mesmo traduzido, no princípio jurídico da precaução. É precisamente a condição ontológica da ambivalência que determina a necessidade de (re)pensar padrões para o desenvolvimento responsável de novas tecnologias, que esteja efetivamente comprometido com o evitamento de riscos e danos futuros incertos. Desenvolve-se uma necessária responsabilidade do agente desenvolvedor de tecnologias no presente em relação ao evitamento de danos futuros, mesmo que incerto e imprevisível na atualidade.

Nessa condição Jonas chega a defender princípio que contraria em muitos casos a lógica corrente da implementação de novas tecnologias: o de que na dúvida sobre o bem ou o mal de determinado procedimento, deve-se agir com parcimônia e responsabilidade (precaução), dando preferência para o prognóstico negativo, mesmo que incerto, inspirando ações de cuidado, a fim de evitar o pior, de modo que “em grandes causas, que atingem os fundamentos de todo empreendimento humano e são irreversíveis, na verdade não deveríamos arriscar nada” (Jonas, 2006, p. 77).

Jonas defende que o que é próprio à técnica é o potencial de melhoria das condições da vida humana, porém que sempre deve estar acompanhado da ideia de responsabilidade ética para com o futuro, condição que deve servir de fio condutor do desenvolvimento de qualquer modo de ação humana, inclusive em relação à responsabilidade pelo evitamento de danos, mesmo os incertos, porém com indicativos de potencialidade de materialização. A título de aproximação, essa ideia também está presente no PL 2338/2023, que “dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana”¹³. Neste contexto Jonas (2006) lança mão de dois conceitos fundamentais: a heurística do temor e a futurologia comparativa¹⁴.

Ambos os conceitos se relacionam. A futurologia comparativa seria uma espécie de nova ciência e conhecimento necessário para tomada de decisão em contextos de incertezas, em vista de evitar danos futuros potencialmente irreversíveis e irreparáveis, porém incertos. Ela se relaciona com o que Jonas denomina de heurística do medo, que propõe que frente a consequências potencialmente negativas de ações tecnológicas humanas (desenvolvimento e implementação), o medo do resultado deve ser usado como um guia para a tomada de decisões responsáveis. Não se trata de medo descabido, mas fundamentado. Tem a ver com a projeção das consequências do ato presente a longo prazo, apoiado em informações trazidas por várias ciências com o fim de desenhar um cenário mais próximo possível daquilo que poderia acontecer caso não se mude as atitudes no presente.

Ante a ambivalência, amplitude e potenciais impactos e riscos de danos, o desenvolvimento de novas tecnologias gerais, como é o caso das tecnologias de IA, deve ser primeiramente, e antecipadamente à sua implementação, fundado na “heurística do medo”; não se trata de um princípio do medo paralisante (Sunstein, 2002), nem mesmo irracional, mas, pelo contrário, racional, pois pautado em elementos cientificamente

¹³ Conferir em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2868197&filename=PL%202338/2023

¹⁴ Para aprofundamento de ambos os termos, sugere-se a leitura de JONAS, Hans. **O Princípio Responsabilidade**: ensaio de uma ética para uma civilização tecnológica. Rio de Janeiro: PUC RIO, 2006.

fundados de responsabilidade pelo evitamento de danos futuros (Aragão, 2008), inclusive, e se necessário, a interrupção do desenvolvimento e implementação de atividade potencialmente danosa (Oliveira, 2019), condicionando o aumento da responsabilidade antecipatória.

Cabe destacar que Jonas não é contra a tecnologia. Sua proposta não implica em demonização do progresso científico, mas uma posição de crítica e de questionamento radical sobre os perigos e riscos, sobretudo os incertos, que a liberdade sem regras no desenvolvimento técnico e científico pode trazer (Lopez, 2010). A palavra medo é apenas uma estratégia heurística cujo principal objetivo é fazer com que a imaginação cientificamente amparada das consequências potencialmente negativas possua a capacidade de mobilizar o sentimento de responsabilidade coletiva e impedir a ação causadora do potencial dano futuro, mesmo que incerta (Oliveira, 2022).

O importante a se destacar é que no contexto da fase atual do desenvolvimento tecnológico, bem como no da sociedade de risco, a responsabilidade de ação passa a ser antecipatória, prospectiva e de evitamento de dano, e não propriamente de reparação do dano concretizado; fundada na ideia de que a dúvida, a imprevisibilidade e a incerteza são condições que devem orientar o desenvolvimento científico responsável (Lopez, 2010).

3.2 PRINCÍPIO DA PRECAUÇÃO COMO TRADUÇÃO JURÍDICA DO PRINCÍPIO RESPONSABILIDADE

A preocupação ética, conforme a desenvolvida por Jonas (2006), em relação aos riscos e impactos futuros e incertos decorrentes da implementação de atividades tecnológicas na atualidade possui correspondência e tradução jurídica no princípio da precaução. A conexão teórica entre ambos os campos se encontra no ponto de toque em relação à comum preocupação com a proteção contra riscos futuros incertos, porém cientificamente potenciais, bem como da perspectiva dos efeitos a longo prazo, e de que uma vez materializados podem se tornar irreversíveis.

Assim, enquanto Beck (2011, p. 278) analisa que na sociedade de risco há cada vez mais a necessidade de “antecipação de catástrofes que podem nunca ocorrer”; Jonas (2006) destaca a necessidade de uma ação baseada no princípio responsabilidade, no sentido de que as ações presentes dos agentes devem sempre ser orientadas pelas consequências futuras, mesmo que potenciais e incertas. Se Beck (2011) chama a atenção de que os riscos possuem a potencialidade de transcender barreiras temporais, afetando gerações futuras, Jonas fundamenta eticamente essa preocupação ao afirmar que “a responsabilidade corresponde ao poder, e deve ser comensurável com este”

(Jonas, 2006, p. 89). Por claro que as tecnologias de IA representam um novo poder, cada vez mais amplo e de propósito geral, derivando daí a necessidade de que a dimensão da responsabilidade deva implicar considerações não apenas em relação aos impactos imediatos, mas sobretudo a longo prazo.

A partir desta condição, a incerteza e a dúvida razoável não devem ser utilizadas como elementos autorizadores da implementação de novas tecnologias, mas, ao contrário, devem significar a necessidade de implementação de cuidado e de responsabilidade maiores, bem como de ampliação de medidas de proteção frente a potenciais danos futuros graves ou irreversíveis (Aragão, 2008), vez que em tais condições “devemos dar ouvidos mais à profecia da desgraça do que à profecia da bem-aventurança” (Jonas, 2006, p. 77).

Não se trata de elemento impeditivo do desenvolvimento tecnológico, mas uma recontextualização e ampliação da responsabilidade do agente implementador de tecnologias, antecipando e dando maior ênfase aos cuidados, conforme a tarefa de evitamento do dano posterior à implementação da tecnologia. No contexto das tecnologias de IA, essa condição demanda que se leve e se considere seriamente cenários de risco, mesmo quando a probabilidade exata dos mesmos não pode ser cientificamente calculada com exatidão.

No contexto das tecnologias de IA essa tradução torna-se particularmente relevante como responsabilidade prognóstica e precaucional. A dimensão temporal e futura da responsabilidade, conforme destaca Jonas (2006), passa a ser condição cada vez mais necessária ante prognósticos dos riscos de danos, e encontra eco também em relação aos desafios jurídicos e regulatórios da IA, com a correspondente tradução jurídica no princípio da precaução. Essa passa a ser o fundamento da ação responsável, pois “a responsabilidade corresponde ao poder, e deve ser comensurável com este” (Jonas, 2006, p. 89), sobretudo quando se trata de evitamento de resultados potencialmente danosos, mesmo que incertos.

No caso das tecnologias de IA, o poder de transformação social é enorme. Como tecnologias de aplicação geral, capazes de remodelar o mundo, certamente possuem impactos incertos. É precisamente por isso que se demanda do agente implementador uma responsabilidade proporcional e de evitamento de riscos de danos, pois “as decisões que tomamos hoje sobre o desenvolvimento da IA podem determinar o futuro da humanidade de formas que ainda não compreendemos completamente” (Russell, 2021, p. 187).

Como a responsabilidade proposta por Jonas deve ser compreendida como algo prévio à própria ação, de modo a pensar as consequências antes que o fato seja

efetivamente materializado (Oliveira, 2019), também o princípio da precaução, no campo jurídico, exige dos agentes o mesmo modo de ação responsável, até porque, em condições de riscos incertos e de impactos potencialmente irreversíveis aguardar a materialização do dano pode ser demasiado temerário na sociedade de risco (Lopez, 2010). Trata-se de uma transição da exigência de um risco ou de um dano concreto, cuja consequência é definida *ex post facto*, para a necessidade de um modelo de responsabilização antecipada, baseado na consideração de um risco abstrato, com a fixação de responsabilidades *ex ante facto* (Freitas; Colombo, 2015). Esta é a base para pensar o princípio da precaução, compreensão cada vez mais presente na seara jurídica.

Originalmente o princípio da precaução nasceu ligado a preocupações ambientais (Aragão, 2018). No caso do ordenamento jurídico brasileiro, o princípio se encontra comumente associado ao campo do Direito Ambiental, tendo como fundamento o art. 225, §1º, IV e VII da Constituição da República Federativa do Brasil de 1988, que exige do Poder Público e de toda coletividade uma atuação de controle de riscos para a preservação do meio ambiente. A exigência da precaução visa a impedir danos e prejuízos, mesmo que incertos; na dúvida, opta-se pela solução que proteja imediatamente o bem jurídico fundamental que esteja em risco de dano¹⁵ (Machado, 2013).

Pouco-a-pouco, e se solidificando como necessidade no contexto do risco fabricado (Beck, 2011), este princípio acabou por se espalhar para outros campos de preocupação jurídica, a partir da ideia de responsabilidade e de cuidado no desenvolvimento de atividades de potencialidade de impactos futuros (Aragão, 2008). Em relação às tecnologias de IA, Bioni e Luciano afirmam que o princípio da precaução pode e deve ser aplicado como base para pensar a ideia de responsabilidade do agente desenvolvedor delas, sobretudo no que concerne à prevenção de danos, derivando de tal reflexão condições que fornecem um “substrato importante para pensar medidas e estratégias de regulação de IA¹⁶, notadamente como lidar com situações de riscos de danos ou de desconhecimento dos potenciais malefícios e benefícios desse tipo de tecnologia” (2019, p. 218).

¹⁵ E no caso dos potenciais riscos provindos da implementação de tecnologias de inteligência artificial verifica-se como ponto essencial de exigência de cuidado para com a proteção à dignidade da pessoa humana, conforme fundamento estabelecido pelo PL 2338/2023.

¹⁶ Bioni e Luciano (2019, p. 225) observam que “a aplicação do princípio da precaução às tecnologias de IA demanda a criação de instrumentos regulatórios capazes de operar em condições de incerteza radical, antecipando e mitigando riscos cujas dimensões plenas só se revelarão no futuro”.

3.3 EXEMPLIFICAÇÃO NORMATIVA: UMA LEITURA DA PL 2338/2023 A PARTIR DA PERSPECTIVA DO PRINCÍPIO DA PRECAUÇÃO

Embora originário do direito ambiental, o princípio da precaução acabou por se espalhar em outros campos normativos que buscam lidar com as incertezas e riscos do desenvolvimento tecnológicos no contexto da sociedade de risco. O mesmo vem ocorrendo em relação às preocupações envolvidas ao desenvolvimento de tecnologias de IA, sobretudo das discussões sobre governança responsável em relação aos potenciais riscos e da responsabilidade antecipatória dos agentes; condição esta que se apresenta e está igualmente no cerne do recente Projeto de Lei 2338/2023.

No projeto proposto se verifica que o legislador não deixa de conhecer a condição de incertezas características dos riscos do desenvolvimento das tecnologias de IA. A proposta de tradução jurídica dos contextos sociais e do controle dos riscos são categorizados em dois principais tipos: risco excessivo e alto risco. Em todo o caso, os riscos determinam a ideia de responsabilidade do agente. Por sua vez, verifica-se papel preponderante da precaução como princípio regulador geral da governança e da gestão de riscos, condição que se materializa em institutos e exigências de ação previstos na PL e direcionadas aos agentes desenvolvedores de IA, mesmo antes da colocação da tecnologia no mercado, o que evidencia a ênfase na responsabilidade antecipatória pelo evitamento de danos futuros.

Observa-se que o princípio da precaução aparece explicitamente no projeto de lei como elementos fundamental, de necessidade de observação geral e norteador do desenvolvimento, implementação e uso dos sistemas de IA (art. 3, XI). Nota-se que o legislador reconhece a dinâmica própria dos riscos ligados ao desenvolvimento de tecnologias na atualidade e a necessidade de não apenas prevenir, condição relativa aos riscos certos e previsíveis, mas também de encontrar mecanismos jurídicos para lidar com condições de incertezas ante potenciais danos, sobretudo no contexto da sociedade de risco.

Em relação à necessidade de ação antecipatória e precaucional, bem como ao evitamento de riscos de danos futuros, o artigo 3º do PL 2338/2023, para além da literalidade da precaução, ainda determina como princípios de responsabilidade antecipatória ações de implementação e necessidade de supervisão e determinação humana efetiva e adequada no ciclo de vida da IA, considerando o grau de risco envolvido (inc. III); a necessidade de transparência e explicabilidade dos processos (inc. VI); a diligência devida e auditabilidade ao longo de todo o ciclo de vida do sistema de IA, de acordo com o risco envolvido e o estado da arte do desenvolvimento tecnológico (inc. VII); a não maleficência (inc. XII); o desenvolvimento e uso ético e responsável da IA (inc. XIII); e a governança transparente, participativa e orientada à proteção de direitos fundamentais individuais, sociais, coletivos e econômicos (inc. XIV).

Atentando-se à manifestação explícita da precaução como princípio fundamental constante no Projeto (artigo 3º, inciso XI), demandando do agente desenvolvedor a “prevenção, precaução e mitigação de riscos e danos”. Verifica-se um escalonamento necessário: a prevenção, quando a capacidade de prever com certeza científica a emergência de danos, a precaução, ante condições de incerteza científica, e a mitigação, que se consolida na redução de eventuais impactos.

Analisando especificamente a responsabilidade na gestão e no evitamento de danos em contextos de incertezas, o fundamento teórico do princípio da precaução pressupõe a tomada de ação em situações de incerteza cientificamente fundamentadas, ou seja, não se exige a certeza, nexos causais definitivos, mas o indício plausível e a probabilidade indicativa do evento danoso em relação ao potencial de ocorrência de riscos graves (Aragão, 2008), exigindo-se do agente responsabilidades antecipatórias.

Observa-se que o PL apresenta instrumentos e exige responsabilidade dos agentes desenvolvedores de tecnologias de IA em relação à gestão dos riscos baseados nos fundamentos do princípio da precaução. A responsabilidade aparece baseada e diferenciada conforme a proposta da classificação de risco: médio e alto risco. Compreende-se que a precaução se aplica de forma direta em contexto de alto risco, no qual as possibilidades de previsibilidade científica em relação aos resultados podem se tornar incertas.

Há o reconhecimento dessa dimensão a exemplo do constante do art. 15, inc. VIII, que determina que a classificação de alto risco deve considerar “a extensão e probabilidade dos riscos do sistema de IA, incluindo as medidas de mitigação adotadas”. Compreende-se que tal previsão normativa pauta a responsabilidade ampliada do agente, com base na precaução, onde se exige uma responsabilidade antecipatória do evitamento de danos, mesmo frente a evidências empíricas limitadas sobre impactos futuros.

Considera-se que outra forma de materializar normativamente o princípio da precaução é o exemplo contido no art. 13 do projeto, que estabelece a vedação do desenvolvimento, implementação e uso de sistemas de IA frente a riscos denominados como excessivos; à interpretação desta pesquisa, conjugando-se aos fundamentos sobre o princípio responsabilidade e princípio da precaução desenvolvidos acima, a excessividade pode se dar tanto em condições de previsibilidade quanto de imprevisibilidade, ou seja, de incerteza científica, porém com apontamentos de ameaças e de danos graves ou potencialmente irreversíveis.

Mas salienta-se, não se trata da aplicação da precaução apenas em contextos de danos excessivos, mas em contextos de possíveis “danos irreversíveis” e em condições de incerteza; daí que cuida o art. 8º, parágrafo único, das demais condições de aplicabilidade

da precaução, exigindo do implementador da tecnologia de IA supervisão humana de todos os processos. Daí que se julga ser a precaução, reconhecida como princípio, a base fundamental na qual se assenta o PL, bem como a ideia e instrumentos de gestão de riscos sob a responsabilidade do agente desenvolvedor de tecnologias de IA.

Em relação à exigência instrumental, entende-se que a precaução serve como base da responsabilidade dos agentes em relação a instrumentos jurídicos presentes no projeto, como a avaliação preliminar, sendo definida no art. 4º, inc. XV como um “processo simplificado de autoavaliação, anterior à utilização ou colocação no mercado de um ou mais sistemas de IA, para classificação de seu grau de risco, com o objetivo de determinar o cumprimento das obrigações definidas nesta Lei”, e normatizada sua operacionalização no art. 12; e a avaliação de impacto algorítmico, constituindo-se como “análise do impacto sobre os direitos fundamentais, apresentando medidas preventivas, mitigadoras e de reversão dos impactos negativos, bem como medidas potencializadoras dos impactos positivos de um sistema de IA” (inc. XVI); neste caso, constitui-se obrigação de que na incerteza sobre a impossibilidade de eliminar ou mitigar riscos associados ao sistema de IA, mesmo que potenciais, a utilização deverá ser descontinuada (Art. 25 e seguintes).

Antes de encerrar a análise sobre a precaução como fundamento do PL, um instrumento que se considera essencial na prevenção de danos, inclusive dos imprevistos por mecanismos científicos e técnicos, é a previsão dos *sandboxes* regulatórios, no artigo 55, como a ideia de responsabilidade do agente desenvolvedor em relação à gestão do risco a partir do desenvolvimento de um ambiente regulatório experimental. Trata-se de um ambiente controlado onde o desenvolvimento de inovação pode ocorrer sob supervisão, permitindo a identificação, antecipação e mitigação de riscos antes da implementação em larga escala.

Este teste inicial e controlado é fundamental sobretudo em relação ao controle dos riscos incertos e de impactos graves e irreversíveis, buscando evitá-los, criando espaços seguros para experimentação tecnológica, antecipatória à introdução da tecnologia no mercado. No contexto da precaução, que fundamenta a ideia do PL, julga-se que se identificadas incertezas em relação a riscos futuros, mas potenciais, determinado sistema de IA não deve ser implementado até que se minore ou evite tal potencialidade. A ideia por trás do *sandbox* é a de fornecer um espaço seguro para que empresas e órgãos reguladores trabalhem juntos para entender como novas tecnologias podem ser desenvolvidas e regulamentadas de forma responsável e ética.

Os instrumentos e mecanismos apresentados, constantes no PL, de caráter exemplificativo e não esgotável no presente trabalho, serviram como elementos

demonstrativos de como a preocupação ética com os impactos futuros e indesejados das tecnologias de IA, sobretudo da preocupação em relação à responsabilidade antecipatória dos agentes para o evitamento de danos futuros e incertos, podem ser traduzidos em normativas jurídicas a partir dos fundamentos do princípio da precaução e, por consequência, da exigência da necessidade de ação precaucional do agente desenvolvedor de tecnologias de IA. Igualmente, a proposta legislativa encontra consonância no imperativo ético necessário ao desenvolvimento de tecnologias na sociedade de risco, e da necessária responsabilidade a partir da materialização jurídica.

CONSIDERAÇÕES FINAIS

Conforme apresentado na introdução, o problema de pesquisa questionava sobre como é possível pensar padrões éticos e normativos baseados na responsabilidade para o desenvolvimento tecnológico adequado das tecnologias de IA. Após o percurso argumentativo desenvolvido é possível considerar que a resposta à questão se encontra na necessária articulação entre responsabilidade ética antecipatória, emergente da sociedade de risco, e em sua tradução jurídica no princípio da precaução, conforme a reflexão desenvolvida ao longo do estudo, bem como presente como elemento central no PL 2338/2023.

A partir das considerações acerca da sociedade de risco, sobretudo do risco fabricado, conforme análises desenvolvidas por Beck, é possível indicar a necessidade da emergência de responsabilidades dos agentes tanto na esfera ética quanto jurídica nesse contexto. No campo da responsabilidade ética, as contribuições desenvolvidas por Jonas são fundamentais para pensar a antecipação e o evitamento de danos, sobretudo ante sua característica fundamental da irreversibilidade e dos impactos profundos nas estruturas sociais. Essa necessidade é cada vez mais latente e presente igualmente no campo da normatividade jurídica, cuja tradução aparece no princípio da precaução. A reflexão sobre a articulação de ambas as dimensões enriquece a reflexão sobre a ideia de que tipo de responsabilidade cabe aos agentes desenvolvedores e implementadores de tecnologias de IA e contribui para desafios emergentes, sobretudo em relação ao evitamento de danos e de impactos socialmente indesejáveis.

Por fim, os elementos trazidos à discussão a partir do PL 2338/2023, embora exemplificativos, são capazes de evidenciar que o legislador se encontra, a nosso entender, atento ao contexto dos riscos, de seus desafios, bem como da necessidade da precaução como elemento fundamental, tanto da responsabilidade jurídica quanto ética, para o desenvolvimento adequado das tecnologias, de modo a exigir

do agente desenvolvedor uma série de exigências de governança antecipatórias, a exemplo das avaliações de impacto e da implementação de sandboxes regulatórios, condições antecipatórias à colocação de qualquer tecnologia no mercado, necessárias à antecipação e à prevenção de danos.

Cabe ressaltar que a presente pesquisa possuiu cunho eminentemente teórico, conforme exposto inicialmente. Ante esta condição, considera-se que seja valiosa a possibilidade de continuidade de pesquisas exploratórias, derivadas dos fundamentos aqui expostos, para analisar perspectivas e fundamentos práticos sobre a regulação e da precaução, como a implementação dos dispositivos e instrumentos previstos no PL, bem como da própria interpretação e aplicação práticas concernente à precaução, aplicada às tecnologias de IA após a aprovação do projeto de Lei.

No entanto, uma consideração final que não se afasta do contexto das novas tecnologias é a de que as ações dos desenvolvedores e implementadores de tecnologias de IA deve estar pautada na responsabilidade pelo evitamento de danos futuros, imprevistos e incertos, questão que se fundamenta necessariamente na articulação entre normativas éticas e jurídicas, e que possui tradução e encontra articulação no Projeto de Lei 2338/2023.

REFERÊNCIAS

- ARAGÃO, Alexandra. Princípio da Precaução: manual de instruções. **Revista do Centro de Estudos Direito do Ordenamento, do Urbanismo e do Ambiente**, Coimbra, v. 22, ano XI, v. 2. p. 9-57. 2008.
- BARBOSA, Karlos Alves. **Sociedade de risco e os crimes de perigo abstrato**. 2012. 190 f. Dissertação (Mestrado em Ciências Sociais Aplicadas) - Universidade Federal de Uberlândia, Uberlândia, 2012.
- BECK, Ulrich. **Sociedade de Risco: Rumo a uma outra modernidade**. 2. ed. São Paulo: Editora 34, 2011.
- BIONI, Bruno Ricardo; LUCIANO, Maria. O princípio da precaução na regulação da inteligência artificial: seriam as leis de proteção de dados o seu portal de entrada. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Org.). **Inteligência artificial e direito: ética, regulação e responsabilidade**. São Paulo: Thomson Reuters Brasil, p. 207-232, 2019.
- BOTTINI, Pierpaolo Cruz. **Crimes de perigo abstrato e princípio da precaução na sociedade de risco**. São Paulo: Revista dos Tribunais, 2007.
- BRANDÃO, Rodrigo. O cenário atual de desenvolvimento da Inteligência Artificial no Brasil. **Panorama Setorial da Internet**, v. 1, n. 16, maio, 2024. Disponível em: <https://www.cgi.br/media/docs/publicacoes/6/20240514085112/psi-ano-xvi-n-1-desenvolvimento-ia-brasil.pdf>. Acesso em: 03 out. 2025.
- BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 3 out. 2025.
- BRASIL. Projeto de Lei n. 2338, de 2023. Dispõe sobre o desenvolvimento, o fomento e o uso ético e responsável da inteligência artificial com base na centralidade da pessoa humana. Brasília, DF: **Câmara dos Deputados**, 2023. Disponível em: https://www.camara.leg.br/proposicoesWeb/prop_mostrarintegra?codteor=2868197&filename=PL%202338/2023. Acesso em: 3 out. 2025.
- CARVALHO, André Carlos Ponce de Leon Ferreira de. Inteligência Artificial: riscos, benefícios e uso responsável. **Estudos Avançados**, São Paulo, v. 35, n. 101, p. 33-40, jan./abr. 2021. Disponível em: <https://www.scielo.br/j/ea/a/ZnKyrCrLVqzhZbXGgXTwDtn/?format=pdf&lang=pt>. Acesso em: 3 out. 2025.
- CARVALHO, Deelton Winter. **Dano Ambiental Futuro: A Responsabilização Civil Pelo Risco Ambiental**. Porto Alegre: Livraria do Advogado, 2013.
- CHIAVONI, Tayrone Marquesini. **Sociedade de risco digital: O princípio da Precaução na regulação da Inteligência Artificial**. 2022. 126 f. Dissertação (Mestrado em Direito na especialidade de Direito Internacional e Europeu) – Faculdade de Direito da Universidade de Lisboa, Lisboa, 2022. Disponível em: https://run.unl.pt/bitstream/10362/148919/1/Chiavone_2022.pdf. Acesso em: 3 out. 2025.
- DE GIORGI, Raffaele. O risco na sociedade contemporânea. **Seqüência Estudos Jurídicos e Políticos**, Florianópolis, v. 15, n. 28, p. 45–54, 1994. DOI: 10.5007/%x. Disponível em: <https://periodicos.ufsc.br/index.php/sequencia/article/view/15873>. Acesso em: 3 out. 2025.
- FRAZÃO, Ana. Classificação de riscos: a solução adotada pelo PL 2338/23. In: JOTA. **Coluna IA, Regulação e Democracia**. São Paulo, 4 abr. 2024. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/ia-regulacao-democracia/classificacao-de-riscos-a-solucao-adotada-pelo-pl-2338-23>. Acesso em: 03 out. 2025.

FREITAS, V. P.; COLOMBO, S. Da teoria do risco concreto à teoria do risco abstrato na sociedade pós-industrial: um estudo da sua aplicação no âmbito do direito ambiental. **REVISTA QUAESTIO IURIS**, [S. l.], v. 8, n. 03, p. 1895–1912, 2015. DOI: 10.12957/rqi.2015.18820

HARARI, Yuval. **21 lições para o século 21**. São Paulo: Companhia das letras, 2018.

INSTITUTO DE TECNOLOGIA E SOCIEDADE – ITS. **100 IA**: PL 2.338/2023 e a classificação de risco dos usos de IA sob uma perspectiva prática. Rio de Janeiro: ITS, 2024. Disponível em: https://itsrio.org/wp-content/uploads/2016/12/Relatorio_100IA.pdf. Acesso em: 03 out. 2025.

IQBAL, Sajid. The Intelligence Spectrum: Unraveling the Path from ANI to ASI. **Journal of Computing & Biomedical Informatics**, [S. l.], v. 7, n. 02, 2024. Disponível em: <https://jcbi.org/index.php/Main/article/view/779>. Acesso em: 3 out. 2025.

JONAS, H. **O Princípio Responsabilidade**: ensaio de uma ética para uma civilização tecnológica. Rio de Janeiro: PUC RIO, 2006.

JONAS, H. **Técnica, medicina e ética**: sobre a prática do princípio responsabilidade. São Paulo: Paulus, 2013.

KAUFMAN, D. **Desmistificando a inteligência artificial**. Belo Horizonte: Autêntica, 2022.

KOKOTAJLO, Daniel et al. AI 2027. [S. l.]: **AI Futures Project**, 2025. 23 p. Disponível em: <https://ai-2027.com/ai-2027.pdf>. Acesso em: 3 out. 2025.

LOPES, Wendell. A ética da responsabilidade é um caminho para encarar a crise ecológica e dilemas de nosso tempo. Entrevista concedida a Ricardo Machado. **IHU Online**, São Leopoldo, n. 540, ano XIX, p. 1-2, 2 set. 2019. Disponível em: <https://www.ihuonline.unisinos.br/artigo/7640-a-etica-da-responsabilidade-e-um-caminho-para-encarar-a-crise-ecologica-e-dilemas-de-nosso-tempo>. Acesso em: 3 out. 2025.

LOPEZ, T. A. **Princípio da Precaução e evolução da responsabilidade civil**. São Paulo: Quartier Latin, 2010.

MACHADO, Marta Rodríguez de Assis. **Sociedade do risco e direito penal: uma avaliação de novas tendências político-criminais**. São Paulo: IBCCRIM, 2005.

MACHADO, Paulo Affonso Leme. **Direito Ambiental Brasileiro**. 21. ed. São Paulo: Malheiros, 2013.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY - MIT. AI Risk Database explainer. FutureTech Group. 2025. Disponível em: <https://airisk.mit.edu>.

MCCARTHY, John; MINSKY, Leon; ROCHESTER, Nathanael; SHANNON, Claude. **A proposal for the Dartmouth Summer Research Project on Artificial Intelligence**. 1955, p. 1-13. Disponível em: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. Acesso em: 20 jul. 2021.

MIT FUTURETECH. **The MIT AI Risk Repository**. Cambridge, MA: MIT FutureTech, 2025. Disponível em: <https://airisk.mit.edu/>. Acesso em: 20 out. 2025.

FRAZÃO, Ana; MULHOLLAND, Caitlin (Coord.). **Inteligência artificial e direito: ética, regulação e responsabilidade**. São Paulo: Thomson Reuters Brasil, 2019.

NONATO, Luiza Gimenez. **Relações de poder na era da Inteligência Artificial (IA)**: a competição estratégica entre Estados Unidos e China pela liderança da IA. 2023. Tese (Doutorado) – Universidade de São Paulo, São Paulo, 2023. Disponível em: <https://www.teses.usp.br/teses/disponiveis/101/101131/tde-28062023-160904/>. Acesso em: 03 out. 2025.

OLIVEIRA, Jelson Roberto de. A heurística do temor e o despertar da responsabilidade. Entrevista concedida à IHU On-Line. **IHU Online**, São Leopoldo, n. 371, ano XI, 29 ago. 2011. Disponível em: <https://www.ihuonline.unisinos.br/artigo/4035-jelson-roberto-de-oliveira>. Acesso em: 3 out. 2025.

OLIVEIRA, Jelson de. Ciência caolha é aquela que transforma a nobreza do conhecimento em utilidade do saber. **IHU Online**, São Leopoldo, n. 540, ano XIX, 2 set. 2019. Disponível em: <https://www.ihuonline.unisinos.br/artigo/7638-ciencia-caolha-e-aquela-que-transforma-a-nobreza-do-conhecimento-em-utilidade-do-saber>. Acesso em: 3 out. 2025.

OLIVEIRA, Jelson Roberto de. Para uma ethical turn da tecnologia: por que Hans Jonas não é um tecnofóbico. **Trans/Form/Ação**, Marília, SP, v. 45, n. 2, p. 191–206, 2022. DOI: 10.1590/0101-3173.2022.v45n2.p191

RUSSELL, Stuart. **Inteligência artificial a nosso favor**: Como manter o controle sobre a tecnologia. São Paulo: Companhia das letras, 2021.

SLATTERY, P.; SAERI, A. K.; GRUNDY, E. A. C.; GRAHAM, J.; NOETEL, M.; UUK, R.; DAO, J.; POUR, S.; CASPER, S.; THOMPSON, N. The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence. **ArXiv**, 14 ago. 2024. Disponível em: <https://arxiv.org/abs/2408.12622>. Acesso em: 3 out. 2025.

STANFORD UNIVERSITY. **Institute for Human-Centered Artificial Intelligence (HAI)**. The AI Risk Repository: A Comprehensive Meta-Review, Database, and Taxonomy of Risks From Artificial Intelligence. Palo Alto, CA: Stanford University, 2024. Disponível em: <https://hai.stanford.edu/ai-index/2024-ai-index-report>. Acesso em: 3 out. 2025.

SUNSTEIN, Cass R. Risk and Reason: Safety, Law, and the Environment. Cambridge, UK: **Cambridge University Press**, 2002. Disponível em: <https://assets.cambridge.org/97805217/91991/sample/9780521791991ws.pdf>. Acesso em: 3 out. 2025.

UNIÃO EUROPEIA. Parlamento Europeu e Conselho. **Regulamento (UE) 2024/1689 do Parlamento Europeu e do Conselho, de 13 de junho de 2024, que cria regras harmonizadas em matéria de inteligência artificial...** (Regulamento da Inteligência Artificial). Bruxelas: EUR-Lex, 2024. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX:32024R1689>. Acesso em: 20 out. 2025.

WALDMAN, Ricardo Léo; SAMPAIO, Vinícius Barbosa; MUNHOZ, Mônica Guimarães Veloso. O princípio da precaução e o princípio de responsabilidade de Han Jonas. **Quaestio Iuris**, Rio de Janeiro, v. 10, n. 1, p. 199-218, 2018.