Relações Internacionais do Mundo Atual - unicuritiba

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

ALGORITMOS, DISCRIMINAÇÃO E RISCOS EXISTENCIAIS DA INTELIGÊNCIA ARTIFICIAL: O PRINCÍPIO DA PRECAUÇÃO COMO OBRIGAÇÃO POSITIVA NO DIREITO INTERNACIONAL DOS DIREITOS HUMANOS

ALGORITHMS, DISCRIMINATION, AND EXISTENTIAL RISKS OF ARTIFICIAL INTELIGENCE: THE PRECAUTIONARY PRINCIPLE AS A POSITIVE OBLIGATION IN INTERNATIONAL HUMAN RIGHTS LAW

MARIA CELESTE CORDEIRO LEITE DOS SANTOS

Professora Livre Docente em Direito Penal pela USP, Professora Doutora em Filosofia do Direito pela PUC-SP, Mestre em Direito das Relações Sociais, Professora do Programa de Pós-Graduação em Direito (Mestrado e Doutorado) da PUC-SP, Coordenadora do Núcleo de Pesquisas em Percepções Cognitivas na Interpretação da Norma na PUC-SP, Advogada.

FABIO RIVELLI

Doutorando e Mestre em Direito pela PUC-SP, Especialista em Administração de Contencioso de Volume pela GVlaw - Fundação Getúlio Vargas (FGV/SP), MBA pelo INSPER-SP, Pesquisador do Núcleo de Pesquisas em Percepções Cognitivas na Interpretação da Norma da PUC-SP, Advogado e sócio do escritório Lee, Brock, Camargo Advogados.

MARCO ANTÔNIO LIMA DA CRUZ FILHO

Doutorando e Mestre em Direito Internacional pela PUC-SP, Advogado, Especialista, Membro das Comissões de Direito Internacional e Direitos Humanos da OAB/SP.

RESUMO:
Objetivo: analisar a crise estrutural do Direito Internacional dos Direitos Humanos ante os riscos existenciais e a discriminação algorítmica da IA, e demonstrar como o princípio da precaução emerge como obrigação positiva essencial para a sua efetiva governança. O estudo visa propor caminhos para a construção de arcabouço regulatório que traduza as obrigações de proteção em garantias jurídicas concretas

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

para assegurar o a dignidade humana e os direitos fundamentais na era da IA. **Metodologia**: utiliza-se o método dedutivo e adota-se abordagem qualitativa, de natureza exploratória e descritiva, busca-se aprofundar a compreensão sobre os desafios impostos pela IA ao Direito Internacional dos Direitos Humanos. Emprega-se a técnica de revisão bibliográfica, que abrangerá literatura científica especializada, documentos normativos de organizações internacionais, tratados e convenções, bem como jurisprudência relevante sobre IA, direitos humanos e o princípio da precaução. **Resultados**: a proteção efetiva do direito à vida e à dignidade impõem deveres de diligência reforçada em face de riscos emergentes — sejam ambientais, tecnológicos ou híbridos. Essa orientação deve servir como parâmetro para o reconhecimento da IA como objeto de obrigações internacionais de cuidado. A discriminação algorítmica constitui o ponto mais visível da crise de legitimidade do Direito ante a automação. A crença na neutralidade técnica mascara a reprodução digital de hierarquias raciais, de gênero e socioeconômicas, e converte a desigualdade em estatística e o preconceito em cálculo. **Contribuições**: o estudo oferece contribuição à comunidade acadêmica ao expor o dilema estrutural do Direito Internacional dos Direitos Humanos diante da Inteligência Artificial, ao propor o princípio da precaução como imperativo ético e jurídico para mitigar riscos existenciais e a discriminação algorítmica. Para a sociedade civil, o estudo realça a urgência de governança global eficaz e mecanismos de responsabilização, essenciais para transformar a precaução em justiça e garantir que o avanço tecnológico preserve a dignidade humana e os direitos fundamentais, enfrentando os desafios impostos pela opacidade e velocidade do dano algorítmico.

Palavras-chave: Inteligência artificial; Direitos humanos; Princípio da precaução; Discriminação algorítmica; Risco existencial.

ABSTRACT:
Objective*: to analyze the structural crisis of International Human Rights Law in view of existential risks and algorithmic discrimination of Artificial Intelligence (AI), and to demonstrate how the precautionary principle emerges as a positive obligation essential for its effective governance. The study aims to propose ways to build a regulatory framework that translates protection obligations into concrete legal guarantees to ensure human dignity and fundamental rights in the age of AI.* ***Methodology****: the deductive method is used and a qualitative approach is adopted, of an exploratory and descriptive nature, seeking to deepen the understanding of the challenges imposed by AI on International Human Rights Law. The bibliographic review technique is used, which will cover specialized scientific literature, normative documents from international organizations, treaties and conventions, as well as relevant jurisprudence on AI, human rights and the precautionary principle.* ***Results****: The effective protection of the right to life and dignity imposes enhanced due diligence duties in the face of emerging risks – whether environmental, technological or hybrid. This guidance should serve as a parameter for the recognition of AI as an object of international care obligations. Algorithmic discrimination constitutes the most visible point of the crisis of legitimacy of the Law in the face of automation. The belief in technical neutrality masks the digital reproduction of racial, gender, and socioeconomic hierarchies, and converts inequality into statistics and prejudice into calculation.* ***Contributions****: the study offers a contribution to the academic community by exposing the structural dilemma of International Human Rights Law in the face of AI, by proposing the precautionary principle as an ethical and legal imperative to mitigate existential risks and algorithmic*

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

discrimination. For civil society, the study highlights the urgency of effective global governance and accountability mechanisms, essential to transform precaution into justice and ensure that technological advancement preserves human dignity and fundamental rights, addressing the challenges posed by the opacity and speed of algorithmic harm.

Keywords: *Artificial intelligence; Human rights; Precautionary principle; Algorithmic discrimination; Existential risk.*

1 INTRODUÇÃO

O avanço exponencial dos sistemas de Inteligência Artificial (IA) representa uma das inovações mais disruptivas da história humana, com potencial para transformar fundamentalmente todas as esferas da vida social, econômica e política; contudo, essa promessa de progresso tecnológico é acompanhada por riscos igualmente inéditos, que incluem ameaças existenciais não triviais e a perpetuação de vieses discriminatórios profundamente enraizados em dados históricos. A natureza transversal da IA, com impactos difusos e sistêmicos, exige uma reavaliação urgente dos paradigmas regulatórios existentes.

Diante desse cenário complexo, o Direito Internacional dos Direitos Humanos (DIDH) se depara com um dilema estrutural significativo. Historicamente fundamentado em premissas liberais e marcado por uma inerente indeterminação argumentativa, o DIDH revela-se insuficiente para oferecer respostas eficazes e ágeis. Sua incapacidade de lidar com a opacidade técnica inerente aos sistemas de IA e a velocidade do dano algorítmico, que desafiam as estruturas tradicionais de responsabilização, expõe fragilidades que precisam ser superadas.

Nesse contexto desafiador, o princípio da precaução emerge como imperativo ético e jurídico fundamental. Sua aplicação, vinculada tradicionalmente a riscos ambientais, expande-se para mitigar os riscos existenciais e estruturais impostos pela IA, exigindo ação preventiva mesmo na ausência de certeza científica absoluta. Este princípio oferece um caminho para antecipar, prevenir e minimizar os efeitos adversos da tecnologia em rápida evolução, e posiciona-se como pilar essencial para a governança responsável e proativa da IA.

No entanto, a ausência de arcabouço regulatório global coeso e a fragilidade dos mecanismos de responsabilização impedem a concretização efetiva dessas

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

obrigações positivas. A *corrida armamentista de IA*, impulsionada por incentivos competitivos, dificulta a colaboração internacional e a implementação de freios unilaterais, expõe o viés estrutural da argumentação jurídica. Essa lacuna impede a tradução das obrigações de proteção dos direitos humanos em garantias judiciais concretas e deixa a humanidade vulnerável.

Como problema da pesquisa, entende-se que o Direito Internacional dos Direitos Humanos enfrenta um dilema estrutural crítico ante o avanço exponencial da IA, especialmente em relação aos riscos existenciais e à discriminação algorítmica. Sua fundação em premissas liberais e a indeterminação argumentativa limitam sua eficácia diante da opacidade técnica e da velocidade do dano algorítmico. A ausência de um arcabouço regulatório global robusto e de mecanismos de responsabilização eficazes impede a tradução de obrigações positivas em garantias judiciais. Adicionalmente, a aplicação de direitos fundamentais a sistemas complexos de IA revela um viés estrutural na argumentação jurídica internacional, forçando escolhas políticas em cenários de incerteza científica.

Diante da insuficiência do Direito Internacional dos Direitos Humanos em responder aos riscos existenciais e à discriminação algorítmica da IA, como o princípio da precaução pode ser efetivamente integrado e operacionalizado como obrigação positiva no âmbito do Direito Internacional? Quais ajustes normativos, institucionais e jurisprudenciais são necessários para mitigar os impactos da IA e garantir a proteção da dignidade humana e dos direitos fundamentais na era digital?

O objetivo geral da pesquisas é analisar a crise estrutural do Direito Internacional dos Direitos Humanos ante os riscos existenciais e a discriminação algorítmica da IA, e demonstrar como o princípio da precaução emerge como obrigação positiva essencial para a sua efetiva governança. O estudo visa propor caminhos para a construção de arcabouço regulatório que traduza as obrigações de proteção em garantias jurídicas concretas para assegurar o a dignidade humana e os direitos fundamentais na era da IA.

Os objetivos específicos compreendem: (i) examinar as limitações estruturais do DIDH: investigar a inadequação do DIDH, com suas premissas liberais e indeterminação argumentativa, para enfrentar a opacidade técnica e a velocidade do dano algorítmico da IA; (ii) d**emonstrar o princípio da precaução como imperativo jurídico:** analisa-lo como obrigação ética e legal para mitigar riscos existenciais e estruturais da IA, enfatizando a ação preventiva; (iii) a**nalisar as**

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

consequências da lacuna regulatória global: avaliar como a ausência de regulação global e a fragilidade dos mecanismos de responsabilização impedem a efetivação das obrigações positivas de proteção; e (iv) i**nvestigar a discriminação algorítmica e o viés estrutural da IA:** discutir como os sistemas de IA reproduzem e amplificam preconceitos históricos, resultando em discriminação algorítmica e reificando desigualdades.

A pesquisa tem metodologia dedutiva e adotará abordagem qualitativa, de natureza exploratória e descritiva, buscará aprofundar a compreensão sobre os desafios impostos pela IA ao Direito Internacional dos Direitos Humanos. Será empregada a técnica de revisão bibliográfica, que abrangerá literatura científica especializada, documentos normativos de organizações internacionais, tratados e convenções, bem como jurisprudência relevante sobre IA, direitos humanos e o princípio da precaução. O estudo analisará criticamente a argumentação jurídica internacional, com foco na capacidade de adaptação do Direito ante a opacidade tecnológica e os riscos existenciais e discriminatórios da IA.

As técnicas de pesquisa envolverão a análise de conteúdo e documental dos materiais selecionados. Realizar-se-á uma leitura crítica de relatórios de organismos internacionais, atos legislativos (como a Lei da IA da União Europeia e a Convenção-Quadro do Conselho da Europa) e artigos acadêmicos.

2 INTELIGÊNCIA ARTIFICIAL E GOVERNANÇA

O desenvolvimento da Inteligência Artificial (IA) carrega promessa imensa de inovação e transformação sem precedentes na história humana graças ao desenvolvimento tecnológico. Seu impacto tecnológico pode ser comparado ao de grandes marcos civilizatórios, desde o fogo, a roda, as revoluções industriais, luz elétrica, automóvel, computador pessoal entre outras inovações marcantes.

A promessa de inovação e transformação é acompanhada de riscos igualmente inéditos, em que muitos especialistas alertam para a existência de uma chance não trivial de que a IA possa representar ameaça existencial para a humanidade. Pesquisas indicam que mais de um terço dos pesquisadores em IA acreditam que esses sistemas poderiam desencadear uma catástrofe comparável a uma guerra nuclear total neste século (GRACE, 2024). A percepção da magnitude do dano potencial, mesmo na ausência de consenso científico exato sobre a probabilidade do risco, impõe a obrigação legal internacional aos Estados de regulamentar o desenvolvimento da IA.

A urgência dessa regulamentação, cuja eficácia é passível de questionamento, decorre do fato de que as iniciativas atuais, tanto em plano nacional quanto internacional, falham em confrontar a IA como ameaça existencial. Em vez disso, concentram-se em riscos parciais e fragmentados —

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

segurança do consumidor, cibersegurança, proteção de dados, privacidade, saúde pública, não discriminação, liberdade de expressão dentre outros. A questão central é que a IA é tecnologia transversal a todas as esferas da vida social, com impacto mais difuso do que o da energia elétrica, que foi o alicerce para o patamar de desenvolvimento que se alcançou. O alerta mostra a insuficiência dos modelos regulatórios atuais e reforça a necessidade de enquadramento jurídico mais abrangente e sistêmico:

> No entanto, a ação regulatória para mitigar a ameaça potencialmente catastrófica representada pelo desenvolvimento da inteligência de máquinas tem sido lenta para se desenvolver. As iniciativas regulatórias existentes não abordam a IA como uma ameaça existencial, mas focam fragmentadamente em riscos específicos relacionados à IA, como saúde pública, segurança do consumidor, não discriminação, privacidade, proteção de dados e liberdade de expressão (nossa tradução) (DRUZIN, BOUTE, RAMSDEM, 2025)[1].

Exemplos de inadequação regulatória incluem a Lei da Inteligência Artificial da União Europeia, que, embora estabeleça obrigações graduadas conforme graus de risco e proíba sistemas classificados como de risco inaceitável, não enfrenta os problemas estruturais relacionados à ausência de governança global e ao uso estratégico da tecnologia. De modo semelhante, a Convenção-Quadro do Conselho da Europa sobre IA, adotada em maio de 2024, constitui o primeiro tratado internacional vinculativo juridicamente sobre o tema, buscando assegurar a compatibilidade dos sistemas com os direitos humanos, a democracia e o Estado de Direito. Todavia, não aborda de maneira direta os riscos derivados da instrumentalização militar da IA nem da crescente competição geopolítica em torno de seu desenvolvimento.

Parte significativa da literatura internacional enfatiza cenários de explosão de inteligência e da chamada *Artificial General Intelligence* (AGI), definidos como sistemas dotados de capacidades cognitivas iguais ou superiores às humanas e potencialmente capazes de autoaperfeiçoamento recursivo (BOSTROM, 2016; RUSSELL, 2019; HENDRYCKS *et al.*, 2023). Tais narrativas projetam riscos existenciais baseados na hipótese de superinteligência emergente, autônoma e incontrolável. No entanto, essa hipótese é altamente especulativa e tem sido objeto de contestação, com estudos recentes que apontam probabilidade extremamente baixa de ocorrência em horizontes previsíveis.

A presente análise afasta-se dessa linha argumentativa. O risco considerado mais grave não decorre de uma eventual consciência da máquina, mas da ação humana. O verdadeiro perigo está no avanço descoordenado de uma tecnologia transversal que já permeia todos os setores da vida social. A ausência de governança internacional efetiva, o emprego militar da IA, a opacidade das grandes corporações e a incapacidade de estabelecer mecanismos regulatórios globais de contenção configuram um cenário de descontrole humano — e não de insurgência algorítmica. Como problema

[1] Texto original em inglês: *"However, regulatory action to mitigate the potential catastrophic threat posed by the development of machine intelligence has been slow to develop. Existing regulatory initiatives do not address AI as an existential threat but instead focus piecemeal on discrete AI-related risks such as public health, consumer safety, non-discrimination, privacy, data protection, and freedom of expression*." DRUZIN; BOUTE; RAMSDEN, 2025)

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

central, a tecnologia avança em velocidade e escala inéditas, sem que existam instrumentos jurídicos e institucionais capazes de disciplinar sua aplicação estratégica, econômica e política.

Em suma, a magnitude do risco impõe ação regulatória imediata. Tanto as mudanças climáticas quanto a perda de biodiversidade já se encontram em estágios críticos, talvez sem tempo hábil para correções eficazes. A inteligência artificial insere-se nesse mesmo horizonte de riscos existenciais: é, ao mesmo tempo, uma ameaça autônoma e um fator de aceleração das demais crises.

O caráter transversal da IA amplia sua gravidade. Ao atravessar todas as esferas da vida social, econômica e política, a tecnologia pode intensificar a crise ambiental pelo seu elevado gasto energético, comprometer a segurança internacional ao ser instrumentalizada em contextos militares, e ampliar a instabilidade social e institucional diante da possibilidade de erros sistêmicos em setores críticos como saúde, finanças e infraestrutura.

Diferentemente de outros riscos que se manifestam em ciclos mais longos, a IA carrega a particularidade de combinar velocidade de desenvolvimento, opacidade técnica e ausência de governança global. Nesse sentido, a humanidade enfrenta desafio duplo: regular a IA como risco em si e, simultaneamente, impedir que seu uso indiscriminado agrave problemas estruturais de sobrevivência presentes no plano ambiental e social.

Apesar da gravidade da ameaça existencial da IA, reconhecida por figuras proeminentes como Stephen Hawking e por diversos líderes globais, a ação regulatória internacional para mitigar seu potencial catastrófico tem sido lenta.

O problema central da governança reside em um dilema estrutural de ação coletiva. O desenvolvimento da IA promete benefícios imensos e vantagens competitivas insuperáveis. As partes interessadas — sejam corporações ou Estados — estão presas em uma verdadeira corrida armamentista de IA, em que os incentivos individuais impedem a aplicação unilateral de freios ao avanço da tecnologia. A IA constitui, assim, uma transformação irreversível, comparável em impacto histórico à introdução da máquina de escrever ou do computador pessoal.

Como observa Hirsh (2023), apesar dos riscos existenciais associados ao desenvolvimento militar da IA, os benefícios estratégicos e a pressão competitiva entre Estados tornam quase impossível interromper essa nova corrida armamentista de IA.

Druzin, Boute e Ramsden afirmam:

> Como os *stakeholders* não podem se autorregular de forma efetiva, o ônus de mitigar os riscos da tecnologia de IA recai necessariamente sobre nossas instituições jurídicas. O Direito é nossa última linha de defesa. Um conjunto compartilhado de regras e salvaguardas regulatórias universalmente acordadas pode difundir esse problema de ação coletiva, criando um framework que facilita a coordenação entre nações, empresas, pesquisadores e outros stakeholders (nossa tradução) (2025)[2].

[2] Texto original em inglês: *Because stakeholders cannot effectively regulate themselves, the burden of mitigating the risks of AI technology necessarily falls on our legal institutions. Law is our last line of defense. A shared set of rules and universally agreed-upon regulatory guardrails can diffuse this collective action problem by creating a framework that facilitates coordination among nations, companies, researchers, and other stakeholders.*

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

Entretanto, as iniciativas internacionais existentes permanecem fragmentadas e concentram-se em aspectos técnicos e sociais não existenciais, sem enfrentar diretamente os riscos estruturais de longo prazo.

As Recomendações da OCDE sobre Inteligência Artificial (OECD *AI* Principles), adotadas em 2019 e atualizadas em 2024, constituem o primeiro padrão intergovernamental sobre IA. Elas promovem o desenvolvimento de sistemas inovadores e confiáveis, em consonância com os direitos humanos e valores democráticos. Embora abordem riscos como discriminação, segurança, privacidade e responsabilidade, fazem-no dentro de um quadro de governança orientado a maximizar benefícios e mitigar riscos genéricos, e não ante o risco existencial (OECD, 2024).

A Lei da Inteligência Artificial da União Europeia (*EU AI Act*), em vigor desde agosto de 2024, estabelece regime regulatório baseado em níveis de risco. O diploma proíbe sistemas de risco inaceitável e determina a inclusão de mecanismos de segurança, como a possibilidade de desativação em caso de falha; todavia, não confronta diretamente o risco fundamental de perda de controle humano ou mesmo de extinção, e se limita a uma abordagem técnico-setorial (EUROPEAN UNION, 2024).

De modo semelhante, a Convenção-Quadro do Conselho da Europa sobre Inteligência Artificial, Direitos Humanos, Democracia e Estado de Direito, adotada em maio de 2024, representa o primeiro tratado internacional vinculativo juridicamente sobre o tema. Seu objetivo é assegurar que o ciclo de vida dos sistemas de IA seja compatível com os direitos fundamentais, mas, novamente, a ênfase recai sobre princípios gerais e medidas graduadas, sem previsão de mecanismos específicos contra riscos de ordem existencial (COUNCIL OF EUROPE, 2024). A falta de consenso e a corrida competitiva significam que, no plano internacional, a exigência de agir com cautela pode ser interpretada como obrigação dos Estados de cooperar no enfrentamento de ameaças potencialmente irreversíveis.

Esse entendimento se vincula ao princípio da precaução, consolidado no Direito Internacional Ambiental, segundo o qual a incerteza científica não justifica a ausência de medidas protetivas. Aplicado ao campo da IA, tal princípio implica a necessidade de cooperação internacional para antecipar, prevenir ou minimizar riscos e mitigar seus efeitos adversos.

A dificuldade em regulamentar a IA está ligada de modo intrínseco à sua natureza: a IA é, conforme declaração do Vice-Ministro Wu Zhaohui na *AI Safety Summit*, “incerta, inexplicável” (UNITED KINGDOM 2023, p. 1/2). A complexidade e a opacidade inerentes aos sistemas algorítmicos avançados dificultam a previsibilidade e a compreensão de seus resultados. Nesse sentido, muitos especialistas alertam que os rápidos avanços tecnológicos apresentam riscos significativos para a humanidade que não podem ser plenamente compreendidos pela ciência atual (UNITED KINGDOM, 2023, p. 195/196).

O problema conecta-se ao princípio da precaução no Direito Internacional, que fora formulado diante da necessidade de antecipar riscos em situações de incerteza científica, especialmente em matéria ambiental (UNITED NATIONS, 1992).

O problema da opacidade da IA (a *caixa preta*) é agravado pela tendência de os sistemas complexos de computação tornarem-se cada vez mais invisíveis e centrais à infraestrutura social contemporânea. O mundo tornou-se um código/espaço, em que o ambiente e a experiência dependem

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

criticamente do aplicativo, e revelam-se apenas em falhas. Essa complexidade e ilegibilidade aumentam a dificuldade em responsabilizar e criticar os sistemas (BRIDLE, 2019).

Os sistemas de IA, baseados em aprendizado de máquina e *deep learning*, aprendem a partir de dados históricos, o que gera o risco de reproduzir preconceitos e injustiças do passado. O historiador Walter Benjamin notou que "Não há documento da civilização que não seja documento da barbárie" (2006); ao treinar IAs com dados históricos, essa barbárie é codificada no futuro.

O Direito Internacional dos Direitos Humanos (DIDH) constitui o arcabouço normativo central para enfrentar os desafios trazidos pela IA, sobretudo diante da ameaça existencial que pode representar ao direito à vida. A universalidade das normas de direitos humanos confere-lhes legitimidade e impõe obrigações diretas aos Estados; o direito à vida é o núcleo a partir do qual todos os outros se fundamentam.

Apesar dessa centralidade, o avanço exponencial dos algoritmos confronta o DIDH com um dilema estrutural: sua incapacidade de oferecer respostas unívocas e eficazes diante da opacidade tecnológica e da velocidade do dano algorítmico.

Como prática argumentativa, o DIDH buscou historicamente a despolitização das relações internacionais; contudo, como observa Koskenniemi, isso o torna vulnerável à crítica de funcionar ora como uma utopia moralista desvinculada da realidade, ora como apologia dos interesses estatais (KOSKENNIEMI, 2005, p. 5). A tensão entre ideal normativo e concretude social explica sua indeterminação.

A crise estrutural do DIDH se acentua no contexto da IA, pois o Direito Internacional tradicionalmente repousa sobre premissas liberais e uma relação vertical entre Estado e indivíduo. Nesse cenário, a dignidade humana emerge como o tronco interpretativo a partir do qual direitos como privacidade e não discriminação devem ser compreendidos, e não apenas como derivações secundárias.

Luciano Floridi propõe compreender a dignidade em uma perspectiva *antropo-excêntrica*, que reconhece a incompletude humana como processo em constante *vir-a-ser* (*work-in-progress*) (2016, p. 308). Para o autor, a essência humana é informacional, e nossa dignidade consiste em preservar a abertura das escolhas e identidades (FLORIDI, 2016, p. 310/311). Nesse sentido, a IA, ao fixar e classificar indivíduos em perfis (*mounting board of a profile*), ameaça a própria ontologia humana, gerando impactos desumanizadores (FLORIDI, 2016, p. 311/312).

A aplicação de direitos fundamentais a sistemas algorítmicos complexos evidencia o viés estrutural da argumentação jurídica internacional. A indeterminação do DIDH não decorre apenas da ambivalência semântica, mas da colisão de princípios em torno dos fundamentos das normas. Como lembra Alexy, o Direito articula-se dualmente entre regras — que exprimem um dever definitivo — e princípios — que traduzem um dever *prima facie* (2010, p. 167, 174). Nos conflitos, o sistema exige ponderação e revela escolhas inevitavelmente políticas.

No caso do dano algorítmico, a tensão opõe justiça (conteúdo material) e segurança jurídica (dimensão formal), bem como exige decisões em contextos de incerteza científica (ALEXY, 2010, p. 180).

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

Dois problemas se destacam, o viés estrutural e a falha na neutralidade: a crença na neutralidade da IA é ilusória, pois os sistemas de *machine learning* reproduzem preconceitos históricos. Como alerta Cathy O'Neil, os modelos de *Big Data* "reforçam a discriminação, sustentam os afortunados e punem os oprimidos, minando a democracia" (2016).

O segundo problema é a fragilidade da abordagem vertical: o DIDH, centrado na relação Estado–indivíduo, mostra-se insuficiente ante o poder das grandes corporações tecnológicas. Embora exista a noção de responsabilidade de respeitar os direitos humanos por parte das empresas, não há obrigação coercitiva direta, e o ônus recai sobre instituições jurídicas.

O Estado, sob o DIDH, vincula-se a obrigações positivas que exigem ação proativa para proteger o direito à vida, inclusive diante de ameaças incertas ou não humanas, como desastres naturais. Essa lógica aplica-se também aos riscos da IA.

O princípio da precaução, consolidado no Direito Internacional, determina que a ausência de certeza científica não deve justificar a inação quando houver possibilidade de dano grave ou irreversível. Sua relevância para a IA reside no caráter existencial dos riscos, ainda que de baixa probabilidade, que impõem a necessidade de regulação antecipada (ALEXY, 2010, p.194)

Assim, a crise estrutural do DIDH ante a IA exige a incorporação da precaução como critério de legitimação para a criação de marcos jurídicos básicos e mecanismos de responsabilização, inclusive para atores não estatais (ALEXY, 2010, p. 179). A ausência desse aparato compromete a eficácia do sistema em traduzir obrigações de proteção em garantias judiciais concretas.

A crise estrutural do Direito Internacional dos Direitos Humanos (DIDH), diante da Inteligência Artificial (IA), manifesta-se pela exposição do viés estrutural da argumentação jurídica internacional. A indeterminação intrínseca do DIDH, resultante da colisão de princípios e da incerteza científica, obriga a jurisprudência a realizar escolhas políticas em contextos de alta complexidade. Embora o discurso jurídico apresente fluidez argumentativa, na prática tende a reproduzir o status quo, favorecendo interesses estabelecidos e incorporando preferências profundamente enraizadas (KOSKENNIEMI, 2005, p. 600, 607).

O sistema jurídico, à luz do Tridimensionalismo de Miguel Reale, caracteriza-se pela interação dinâmica entre fato, valor e norma, numa dialética de complementaridade (REALE, 2002, p. 558). A norma é sempre uma resposta ao fato social, que concretiza valores. A aplicação do Direito, entretanto, não se reduz a uma concatenação lógica de proposições, mas constitui um processo de integração em que a decisão judicial incorpora o significado dos valores subjacentes aos fatos (REALE, 2002, p. 549).

Essa tensão é visível na articulação entre regras (que exprimem deveres definitivos) e princípios (que traduzem deveres *prima facie*) (REALE, 2002, p. 555). Como observa Alexy, em situações de colisão, a decisão exige ponderação e revela a volubilidade da jurisprudência (2010, p. 174). Assim, a interpretação não é fixa: varia conforme a leitura dos fatos e das circunstâncias, ainda que o texto normativo permaneça idêntico.

Nesse quadro, a indeterminação do DIDH evidencia-se na necessidade de escolhas políticas pelos órgãos jurisdicionais, diante da pressão entre coerência normativa e materialidade do conflito (ALEXY, 2010, p. 167, 174). O Direito, longe de ser um sistema puramente lógico, revela-se permeado por decisões contextuais e contingentes. Esse viés estrutural é amplificado pela IA. Sistemas de

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

machine learning e *deep learning* aprendem a partir de dados históricos para codificar no futuro preconceitos e injustiças do passado.

Cathy O'Neil demonstra que modelos algorítmicos, frequentemente desregulamentados e incontestáveis, reforçam desigualdades. Ao estruturar decisões em larga escala, "sustentam os afortunados, punem os oprimidos e minam a democracia" (2016, p.288) A lógica da eficiência algorítmica, aparentemente neutra, mascara um processo de reificação de crenças e vieses sociais.

Diante do risco existencial da IA, o princípio da precaução surge como imperativo ético e legal. Seu núcleo reside na máxima de que a ausência de certeza científica não pode justificar a inação diante de ameaças potencialmente catastróficas. O princípio da precaução exige a adoção de salvaguardas antecipatórias diante de riscos sistêmicos e incertos.

3 O PARADOXO DA PROTEÇÃO INTERNACIONAL: UTOPIA, APOLOGIA E O RISCO ALGORÍTMICO

O desenvolvimento exponencial da IA inaugura um novo paradoxo para o DIDH. Ao mesmo tempo que amplia as possibilidades de concretização de direitos pela automação de políticas públicas, pela ampliação de acesso a bens e serviços e pela democratização de informações; revela-se também como ameaça existencial à humanidade e converte-se em instrumento de discriminação, controle e exclusão digital.

O dilema se acentua porque o DIDH, historicamente fundado em uma matriz liberal e antropocêntrica, encontra dificuldade em responder às transformações trazidas pela IA. Conforme adverte Martti Koskenniemi, o Direito Internacional oscila entre dois polos inconciliáveis: a apologia — quando se reduz a justificar os interesses estatais — e a utopia — quando pretende encarnar um ideal moral universal (2005).

Na era algorítmica, essa tensão é reatualizada, pois a normatividade universalista colide com a concretude tecnológica dos sistemas de decisão automatizada. A IA, na qualidade de tecnologia autônoma, escapa da lógica estatal tradicional e desloca o centro da responsabilidade para atores privados, cuja atuação global desafia a estrutura normativa vigente (CRESTANE; LEAL, 2024, p. 26-27).

Essa ambiguidade revela a natureza politicamente contingente do DIDH. Robert Alexy ensina que o sistema jurídico é composto por regras — que impõem deveres definitivos — e princípios — que expressam deveres *prima facie*, dependentes de ponderação (2008, p. 88-89).

Diante dos riscos algorítmicos, a ponderação entre liberdade de inovação e proteção da dignidade humana deixa de ser um exercício teórico e passa a constituir uma escolha política essencial. Como adverte Eduardo Cambi e Maria Eduarda Amaral, os algoritmos judiciais, ao classificarem e hierarquizarem padrões de decisão, tendem a reproduzir preconceitos e a comprometer a imparcialidade jurisdicional (2023, p. 203–206).

A crise epistêmica da IA decorre da opacidade e da incompreensibilidade técnica que tornam impossível verificar integralmente as decisões automatizadas. Cathy O'Neil define esses modelos como

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

"armas de destruição matemática", pois se baseiam em dados enviesados e são projetados para escalabilidade e eficiência, e não para justiça (2016, p. 12-13).

Tal constatação implica deslocamento ontológico: o dano algorítmico não resulta apenas de falhas técnicas, mas de estruturas sociais codificadas digitalmente. A IA, portanto, atua como catalisadora da desigualdade estrutural, transforma injustiças históricas em decisões matemáticas aparentemente neutras (FRAZÃO, 2021, p. 45).

Nesse contexto, a dignidade humana assume posição ontológica e normativa central. Luciano Floridi propõe uma concepção antropo-excêntrica de dignidade, segundo a qual o ser humano é um "processo informacional em aberto" (2013, p. 7–8), cuja integridade reside na preservação da liberdade de escolha e de identidade (FLORIDI, 2013, p. 7–8).

Assim, a classificação automatizada e a vigilância algorítmica representam riscos diretos à ontologia humana, pois reduzem o sujeito a um perfil preditivo. Dérique Crestane e Mônia Leal apontam que tal redução traduz a reprodução da discriminação estrutural no espaço digital, que implica a erosão da alteridade e da pluralidade que sustentam o constitucionalismo de direitos (2024, p. 25-27).

A insuficiência do modelo vertical do DIDH — centrado na relação Estado versus indivíduo — torna-se patente diante da atuação das grandes corporações de tecnologia. Essas empresas, responsáveis por infraestruturas digitais que atravessam fronteiras, escapam à jurisdição tradicional e operam sob lógicas próprias de *accountability*.

Segundo Jarbas Cugula, Sandro GODOY e Gabriel ALMEIDA, a concentração do poder informacional nas mãos de poucos agentes privados exige novo pacto social digital, fundado na transparência e na função social da tecnologia (2023, p. 15–16). Sem essa redefinição, a proteção dos direitos humanos corre o risco de converter-se em mero discurso moral, incapaz de enfrentar a materialidade das práticas discriminatórias.

Dessa forma, o paradoxo da proteção internacional dos direitos humanos ante a IA revela a necessidade de superação do binarismo entre utopia e apologia. A utopia neste contexto simboliza a crença em um Direito universal descolado da técnica; a apologia, sua rendição à racionalidade instrumental.

Entre ambas, impõe-se a construção da ética jurídica de precaução e responsabilidade, apta a reconhecer a incerteza científica e a antecipar danos potenciais. Como lembra Cambi e Amaral, o princípio da precaução é expressão de dever de diligência global que vincula Estados e empresas na prevenção de riscos existenciais (2023, p. 215–216).

O risco algorítmico, portanto, não é apenas técnico, mas civilizatório. Ele questiona a capacidade do Direito Internacional de permanecer fiel à sua vocação humanista diante da automação da decisão e da erosão da agência humana.

A utopia de um Direito universal precisa ser reconstruída a partir da realidade concreta das tecnologias emergentes; a apologia da neutralidade técnica precisa ser substituída pela afirmação da responsabilidade coletiva. O futuro da proteção internacional dos direitos humanos dependerá de sua capacidade de traduzir valores morais em arquiteturas regulatórias concretas e eficazes.

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

4 O PRINCÍPIO DA PRECAUÇÃO COMO OBRIGAÇÃO POSITIVA INEVITÁVEL

O avanço da IA inaugura um regime de incerteza radical que tensiona as estruturas clássicas do Direito Internacional dos Direitos Humanos. Não se trata apenas de inovação tecnológica, mas de transformação ontológica da agência humana, dos processos decisórios e da gramática da vida social.

A tecnologia deixa de ser mero instrumento e passa a operar como infraestrutura normativa e epistemológica — capaz de influenciar escolhas, modelar comportamentos e deslocar a autonomia humana para sistemas autônomos e opacos.

Diante desse cenário, o princípio da precaução emerge como obrigação positiva inevitável, constitui fundamento jurídico para atuação preventiva do Estado em contextos de risco tecnológico profundo.

A tradição jurídica internacional reconhece que, em situações de incerteza científica e possibilidade de dano grave, a ausência de ação não configura neutralidade, mas violação estrutural do dever de proteção (SARLET; FENSTERSEIFER, 2021, p. 182-183).

No campo dos direitos humanos, essa exigência se intensifica. A Corte Interamericana de Direitos Humanos consolida o entendimento de que os Estados devem adotar medidas legislativas, administrativas, técnicas e judiciais para garantir a efetividade dos direitos fundamentais em cenários de risco social e tecnológico[3]. A omissão estatal se traduz em responsabilidade internacional quando a falta de regulação expõe indivíduos e coletividades a danos previsíveis e que não foram evitados de maneira eficaz. A IA materializa essa ameaça: opera com velocidade, escala e opacidade inéditas; reproduz assimetrias históricas; produz danos difusos e irreversíveis; e escapa aos mecanismos tradicionais de controle democrático (O'NEIL, p. 27-34).

Nesse contexto, o princípio da precaução adquire *status* de princípio geral do Direito Internacional, e projeta-se sobre a governança tecnológica como critério hermenêutico e político para proteção do humano.

Essa compreensão já se reflete em instrumentos normativos recentes, como a Convenção-Quadro do Conselho da Europa sobre Inteligência Artificial, de 2024, que afirma expressamente a necessidade de antecipação e mitigação de riscos graves mesmo na ausência de certeza empírica (COUNCIL OF EUROPE , 2024).

Em linha semelhante, decisões do Supremo Tribunal Federal (STF) brasileiro vinculam inovação tecnológica à observância do dever constitucional de proteção da vida e da dignidade humana (BRASIL, 2020).

Diante disso, o Direito Internacional não pode mais operar sob paradigma exclusivamente reativo. A urgência e a irreversibilidade dos danos tecnológicos exigem hermenêutica do risco, fundada na precaução como critério orientador para políticas públicas, *design* algorítmico, mecanismos de supervisão e responsabilidade estatal e corporativa.

[3] CORTE INTERAMERICANA DE DIREITOS HUMANOS. *Caso Velásquez Rodríguez vs. Honduras*. Sentença de 29 jul. 1988.

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

A IA, portanto, constitui parâmetro crítico para avaliação da governança tecnológica: falhar na precaução não será falhar em interpretar o Direito — será falhar em proteger a própria condição humana.

5 DISCRIMINAÇÃO E VIÉS ESTRUTURAL: A JURISPRUDÊNCIA DIANTE DA OPACIDADE ALGORÍTIMICA

O avanço acelerado da IA impõe à ciência jurídica um desafio de natureza epistemológica e axiológica, especialmente no tocante à preservação dos direitos fundamentais diante da crescente opacidade dos sistemas algorítmicos. A aplicação da IA Generativa no campo jurídico, embora represente um instrumento de elevada eficiência na automação de tarefas, na sistematização de informações e na análise de grandes volumes de dados, introduz nova categoria de fenômeno juridicamente relevante — um fato jurídico algorítmico. Tal fato decorre da atuação de sistemas que produzem efeitos no mundo jurídico sem a intermediação direta da vontade humana, desloca o centro da imputação de sentido normativo e exige, portanto, redefinição das fronteiras entre ação humana, decisão técnica e responsabilidade jurídica.

Sem a devida contenção epistemológica — isto é, sem a imposição de critérios racionais, verificáveis e transparentes que assegurem a legitimidade do conhecimento produzido pela IA — corre-se o risco de comprometer os fundamentos da justiça e da igualdade, pilares da racionalidade jurídica moderna. A denominada opacidade algorítmica (*black box*) agrava essa vulnerabilidade, ao dificultar a compreensão dos processos de inferência e decisão empregados pelos modelos de IA, o que pode, em última instância, reproduzir e amplificar os vieses estruturais existentes na sociedade. Assim, a crise de confiança na jurisprudência contemporânea manifesta-se não apenas na instabilidade interpretativa, mas também na incapacidade de auditar criticamente as decisões mediadas por sistemas de IA.

A epistemologia simplista que permeia parte da cultura tecnológica do Vale do Silício apoia-se na crença de que a acumulação massiva de dados — denominada hiperinclusão informacional — seria suficiente para compreender e solucionar problemas sociais complexos. Essa postura, frequentemente associada ao chamado solucionismo tecnológico (MOROZOV, 2013), reduz a complexidade institucional e histórica de questões como pobreza, exclusão e discriminação racial a meras variáveis estatísticas manipuláveis por meio de algoritmos e aplicativos. Ao proceder dessa forma, tal racionalidade ignora as dimensões estruturais e normativas do fenômeno social, e substitui a mediação política e jurídica pela lógica da eficiência técnica.

Consequentemente, a política é reconfigurada como um espetáculo individualista e favorável ao consumidor, em que os conflitos coletivos são reinterpretados como falhas de comportamento ou de escolha individual, passíveis de correção por sensores, métricas e sistemas automatizados de decisão. A redução do espaço público a uma arena de consumo de soluções digitais não apenas despolitiza o debate sobre justiça social, como também desloca o *locus* da responsabilidade: o cidadão deixa de ser sujeito de direitos e passa a ser tratado como objeto de monitoramento e ajuste algorítmico.

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

O denominado novo consenso algorítmico não é neutro. Os sistemas de *deep learning*, por serem treinados a partir de dados históricos que refletem desequilíbrios de poder presentes e passados, têm o potencial de incorporar, ocultar e amplificar vieses, remodelar preconceitos culturais, sociais e étnicos em forma de verdades empíricas aparentemente objetivas. Tal processo converte discriminações estruturais em resultados tecnicamente validados, o que reforça desigualdades e legitima práticas discriminatórias sob a aparência de neutralidade estatística.

A política orientada por sistemas de IA tende a gerir efeitos, e não causas, opera com explicações monocausais e reduz a complexidade do real a correlações probabilísticas. A racionalidade contrasta com a natureza dialética da democracia, cuja essência está na deliberação plural e na busca de causas múltiplas que permitam a reconciliação entre conflitos sociais e os ideais de justiça.

Diante desse cenário, a jurisprudência internacional deve reafirmar sua função integradora entre fato, valor e norma, inspirar-se no método tridimensional proposto por Reale (1968), reinterpretado neste contexto em chave universal como fundamento hermenêutico aplicável à proteção internacional dos direitos humanos.

O valor axiológico central da dignidade da pessoa humana orienta a construção normativa dos sistemas de proteção dos direitos humanos, assegura a igualdade material e formal e proíbe qualquer forma de discriminação atentatória à liberdade, à integridade e à autonomia pessoal.

Instrumentos fundamentais, como a Declaração Universal dos Direitos Humanos (1948), o Pacto Internacional sobre Direitos Civis e Políticos (1966) e a Convenção Europeia de Direitos Humanos (1950), consagram o dever positivo dos Estados de prevenir e combater discriminações diretas e indiretas. Sob a ótica contemporânea dos direitos humanos e da interpretação evolutiva desses tratados, esse dever estende-se às práticas discriminatórias mediadas por tecnologias digitais e sistemas de IA. O compromisso é reforçado por instrumentos normativos como o Regulamento Geral de Proteção de Dados da União Europeia (GDPR, 2016) — que concretiza o direito fundamental previsto no artigo 8º. da Carta dos Direitos Fundamentais da União Europeia[4] — e pela Convenção 108+ do Conselho da Europa (1981/2018)[5], que reconhece a proteção de dados pessoais como elemento intrínseco à dignidade da pessoa humana e essencial à tutela das liberdades fundamentais na era digital.

[4] Carta dos Direitos Fundamentais da União Europeia. Artigo 8º — *Proteção de dados de caráter pessoal:* 1. Todas as pessoas têm direito à proteção dos dados de caráter pessoal que lhes digam respeito. 2. Esses dados devem ser tratados de forma leal, para fins determinados e com o consentimento da pessoa em causa ou com outro fundamento legítimo previsto por lei. 3. O cumprimento destas regras será fiscalizado por uma autoridade independente. *(Jornal Oficial da União Europeia, C 326/391, de 26.10.2012).*

[5] Conselho da Europa. **Convenção para a Proteção das Pessoas relativamente ao Tratamento Automatizado de Dados de Carácter Pessoal (Convenção 108+)**, de 1981, modernizada em 2018. Artigo 1º — *Objetivo e âmbito: "*O objetivo da presente Convenção é assegurar, em território de cada Parte, a proteção de todas as pessoas singulares, qualquer que seja a sua nacionalidade ou residência, no que respeita ao tratamento automatizado de dados de caráter pessoal, em consonância com os direitos humanos e liberdades fundamentais, e, em particular, com o direito à vida privada." *(Conselho da Europa, Estrasburgo, 2018).*

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

6 CONSIDERAÇÕES FINAIS

O percurso investigativo deste estudo demonstrou que o avanço da IA, ao mesmo tempo que inaugura novas possibilidades de emancipação humana, expõe de forma aguda os limites estruturais do Direito Internacional dos Direitos Humanos.

O paradoxo fundamental identificado — entre a promessa de progresso técnico e o risco de degradação civilizatória — evidencia que o sistema jurídico internacional se encontra diante de encruzilhada histórica: ou adapta sua normatividade às transformações algorítmicas e informacionais do século XXI, ou permanecerá prisioneiro de racionalidade analógica incapaz de responder à velocidade, escala e opacidade das novas tecnologias.

A IA reconfigura o campo ontológico da ação humana. Ao deslocar o *locus* da decisão para sistemas autônomos e estatísticos, ela dissolve a fronteira entre sujeito e instrumento, e produz um novo tipo de responsabilidade difusa e sistêmica.

Essa mutação epistemológica desafia o modelo liberal do Direito Internacional dos Direitos Humanos, centrado na relação vertical Estado–indivíduo, e exige a construção de hermenêutica transnacional da responsabilidade.

O problema não reside apenas na violação pontual de direitos, mas na emergência de estruturas automatizadas que naturalizam discriminações e reduzem a pessoa a um perfil preditivo.

Nesse sentido, a dignidade humana — núcleo axiológico do Direito Internacional — deve ser reinterpretada como valor informacional e processual, cuja proteção implica garantir a abertura ontológica do humano diante de tecnologias que tendem a fixar, classificar e determinar identidades.

O princípio da precaução revelou-se, nesse contexto, não como simples categoria ambiental, mas como imperativo jurídico-ético universal. Aplicado à IA, ele traduz o dever de agir diante da incerteza, ao impor aos Estados a obrigação positiva de prevenir, mitigar e fiscalizar riscos existenciais, mesmo na ausência de certeza científica.

A precaução não é um obstáculo ao desenvolvimento tecnológico, mas um instrumento de racionalidade coletiva — uma ética da contenção que protege o humano frente à aceleração cega da inovação.

Assim como o princípio da dignidade fundamentou o constitucionalismo dos direitos no pós-guerra, a precaução deve sustentar o constitucionalismo algorítmico do século XXI: um novo pacto normativo em que o risco não seja privatizado, e a responsabilidade, distribuída de forma justa entre Estados, corporações e sociedade civil.

A investigação também revelou que o dilema da utopia e da apologia, identificado por Koskenniemi, permanece atual na era digital. O discurso universalista do Direito Internacional dos Direitos Humanos corre o risco de se tornar utópico quando ignora as materialidades técnicas que condicionam a vida contemporânea; e torna-se apologético quando se submete à lógica econômica e geopolítica das grandes corporações tecnológicas.

Entre esses extremos, impõe-se a terceira via: o realismo normativo da precaução, que reconhece a inevitabilidade do risco, mas insiste na responsabilidade compartilhada de mitigá-lo.

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

Os sistemas de IA operam como novas instituições globais de poder, dotadas de capacidade decisória, preditiva e disciplinar sem precedentes. Diante disso, a omissão regulatória não é mera falha administrativa: é violação do dever de proteção.

O Direito Internacional dos Direitos Humanos, como ordenamento de garantias universais, deve expandir-se para além de sua estrutura vertical, reconhecer o papel central dos atores privados na produção de danos e na reprodução de desigualdades.

A responsabilidade corporativa internacional, ainda incipiente, precisa evoluir para modelos vinculantes de prestação de contas, com mecanismos de supervisão pública e acesso à justiça transnacional.

A jurisprudência internacional tem papel essencial nesse processo. A hermenêutica do risco exige que tribunais e órgãos de direitos humanos adotem interpretações evolutivas dos tratados, ao aplicar o princípio da precaução de modo transversal às novas formas de ameaça.

Como reconheceu a Corte Interamericana de Direitos Humanos, a proteção efetiva do direito à vida e à dignidade impõe deveres de diligência reforçada em face de riscos emergentes — sejam ambientais, tecnológicos ou híbridos. Essa orientação deve servir como parâmetro para o reconhecimento da IA como objeto de obrigações internacionais de cuidado.

Ao longo desta pesquisa, constatou-se que a discriminação algorítmica constitui o ponto mais visível da crise de legitimidade do Direito ante a automação. A crença na neutralidade técnica mascara a reprodução digital de hierarquias raciais, de gênero e socioeconômicas, e converte a desigualdade em estatística e o preconceito em cálculo.

A efetividade do Direito Internacional dos Direitos Humanos dependerá, portanto, de sua capacidade de reconectar a técnica à ética, restituir à justiça sua função crítica e humanizadora.

A reconstrução do Direito Internacional dos Direitos Humanos diante da IA exige três movimentos complementares: (i) redefinição ontológica da dignidade humana em termos informacionais; (ii) institucionalização global do princípio da precaução como fundamento de uma governança tecnológica justa; e (iii) consolidação de uma responsabilidade compartilhada entre Estados, empresas e sociedade para o controle dos riscos sistêmicos.

A IA é o espelho da incompletude moral e institucional. O risco não provém primariamente da autonomia da máquina, mas da omissão humana em governá-la adequadamente.

O futuro dos direitos humanos dependerá da coragem jurídica de antecipar o dano antes que ele se torne irreversível — de transformar a precaução em justiça e a dignidade em prática concreta. A proteção do humano, ante a opacidade da máquina, será o verdadeiro teste da maturidade do Direito Internacional no século XXI.

REFERÊNCIAS

ALEXY, Robert. The Dual Nature of Law. **Ratio Juris**, v. 23, n. 2, p. 167-182, jun. 2010.

AMODEI, Dario *et al*. **Concrete Problems in AI Safety**. arXiv:1606.06565, 21 jun. 2016. Disponível em: https://arxiv.org/abs/1606.06565 . Acesso em: 17 out. 2025.

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

BENJAMIN, Walter. Theses on the Philosophy of History. In: BENJAMIN, Walter. **Selected Writings**, v. 4, p. 1.938-1.940. Cambridge, MA: Harvard University Press, 2006.

BOSTROM, Nick. **Superintelligence**: Paths, Dangers, Strategies. Oxford: Oxford University Press, 2016.

BRIDLE, James. **A nova idade das trevas**: A tecnologia e o fim do futuro. Tradução de Érico Assis. São Paulo: Todavia, 2019.

CHRISTIAN, Brian. **The Alignment Problem**: How Can Machines Learn Human Values? New York: W.W. Norton & Company, 2021.

COUNCIL OF EUROPE. Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. Adotada em 17 maio 2024; aberta para assinatura em 11 set. 2024. **Council of Europe Treaty Series**, n. 225. Disponível em: https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence Acesso em: 17 out. 2025.

DRUZIN, Bryan; BOUTE, Anatole; RAMSDEN, Michael. Confronting catastrophic risk: the international obligation to regulate artificial intelligence. **Michigan Journal of International Law**, Ann Arbor, v. 46, n. 2, p. 173-217, 2025. Disponível em: https://repository.law.umich.edu/mjil/vol46/iss2/2 Acesso em: 17 out. 2025.

EDGECLIFFE-JOHNSON, Andrew. AI poses 'bracing test' to multilateral system, says UK deputy prime minister. **Financial Times**, Londres, 24 set. 2023. Disponível em: https://www.ft.com/content/9d98da0a-14e2-4bbb-a076-91ef131fe2b2 . Acesso em: 17 out. 2025.

EUROPEAN UNION. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence. **Official Journal of the European Union**, L 1689, 12 jul. 2024. Disponível em: https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence. Acesso em: 17 out. 2025.

FLORIDI, Luciano. On Human Dignity as a Foundation for the Right to Privacy. **Philosophy & Technology**, v. 29, p. 307-312, dez. 2016.

GABRIEL, Lason**. Artificial Intelligence, Values and Alignment**. arXiv:2001.09768, 13 jan. 2020. Disponível em: https://arxiv.org/abs/2001.09768 . Acesso em: 17 out. 2025.

HENDRYCKS, Dan *et al*. **An Overview of Catastrophic AI Risks**. arXiv:2306.12001, 21 jun. 2023. Disponível em: https://arxiv.org/abs/2306.12001 . Acesso em: 17 out. 2025.

HIRSH, Michael. How AI will revolutionize warfare. **Foreign Policy**, [S.l.], 11 abr. 2023. Disponível em: https://foreignpolicy.com/2023/04/11/ai-arms-race-artificial-intelligence-chatgpt-military-technology/ Acesso em: 17 out. 2025.

INTERNATIONAL DIALOGUE ON AI SAFETY (IDAIS). **IDAIS Beijing Dialogue Report**. Beijing: IDAIS, 10-11 mar. 2024. Disponível em: https://idais.ai/dialogue/idais-beijing . Acesso em: 17 out. 2025.

RVING, Geoffrey; CHRISTIANO, Paul; AMODEI, Dario. **AI Safety via Debate**. arXiv:1805.00899, 2 maio 2018. Disponível em: https://arxiv.org/abs/1805.00899 . Acesso em: 17 out. 2025.

Submetido em: 10/10/2025
Aprovado em: 24/11/2025
Avaliação: Double Blind Reviewe
ISSN: **2316-2880**

JI, Jiaming *et al*. **AI Alignment**: A Comprehensive Survey. arXiv:2310.19852, 1 maio 2023. Disponível em: https://arxiv.org/abs/2310.19852 . Acesso em: 17 out. 2025.

KISSINGER, Henry A.; SCHMIDT, Eric; MUNDIE, Craig. **Genesis**: Artificial Intelligence, Hope, and the Human Spirit. New York: Little, Brown and Company, 2024.

KOSKENNIEMI, Martti. **From Apology to Utopia**: The Structure of International Legal Argument. Reissue with a new Epilogue. Cambridge: Cambridge University Press, 2005.

OECD. **Recommendation of the Council on Artificial Intelligence**. OECD/LEGAL/0449, adopted on 22 May 2019, updated 2024. Disponível em: https://oecd.ai/en/ai-principles . Acesso em: 17 out. 2025.

O'NEIL, Cathy. **Weapons of Math Destruction**: How Big Data Increases Inequality and Threatens Democracy. New York: Crown, 2016.

REALE, Miguel. **Filosofia do Direito**. São Paulo: Saraiva, 2002.

REALE, Miguel. **O direito como experiência**: introdução à epistemologia jurídica. São Paulo: Saraiva, 1992.

RUSSELL, Stuart. **Human Compatible: Artificial Intelligence and the Problem of Control**. London: Penguin Press, 2019.

UNITED KINGDOM. **The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023**. GOV.UK, 1 nov. 2023. Disponível em: https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 . Acesso em: 17 out. 2025.

UNITED NATIONS. **Rio Declaration on Environment and Development**. Doc. A/CONF.151/26/Rev.1 (Vol. I), annex I, 12 ago. 1992. Disponível em: https://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_CONF.151_26_Vol.I_Declaration.pdf . Acesso em: 17 out. 2025.

VAN VEEN, Christiaan; CATH, Corinne. **Artificial Intelligence**: What's Human Rights Got To Do With It? Data & Society Points, 14 maio 2018. Disponível em: https://datasociety.net/points/artificial-intelligence-whats-human-rights-got-to-do-with-it/ . Acesso em: 17 out. 2025.

WHITING, Kate. From Sam Altman to António Guterres: Here's what 10 leaders said about AI at Davos 2024. **World Economic Forum**, 23 Jan. 2024. Disponível em: https://www.weforum.org/agenda/2024/01/what-leaders-said-about-ai-at-davos-2024 . Acesso em: 17 out. 2025.

ZAKRZEWSKI, Cat. The Davos elite embraced AI in 2023. Now they fear it. **The Washington Post**, Washington, 18 Jan. 2024. Disponível em: https://www.washingtonpost.com/technology/2024/01/18/davos-ai-world-economic-forum . Acesso em: 17 out. 2025.