# Inteligência artificial e direitos humanos: diligência, governança e obrigação do Estado

## *Inteligencia artificial y derechos humanos: diligencia, gobernanza y obligación del Estado*

## *Artificial intelligence and human rights: due diligence, governance, and State obligations*

***Fabiane Gaspar e Armando De Negri Filho***

**Resumo:** O relatório *Human Rights and New and Emerging Digital Technologies: Human Rights Due Diligence for AI and State Obligations*, publicado pelo Alto Comissariado das Nações Unidas para os Direitos Humanos (OHCHR), analisa o papel da devida diligência em direitos humanos como mecanismo central para orientar a atuação dos Estados diante dos riscos associados à inteligência artificial (IA). O documento sustenta que as obrigações estatais de respeitar, proteger e garantir os direitos humanos exigem avaliações contínuas dos impactos da IA ao longo de todo o seu ciclo de vida, incluindo desenvolvimento, aquisição, implementação e monitoramento. Além disso, enfatiza a importância da prevenção, transparência, prestação de contas, supervisão independente, responsabilização corporativa e participação social ao lidar com riscos associados à IA. O relatório propõe um modelo regulatório com base em riscos e fundamentado no direito internacional dos direitos humanos, com ênfase na identificação precoce de danos sistêmicos e discriminatórios. Dessa maneira, a publicação constitui uma referência importante a construção de políticas públicas compatíveis com os direitos fundamentais em sociedades cada vez mais digitais.

**Palavras-chave**: Inteligência artificial; direitos humanos; devida diligência; governança digital.

***Resumen:*** *El informe "Human Rights and New and Emerging Digital Technologies: Human Rights Due Diligence for AI and State Obligations" (Derechos Humanos e Tecnologías Nuevas e Emergentes: Devida Diligencia para IA e Obligaciones Estatales, en libre traducción), publicado por la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos, examina la diligencia debida en materia de derechos humanos como un instrumento esencial para la gobernanza de la inteligencia artificial. El documento sostiene que las obligaciones estatales de respetar, proteger y garantizar los derechos humanos requieren evaluaciones permanentes de los impactos generados por los sistemas de IA durante todo su ciclo de vida, incluidos el desarrollo, la adquisición, la implementación y el seguimiento. Asimismo, destaca la importancia de la prevención, la transparencia, la rendición de cuentas, la supervisión independiente, responsabilidad corporativa y la participación pública en hacer frente a los riesgos relacionados con la IA. El informe propone un modelo regulatorio basado en riesgos y fundamentado en el derecho internacional de los derechos humanos, con énfasis en la identificación temprana de daños sistémicos y discriminatorios. De esta manera, constituye una referencia importante para el desarrollo de políticas públicas y marcos normativos orientados a una innovación tecnológica compatible con la dignidad humana.*

***Palabras clave****: Inteligencia artificial; derechos humanos; debida diligencia; gobernanza digital.*

***Abstract:*** *The report "Human Rights and New and Emerging Digital Technologies: Human Rights Due Diligence for AI and State Obligations", issued by the Office of the United Nations High Commissioner for Human Rights, examines human rights due diligence as a core mechanism for governing artificial intelligence. The report argues that States' obligations to respect, protect and fulfil human rights require continuous impact assessments throughout the entire AI lifecycle, including development, procurement, deployment and monitoring. It highlights the importance of prevention, transparency, accountability, independent oversight, and public participation in addressing AI-related risks.*

*The report proposes a risk-based regulatory model grounded in international human rights law, with an emphasis on the early identification of systemic and discriminatory harms. As such, it serves as an important reference for the development of regulatory frameworks aimed at technological innovation that is compatible with human dignity.*

***Keywords:*** *Artificial Intelligence; Human Rights; Due Diligence; Digital Governance.*

## Introdução

A expansão acelerada da inteligência artificial (IA)[1] tem provocado profundas transformações econômicas, políticas e sociais, reconfigurando a forma como Estados, empresas e cidadãos interagem no ambiente digital. O crescente emprego de sistemas automatizados em áreas como segurança pública, administração da justiça, controle migratório, educação, saúde e políticas sociais ampliou as possibilidades de eficiência institucional, mas também revelou riscos significativos para a proteção dos direitos humanos. Nesse cenário, a formulação de mecanismos jurídicos torna-se relevante na tentativa em garantir que a inovação tecnológica seja compatível com os princípios da dignidade humana, da igualdade e da justiça social.

É nesse contexto que se insere o relatório *Human Rights and New and Emerging Digital Technologies: Human Rights Due Diligence for AI and State Obligations* (Direitos Humanos e Novas as Tecnologias Digitais Emergentes: Devida Diligência em Direitos Humanos para IA e Obrigações dos Estados, em tradução livre) (A/HRC/82/33), elaborado pelo Alto Comissariado das Nações Unidas para os Direitos Humanos (OHCHR) e apresentado durante a 62ª sessão do Conselho de Direitos Humanos da ONU.

O documento representa um marco importante na construção de parâmetros internacionais voltados à governança da inteligência artificial, ao propor que a devida diligência em direitos humanos seja incluída, no cumprimento das obrigações estatais, como um instrumento indispensável para respeitar, proteger e garantir direitos fundamentais.

A argumentação desenvolvida pelo relatório parte da premissa de que os impactos da IA não são exclusivamente individuais, mas coletivos, cumulativos e sistêmicos.

[1] Não existe uma definição única e simples de inteligência artificial. A Organização para a Cooperação e o Desenvolvimento Econômico (OCDE) define IA como sistemas baseados em máquinas que, com objetivos explícitos ou implícitos, inferem, a partir dos dados de entrada que recebem, como "gerar resultados tais como previsões, conteúdo, recomendações ou decisões que possam influenciar ambientes físicos ou virtuais". Disponível em: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 Acesso em: 03 ago. 2026.

Consequentemente, mecanismos tradicionais de supervisão e responsabilização mostram-se insuficientes e ineficazes diante de tecnologias capazes de produzir efeitos amplificados em larga escala. A publicação propõe, portanto, uma abordagem preventiva e contínua, baseada na antecipação de riscos e na avaliação permanente dos possíveis impactos dessas tecnologias sobre diferentes grupos sociais, especialmente aqueles historicamente vulnerabilizados.

Nesse sentido, o relatório não se limita a reiterar princípios gerais de direitos humanos, mas busca operacionalizá-los em um contexto marcado pela complexidade tecnológica, pela opacidade algorítmica[2] e pela atuação transnacional das empresas que compõem o ecossistema da inteligência artificial. O resultado é uma contribuição relevante para o debate contemporâneo sobre regulação digital, servindo como referência para formuladores de políticas públicas, pesquisadores, órgãos reguladores e instituições internacionais.

**Relatório "Direitos Humanos e Novas as Tecnologias Digitais Emergentes: Devida Diligência em Direitos Humanos para IA e Obrigações dos Estados"**

**Inteligência artificial e devida diligência em direitos humanos – panorama geral**

O relatório parte do reconhecimento de que a inteligência artificial se tornou uma das tecnologias mais influentes do século XXI, produzindo impactos significativos sobre direitos civis, políticos, econômicos, sociais e culturais. Nesse contexto, a devida diligência em direitos humanos (*human rights due diligence*) emerge como um dos principais instrumentos normativos para orientar a atuação de Estados e empresas diante dos riscos associados à crescente digitalização da vida social.

Segundo o Alto Comissariado das Nações Unidas para os Direitos Humanos (OHCHR):

> *... a devida diligência consiste em um processo contínuo voltado à identificação, prevenção, mitigação e prestação de contas sobre impactos adversos relacionados aos direitos humanos decorrentes do desenvolvimento e da utilização de tecnologias digitais, especialmente sistemas de inteligência artificial.*

O documento destaca que a centralidade da devida diligência no debate sobre governança digital resulta de um processo gradual de consolidação dentro do sistema internacional de direitos humanos. Relatórios anteriores do OHCHR já haviam enfatizado a necessidade de avaliações permanentes dos impactos das tecnologias digitais sobre a privacidade, a liberdade de expressão, a igualdade e outros direitos fundamentais.

A inovação do relatório de 2026 reside em aprofundar a relação entre a devida diligência e as obrigações jurídicas dos Estados, estabelecendo que a proteção dos direitos humanos em

[2] A opacidade algorítmica (ou do algoritmo) refere-se ao fenômeno em que o funcionamento interno de um sistema de IA é opaco (obscuro, não transparente) para seus usuários ou reguladores, uma questão comum em modelos de aprendizado profundo (*deep learning models*). De acordo com o Artigo 13 da Lei de IA da União Europeia e os Artigos 13 a 15 do Regulamento Geral de Proteção de Dados da União Europeia, os sistemas de IA de alto risco devem ser transparentes e interpretáveis. Disponível em https://winners.com.tw/en/glossary/algorithmic-opacity Acesso em: 03 ago. 2026.

contextos de inteligência artificial não pode depender exclusivamente da boa vontade das empresas ou de mecanismos voluntários de autorregulação.

Nesse sentido, o relatório argumenta que a devida diligência deve ser compreendida como um mecanismo capaz de integrar a proteção dos direitos humanos em todas as etapas do ciclo de vida da inteligência artificial. Isso inclui a coleta de dados, o treinamento dos modelos, o desenvolvimento de algoritmos, os testes de segurança, a implementação prática dos sistemas, sua utilização por entidades públicas ou privadas e o monitoramento permanente de seus efeitos concretos.

A análise parte da premissa de que os riscos associados à IA não são eventos isolados ou ocasionais, mas fenômenos dinâmicos que podem surgir ou se agravar ao longo do tempo, exigindo respostas igualmente contínuas e adaptáveis.

Outro aspecto importante abordado pelo relatório refere-se à crescente incorporação da devida diligência em instrumentos internacionais recentes. O documento menciona o Pacto Global Digital das Nações Unidas, recomendações da UNESCO sobre ética da inteligência artificial e diversas resoluções da Assembleia Geral e do Conselho de Direitos Humanos que já reconhecem a necessidade de mecanismos preventivos de gestão de riscos. Essa convergência normativa evidencia a formação gradual de um consenso internacional segundo o qual a inovação tecnológica deve estar subordinada aos princípios da dignidade humana, da não discriminação e da responsabilidade pública (*public accountability*).

O relatório também enfatiza que a devida diligência não deve ser vista apenas como uma ferramenta de prevenção de danos, mas como um instrumento de governança democrática. Isso ocorre porque os processos de avaliação de impacto exigem transparência, participação social e diálogo com grupos potencialmente afetados pelas decisões algorítmicas. Dessa forma, a devida diligência contribui para ampliar a legitimidade das políticas tecnológicas, promovendo formas mais inclusivas e participativas de regulação da inteligência artificial.

Em síntese, o panorama apresentado pelo OHCHR demonstra que a devida diligência em direitos humanos deve ser considerada como um modelo central para a governança global da inteligência artificial. Mais do que um procedimento administrativo, trata-se de uma abordagem normativa que visa assegurar que os avanços tecnológicos sejam compatíveis com os compromissos internacionais assumidos pelos Estados em matéria de direitos humanos.

**Sistemas de inteligência artificial e desafios específicos aos direitos humanos**

Uma das contribuições centrais do relatório consiste em demonstrar que a inteligência artificial apresenta riscos qualitativamente distintos daqueles observados em tecnologias digitais anteriores. Segundo o OHCHR, os impactos da IA não se restringem a violações individuais facilmente identificáveis, mas frequentemente assumem caráter coletivo, sistêmico, cumulativo e transnacional. Isso ocorre porque os sistemas inteligentes operam a partir da coleta massiva de dados, do processamento automatizado de informações e da capacidade de realizar inferências, previsões, recomendações e decisões que influenciam diretamente ambientes físicos e virtuais.

Uma característica fundamental destacada pelo relatório é a natureza intensiva em dados dos sistemas de IA. A coleta, o armazenamento e o processamento de grandes volumes de informações pessoais ampliam significativamente os riscos relacionados à privacidade e à proteção de dados. Além disso, a concentração dessas informações em plataformas digitais,

combinada ao uso de mecanismos automatizados de vigilância e monitoramento, cria condições para formas de controle social sem precedentes. Os sistemas de IA frequentemente dependem de correlações estatísticas e padrões comportamentais que permitem classificar indivíduos e grupos, produzindo consequências concretas no acesso a direitos, oportunidades e serviços.

O documento enfatiza que a dependência de dados históricos também favorece a reprodução de desigualdades estruturais. Como os algoritmos aprendem a partir de informações produzidas em contextos sociais já marcados por discriminações raciais, étnicas, econômicas e de gênero, existe o risco de que esses sistemas internalizem e ampliem tais distorções.

O problema torna-se ainda mais preocupante porque a aparente objetividade matemática das decisões algorítmicas pode mascarar preconceitos presentes tanto nas bases de dados quanto nos critérios de desenvolvimento dos modelos computacionais. Como consequência, grupos historicamente vulnerabilizados, incluindo mulheres, minorias raciais e étnicas, pessoas com deficiência, migrantes e populações socialmente marginalizadas, podem sofrer impactos desproporcionais em áreas como recrutamento profissional, concessão de crédito, acesso a serviços públicos e processos decisórios estatais.

Outra preocupação central do relatório refere-se à capacidade dos sistemas de IA de operar em larga escala. Diferentemente de decisões humanas isoladas, decisões automatizadas podem ser replicadas simultaneamente para milhões de pessoas. Assim, pequenos erros estatísticos, falhas de programação ou vieses aparentemente insignificantes podem produzir consequências amplificadas e disseminadas em toda a sociedade. Por essa razão, o OHCHR argumenta que a análise de riscos não deve limitar-se à avaliação de prejuízos individuais, mas considerar também os efeitos coletivos, sistêmicos e cumulativos que podem emergir da utilização continuada dessas tecnologias.

O relatório destaca ainda o impacto da inteligência artificial sobre o ambiente informacional contemporâneo. Algoritmos utilizados por plataformas digitais influenciam diretamente quais conteúdos são exibidos, recomendados ou ocultados aos usuários, moldando fluxos de informação e processos de formação de opinião. Essa capacidade de mediação afeta direitos relacionados à liberdade de expressão, ao acesso à informação e à participação democrática. Em contextos políticos, mecanismos automatizados de recomendação podem facilitar a segmentação excessiva de audiências, a manipulação informacional e a disseminação direcionada de conteúdos enganosos. Esses riscos são agravados pela opacidade algorítmica, uma vez que os critérios de funcionamento dos sistemas frequentemente permanecem inacessíveis ao escrutínio público e à fiscalização regulatória.

Especial atenção é dedicada às tecnologias de vigilância habilitadas por inteligência artificial, especialmente aquelas baseadas em reconhecimento facial, identificação biométrica, rastreamento comportamental e monitoramento automatizado. O relatório observa que essas ferramentas não afetam apenas o direito à privacidade, mas também produzem impactos indiretos sobre diversas outras liberdades fundamentais.

A percepção de estar permanentemente monitorado pode desencorajar a participação em manifestações públicas, atividades políticas, movimentos sociais e iniciativas de defesa de direitos, gerando efeitos inibidores sobre a liberdade de expressão, a liberdade de associação e a participação democrática. Dessa forma, os impactos da vigilância algorítmica ultrapassam a esfera individual e assumem dimensão coletiva.

No campo das relações de trabalho, o documento aponta que a crescente adoção de sistemas automatizados está transformando profundamente a organização laboral. Ferramentas de monitoramento algorítmico são utilizadas para avaliar desempenho, controlar produtividade e coordenar fluxos de trabalho. Embora possam gerar ganhos de eficiência, essas tecnologias também têm potencial para ampliar a vigilância sobre trabalhadores, reduzir sua autonomia e afetar o direito a condições dignas e justas de trabalho.

O relatório também chama atenção para a utilização crescente de sistemas de IA em setores estratégicos da atuação estatal, como administração pública, justiça, segurança pública, gestão migratória, assistência social e defesa militar. Nesses contextos, erros algorítmicos podem consolidar preconceitos, reproduzir falhas institucionais e provocar restrições indevidas de direitos em larga escala. Como os sistemas automatizados tendem a aplicar os mesmos critérios decisórios a grandes conjuntos populacionais, uma única falha de concepção ou implementação pode afetar milhares de pessoas simultaneamente. Além disso, decisões algorítmicas em diferentes setores podem reforçar mutuamente padrões de exclusão e vulnerabilidade, criando ciclos acumulativos de desvantagens sociais.

Diante desse cenário, o relatório conclui que os desafios impostos pela inteligência artificial exigem uma abordagem regulatória específica, baseada nos direitos humanos e orientada pela devida diligência. Diferentemente de tecnologias convencionais, os sistemas de IA apresentam características estruturais que dificultam a aplicação dos mecanismos tradicionais de prevenção, supervisão e responsabilização. Entre esses fatores destacam-se a opacidade algorítmica, a escalabilidade dos efeitos produzidos pelos sistemas automatizados, a intensa dependência de dados, a crescente autonomia dos modelos de IA e a complexidade das cadeias globais de desenvolvimento, fornecimento e utilização dessas tecnologias.

A **opacidade algorítmica** constitui um dos principais desafios apontados pelo relatório. Muitos sistemas contemporâneos de inteligência artificial, especialmente aqueles desenvolvidos com técnicas de aprendizado de máquina e aprendizado profundo (*machine learning* e *deep learning*), funcionam de forma que nem mesmo seus desenvolvedores conseguem explicar plenamente como determinadas conclusões ou recomendações foram produzidas. Essa característica compromete princípios fundamentais do Estado Democrático de Direito, como transparência, previsibilidade, motivação das decisões e possibilidade de controle social.

Quando uma pessoa é afetada por uma decisão automatizada relacionada à concessão de benefícios sociais, à seleção para uma vaga de emprego, à obtenção de crédito ou até mesmo à avaliação de risco criminal, a dificuldade de compreender os critérios utilizados pelo algoritmo cria obstáculos para o exercício do contraditório e da ampla defesa. O relatório argumenta que essa falta de explicabilidade torna especialmente difícil identificar discriminações, corrigir erros e atribuir responsabilidades quando ocorrem violações de direitos humanos.

Outro aspecto crítico é a **escalabilidade dos efeitos da inteligência artificial**. Enquanto decisões humanas costumam produzir impactos limitados a casos específicos, sistemas algorítmicos podem replicar instantaneamente suas decisões para milhões de pessoas. Um erro em um modelo estatístico, uma falha de programação ou um viés presente no treinamento do algoritmo pode ser reproduzido em larga escala, afetando simultaneamente grupos inteiros da população. Nesse sentido, o relatório destaca que a IA possui a capacidade de transformar pequenos problemas técnicos em grandes problemas sociais.

Um critério discriminatório incorporado ao algoritmo de seleção de candidatos, por exemplo, pode excluir sistematicamente milhares de pessoas de oportunidades de emprego; da mesma forma, um sistema automatizado utilizado em programas governamentais pode negar benefícios sociais a grandes contingentes populacionais antes mesmo que a falha seja identificada. Essa capacidade de amplificação dos efeitos faz com que os riscos da IA sejam potencialmente mais graves e difíceis de corrigir do que aqueles decorrentes de decisões humanas isoladas.

A **intensa dependência de dados** representa outro desafio estrutural destacado pelo OHCHR. Os sistemas de inteligência artificial dependem de enormes quantidades de informações para treinamento, validação e funcionamento. Entretanto, os dados utilizados não são neutros. Eles refletem contextos sociais, econômicos e políticos específicos, frequentemente marcados por desigualdades históricas e discriminações estruturais.

Quando esses dados são utilizados para treinar algoritmos, preconceitos e assimetrias existentes podem ser incorporados ao sistema e reproduzidos de maneira automatizada. Além disso, a coleta massiva de dados necessária para alimentar sistemas de IA levanta preocupações relacionadas ao direito à privacidade, à proteção de dados pessoais e à autodeterminação informativa. O relatório ressalta que muitos dos impactos negativos da IA decorrem não apenas dos algoritmos em si, mas também da qualidade, da representatividade e da governança dos dados utilizados em seu desenvolvimento.

O documento chama atenção ainda para a crescente utilização de **sistemas autônomos e de IA agêntica**, nos quais múltiplos agentes artificiais atuam de forma relativamente independente e interagem entre si para atingir determinados objetivos. Nessas situações, torna-se cada vez mais difícil identificar a origem de uma decisão ou atribuir responsabilidade por determinado resultado. A atuação simultânea de vários sistemas autônomos pode gerar efeitos inesperados e comportamentos emergentes que não foram explicitamente programados pelos desenvolvedores. Esse fenômeno desafia mecanismos tradicionais de responsabilização jurídica, que normalmente pressupõem a existência de um agente claramente identificável responsável pela conduta analisada.

Além disso, o relatório destaca a complexidade das **cadeias globais de valor da inteligência artificial**. Um único sistema pode ser resultado da interação entre fabricantes de hardware, empresas de computação em nuvem, desenvolvedores de modelos fundacionais, fornecedores de bases de dados, trabalhadores responsáveis pela rotulação de informações, integradores de sistemas e usuários finais localizados em países distintos. Da mesma forma, os dados podem ser coletados em uma região do mundo, armazenados em outra, processados em servidores situados em diferentes jurisdições e utilizados globalmente. Essa fragmentação dificulta a identificação de responsabilidades e favorece o surgimento de lacunas regulatórias, especialmente quando diferentes países adotam níveis distintos de proteção aos direitos humanos e de fiscalização tecnológica.

Segundo o relatório, a combinação desses fatores produz um cenário em que os riscos associados à inteligência artificial tendem a ser não apenas amplificados, mas também mais difíceis de detectar, compreender e corrigir. Por essa razão, o OHCHR sustenta que os mecanismos tradicionais de supervisão estatal são insuficientes para responder adequadamente aos desafios contemporâneos.

Em seu lugar, defende a implementação de processos contínuos e adaptativos de devida diligência em direitos humanos, capazes de acompanhar todas as etapas do ciclo de vida da tecnologia, desde a concepção e o treinamento dos modelos até sua implementação, monitoramento e eventual descontinuação. Nessa perspectiva, quanto maior for a complexidade, a opacidade e o potencial de impacto de um sistema de inteligência artificial, maior deverá ser o nível de cautela, transparência, supervisão e responsabilidade exigido dos Estados e das empresas envolvidas em seu desenvolvimento e utilização.

**Como a devida diligência em direitos humanos pode enfrentar os desafios específicos da IA**

**Prevenção, identificação de riscos, avaliação e previsibilidade ao longo do ciclo de vida da IA**

O relatório estabelece que a prevenção constitui o elemento central da devida diligência em direitos humanos. Em vez de reagir aos danos apenas após sua ocorrência, os Estados devem identificar e avaliar riscos potenciais antes do desenvolvimento, da aquisição ou da implementação de sistemas de inteligência artificial.

Essa abordagem exige a realização de **avaliações de impacto em direitos humanos** (*Human Rights Impact Assessments*) em todas as fases do ciclo de vida da IA. Tais avaliações devem examinar possíveis interferências em direitos como igualdade, privacidade, liberdade de expressão, devido processo legal, acesso a serviços públicos e integridade física.

O documento enfatiza que essas avaliações precisam basear-se nas melhores evidências científicas e técnicas disponíveis. Não basta uma análise jurídica formal; torna-se necessário envolver especialistas em ciência de dados, tecnologia, ética, direitos humanos, sociologia e áreas correlatas para compreender adequadamente os potenciais impactos dos sistemas.

O relatório também incorpora o **princípio da precaução**, tradicionalmente aplicado ao direito ambiental. Quando os riscos são elevados e não podem ser adequadamente avaliados ou mitigados, a ausência de certeza científica não pode justificar a omissão estatal. Em determinadas circunstâncias, especialmente em áreas como segurança pública, gestão migratória, proteção social ou aplicações militares, a devida diligência pode exigir a suspensão ou mesmo a proibição do uso de determinadas tecnologias até que garantias suficientes sejam estabelecidas.

Outro ponto importante refere-se à necessidade de atenção especial às populações vulneráveis. O relatório reconhece que mulheres, minorias raciais e étnicas, pessoas com deficiência, migrantes e grupos historicamente discriminados podem sofrer impactos desproporcionais decorrentes da utilização de sistemas automatizados. Por isso, a avaliação de riscos deve incorporar uma perspectiva de igualdade substancial e de não discriminação.

Além disso, o documento enfatiza que a diligência preventiva não deixa de existir em emergências. Mesmo em contextos de crises humanitárias, desastres naturais ou operações de segurança, a utilização de IA deve ser acompanhada por monitoramento contínuo e mecanismos capazes de corrigir rapidamente falhas identificadas após a implementação.

**Regulação das empresas e a combinação inteligente (*smart mix)* de medidas**

O relatório dedica especial atenção à responsabilidade dos Estados na regulação das empresas que participam do ecossistema da inteligência artificial. Considerando que grande parte das tecnologias atualmente utilizadas por governos é desenvolvida por agentes privados,

o cumprimento das obrigações estatais de proteção dos direitos humanos depende diretamente da capacidade regulatória do poder público.

Nesse sentido, o OHCHR adota a abordagem proposta pelos Princípios Orientadores das Nações Unidas sobre Empresas e Direitos Humanos (U*N Guiding Principles on Business and Human Rights – UNGPs*), defendendo a implementação de uma **combinação inteligente** (***smart mix)*** **de medidas regulatórias**. Essa estratégia consiste na combinação de mecanismos voluntários e obrigatórios destinados a promover comportamentos empresariais compatíveis com os direitos humanos.

As iniciativas voluntárias incluem códigos de conduta, diretrizes éticas, certificações e programas de autorregulação. Embora importantes, o relatório afirma que elas não são suficientes para enfrentar riscos estruturais de grande magnitude.

Por essa razão, os Estados devem complementar essas iniciativas com medidas juridicamente vinculantes, como:

- legislações obrigatórias de devida diligência em direitos humanos;
- exigência de auditorias independentes;
- avaliações obrigatórias de impacto;
- fiscalização por autoridades especializadas;
- sanções administrativas e judiciais;
- mecanismos de monitoramento contínuo.

Em aplicações consideradas de alto risco, o documento recomenda que autoridades reguladoras possuam poderes efetivos de investigação e intervenção, incluindo a possibilidade de restringir ou suspender sistemas que apresentem riscos significativos aos direitos fundamentais.

**Responsabilização (*Accountability*) e acesso a recursos efetivos**

A responsabilização constitui um dos pilares centrais da estrutura de devida diligência proposta pelo relatório. O documento observa que muitos danos provocados pela inteligência artificial permanecem invisíveis ou difíceis de identificar em razão da complexidade técnica e da opacidade dos algoritmos.

Para enfrentar esse problema, o relatório recomenda que os Estados implementem mecanismos robustos de transparência e supervisão. Entre as medidas sugeridas destacam-se a criação de registros públicos de sistemas de IA utilizados pelo setor público, a divulgação de avaliações de impacto em direitos humanos e a realização periódica de auditorias independentes.

O documento ressalta que indivíduos afetados por decisões automatizadas devem ter acesso a mecanismos efetivos de contestação e revisão. Isso inclui o direito de obter explicações compreensíveis sobre decisões algorítmicas (direito à explicabilidade), solicitar revisão por seres humanos e recorrer a instâncias administrativas ou judiciais.

O relatório também demonstra preocupação com o uso excessivo de alegações de segredo comercial para impedir a fiscalização dos sistemas. Embora reconheça a legitimidade da proteção à propriedade intelectual, afirma que tais mecanismos não podem impedir a investigação de violações de direitos humanos nem restringir o acesso à justiça.

Outro aspecto relevante refere-se aos desafios transnacionais da responsabilização. Como muitos sistemas operam simultaneamente em diferentes países, o relatório sugere a construção de mecanismos de cooperação internacional que permitam o compartilhamento de informações e a responsabilização de atores envolvidos em danos que ultrapassem fronteiras nacionais.

**Participação e cooperação internacional**

O relatório conclui a seção destacando a participação social e a cooperação internacional como componentes indispensáveis da devida diligência em direitos humanos para inteligência artificial.

Segundo o OHCHR, avaliações de risco somente podem ser consideradas adequadas quando incluem as perspectivas das pessoas potencialmente afetadas. A participação de usuários, comunidades vulneráveis, organizações da sociedade civil, instituições acadêmicas e especialistas técnicos permite identificar impactos que poderiam passar despercebidos em avaliações exclusivamente governamentais ou corporativas.

O documento distingue duas dimensões complementares da participação. A primeira refere-se ao engajamento durante os processos de avaliação de impactos, permitindo que diferentes atores contribuam para a identificação de riscos concretos. A segunda relaciona-se à participação na elaboração das políticas públicas e marcos regulatórios que irão orientar a governança da inteligência artificial.

Além da participação interna, o relatório enfatiza que a natureza global da IA torna indispensável a cooperação internacional. Nenhum Estado é capaz de regular sozinho sistemas tecnológicos que operam por meio de infraestruturas globais e cadeias produtivas distribuídas em várias jurisdições.

Por essa razão, o documento defende o fortalecimento de iniciativas voltadas para:

- harmonização de normas regulatórias;
- compartilhamento internacional de informações sobre riscos e incidentes;
- criação de padrões técnicos comuns;
- cooperação entre autoridades reguladoras;
- proteção transnacional dos direitos humanos na era digital.

Para o OHCHR, a participação multissetorial e a cooperação internacional não constituem apenas boas práticas de governança, mas requisitos fundamentais para que os Estados possam cumprir efetivamente suas obrigações de respeitar, proteger e realizar os direitos humanos em um contexto de crescente expansão da inteligência artificial.

**Conclusões e recomendações do relatório**

Reprodução das recomendações com assistência do tradutor *DeepL.com* e revisão pelos autores:

> *Reconhecendo que o cumprimento das obrigações dos Estados de proteger, respeitar e garantir os direitos humanos no contexto da inteligência artificial e dos riscos a ela associados exige uma ação estatal cuidadosa, proativa e diligente, o ACNUDH recomenda que os Estados:*

*(a) Realizem, de forma sistemática e contínua, a devida diligência em matéria de direitos humanos, incluindo avaliações de impacto sobre os direitos humanos que também levem em conta danos sistêmicos e cumulativos, ao longo de todo o ciclo de vida dos sistemas de IA que projetam, desenvolvem, implantam, vendem, adquirem ou operam, e que o façam por meio de processos consultivos inclusivos envolvendo a sociedade civil, empresas de tecnologia e outras partes interessadas relevantes;*

*(b) Assegurem que seus processos de devida diligência e auditorias sejam capazes de identificar e abordar danos sistêmicos, baseados em grupos, coletivos e cumulativos decorrentes da IA, levando em consideração, particularmente, os grupos historicamente sujeitos à discriminação;*

*(c) Nos casos em que os riscos não possam ser efetivamente mitigados em aplicações de IA de alto risco e sensíveis aos direitos, incluindo aplicação da lei, gestão de fronteiras, migração, administração de programas de assistência social e aplicações militares da IA, os Estados devem, em conformidade com o princípio da precaução, abster-se de implantar ou proibir totalmente usos específicos da IA;*

*(d) Preencher as lacunas de governança nas compras públicas e nas parcerias entre o Estado e o setor privado; garantir que as compras públicas e as parcerias público-privadas envolvendo sistemas de IA estejam plenamente sujeitas à devida diligência em matéria de direitos humanos, o que inclui avaliações independentes de impacto sobre os direitos humanos e transparência quanto à capacidade do sistema monitoramento contínuo dos impactos e salvaguardas contratuais que garantam a prestação de contas, a supervisão e o acesso a vias de recurso;*

*(e) Garantir a transparência e assegurar o acesso à informação ao longo de todo o ciclo de vida da IA, particularmente no que diz respeito à IA em funções públicas, e incorporar a participação significativa no desenvolvimento e na adoção de leis e políticas que regem a IA;*

*(f) Dotar os órgãos de supervisão do mandato e dos recursos necessários para monitorar e investigar violações dos direitos humanos relacionadas à IA, garantindo o direito de contestar e obter revisão humana de decisões assistidas por IA, salvaguardando os direitos processuais à explicação e removendo barreiras legais, tais como a dependência excessiva de alegações de segredo comercial que impedem o acesso à justiça;*

*(g) Investir e garantir as capacidades institucionais e técnicas necessárias para realizar a devida diligência em matéria de direitos humanos para sistemas de IA, inclusive assegurando que as autoridades competentes tenham a capacidade de (i) realizar avaliações independentes e multidisciplinares de riscos e impactos; e (ii) envolver sistematicamente as partes interessadas afetadas para identificar riscos e impactos no mundo real;*

*(h) Fortalecer a cooperação internacional e o desenvolvimento de padrões comuns para lidar com riscos transfronteiriços aos direitos humanos; prevenir lacunas jurisdicionais e regulatórias que prejudiquem a proteção dos direitos humanos; e apoiar processos inclusivos envolvendo múltiplas partes interessadas na governança da IA;*

*(i) Proteger os indivíduos contra violações dos direitos humanos decorrentes de atividades corporativas envolvendo IA, por meio da aplicação de uma combinação inteligente de medidas regulatórias e incentivos políticos, em conformidade com os Princípios Orientadores das Nações Unidas (UNGPs) e o documento A/HRC/59/32, recomendação 55 (a); estabelecer estruturas que exijam a devida diligência em matéria de direitos humanos por parte das empresas e, em contextos que apresentem riscos elevados, conferir às autoridades de supervisão poderes de investigação e fiscalização, incluindo a auditoria independente de sistemas de IA de alto risco;*

> *(j) Colaborar com o serviço de assessoria do ACNUDH sobre direitos humanos no espaço digital para obter apoio na elaboração de leis e políticas relacionadas à devida diligência em matéria de IA.*

**Considerações finais**

O relatório do Alto Comissariado das Nações Unidas para os Direitos Humanos representa uma importante formulação internacional sobre a relação entre inteligência artificial e obrigações estatais em matéria de direitos humanos. Sua principal contribuição reside na afirmação de que a devida diligência em direitos humanos não deve ser tratada como uma prática facultativa ou meramente administrativa, mas como uma exigência decorrente das obrigações jurídicas dos Estados de respeitar, proteger e garantir os direitos fundamentais em ambientes cada vez mais mediados por tecnologias digitais.

O documento apresenta uma mudança de ponto de vista ao deslocar o debate da simples regulação tecnológica para uma perspectiva centrada em direitos humanos. Essa abordagem permite incorporar preocupações frequentemente negligenciadas nos debates sobre inovação, como discriminação algorítmica, vigilância em massa, exclusão digital, impactos laborais da automação e restrições ao acesso à justiça.

Entretanto, a implementação das recomendações propostas pelo Alto Comissário enfrenta desafios significativos. Em primeiro lugar, muitos governos ainda carecem de capacidade técnica e institucional para realizar avaliações independentes de sistemas complexos de IA. A ausência de especialistas, infraestrutura de auditoria e órgãos reguladores adequadamente estruturados pode comprometer a efetividade das medidas sugeridas.

Outra dificuldade a ser considerada decorre da própria natureza complexa e global da indústria de inteligência artificial. Sistemas frequentemente são desenvolvidos em uma jurisdição, treinados em outra e utilizados em diversos países simultaneamente. Essa fragmentação dificulta a definição de responsabilidades e cria lacunas regulatórias que podem ser exploradas por empresas transnacionais.

A cooperação internacional proposta pelo relatório surge, portanto, como requisito indispensável, embora sua concretização dependa de elevado grau de coordenação política entre Estados com diferentes níveis de desenvolvimento tecnológico e interesses econômicos.

Há ainda obstáculos relacionados à transparência. Grande parte das ferramentas de IA utilizadas atualmente encontra-se protegida por segredos comerciais e direitos de propriedade intelectual. A exigência de explicabilidade e auditoria independente pode enfrentar resistência do setor privado, sobretudo em mercados altamente competitivos.

Conciliar inovação, proteção de investimentos e direitos fundamentais é incompatível com o atual modelo capitalista e regulatório. Por outro lado, algumas condições favorecem a implementação das recomendações do relatório, pois há um movimento de reconhecimento internacional dos riscos da IA, aliado à consolidação de marcos regulatórios como o AI Act europeu e às iniciativas normativas da UNESCO e das Nações Unidas, criando um repertório e um ambiente cada vez mais propício para a adoção de políticas de governança responsável.

Além disso, o avanço das metodologias de avaliação de impacto algorítmico e das auditorias de sistemas automatizados oferece instrumentos concretos para operacionalizar a devida diligência defendida pelo OHCHR.

Em síntese, o relatório demonstra que a governança da inteligência artificial não pode ser reduzida a uma questão técnica. Trata-se de um desafio eminentemente jurídico, político e ético, que exige a construção de instituições capazes de compatibilizar inovação e proteção dos direitos humanos.

Um mérito do relatório a ser reconhecido é o de fornecer uma estrutura consistente normativa e epistêmica para orientar esse processo, oferecendo aos Estados um roteiro para que a inteligência artificial seja utilizada como instrumento de promoção do bem-estar coletivo, e não como um vetor de aprofundamento das desigualdades e vulnerabilidades existentes.

**Referências**


ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). Alto Comissariado das Nações Unidas para os Direitos Humanos (OHCHR). Human rights and new and emerging digital technologies: human rights due diligence for AI and State obligations: report of the United Nations High Commissioner for Human Rights. Geneva: United Nations Human Rights Council, 2026. Documento A/HRC/62/33. Disponível em: <https://www.ohchr.org/en/documents/thematic-reports/ahrc6233-human-rights-due-diligence-ai-and-state-obligations-under>. Acesso em: 30 jul. 2026.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). Guiding Principles on Business and Human Rights: implementing the United Nations “Protect, Respect and Remedy” Framework. New York; Geneva: United Nations, 2011. Disponível em: <https://www.ohchr.org>. Acesso em: 30 jul. 2026.

UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION (UNESCO). Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO, 2021. Disponível em: <https://unesdoc.unesco.org/>. Acesso em: 30 jul. 2026.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT (OECD). Recommendation of the Council on Artificial Intelligence. Paris: OECD, 2019. Atualizada em 2024. Disponível em: <https://legalinstruments.oecd.org>. Acesso em: 30 jul. 2026.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS (ONU). Assembleia Geral. Global Digital Compact. New York: United Nations, 2025. Disponível em: <https://www.un.org>. Acesso em: 30 jul. 2026.

CORTE INTERAMERICANA DE DIREITOS HUMANOS. Emergência climática e direitos humanos: parecer consultivo OC-32/25. San José: Corte Interamericana de Direitos Humanos, 2025. Disponível em: <https://www.corteidh.or.cr>. Acesso em: 30 jul. 2026.

COMISSÃO EUROPEIA. Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, Brussels, 2024. Disponível em: <https://eur-lex.europa.eu>. Acesso em: 30 jul. 2026.

UNIÃO EUROPEIA. Directive (EU) 2024/1760 on Corporate Sustainability Due Diligence. Official Journal of the European Union, Brussels, 2024. Disponível em: <https://eur-lex.europa.eu>. Acesso em: 30 jul. 2026.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO. Recommendation of the Council on Artificial Intelligence. OECD/LEGAL/0463. Disponível em: < https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> Acesso em: 03 ago. 2026.