REVISTA DIREITO PÚBLICO

INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS: UMA AVALIAÇÃO ÉTICA NOS NEGÓCIOS[1]

ARTIFICIAL INTELLIGENCE AND HUMAN RIGHTS: A BUSINESS ETHICAL ASSESSMENT

ALEXANDER KRIEBITZ[2]
I Technical University Munich. Munich. Alemanha

CHRISTOPH LÜTGE[3]
II Technical University Munich. Munich. Alemanha

MIRIAM WIMMER[4] **(Trad.)**
III Instituto De Ensino, Desenvolvimento E Pesquisa (Idp). Brasília (Df). Brasil

RESUMO: A inteligência artificial (IA) evoluiu como uma tecnologia disruptiva, gerando impactos para uma ampla gama de questões relacionadas aos direitos humanos, que vão desde a discriminação até a devida diligência na cadeia logística. Dadas as crescentes obrigações de direitos humanos das empresas e a intensificação do discurso sobre IA e direitos humanos, lançamos luz sobre as responsabilidades dos atores corporativos em termos de padrões de direitos humanos no contexto do desenvolvimento e uso da IA. Quais são as implicações das obrigações de direitos humanos para as empresas que desenvolvem e usam IA? Em nosso artigo, discutimos, primeiramente, se a IA possui um conflito intrínseco com os direitos humanos e com a autonomia humana. A seguir, discutimos como a IA pode estar ligada ao critério de beneficência, que integra a ética da IA, e como a IA pode ser aplicada em áreas relacionadas aos direitos humanos. Por fim, aprofundamo-nos sobre aspectos particulares do que significa estar em conformidade com os direitos humanos, abordando áreas problemáticas específicas da IA.

Palavras-chave: inteligência artificial, ética corporativa, privacidade, digitalização, direitos humanos.

ABSTRACT: Artificial intelligence (AI) has evolved as a disruptive technology, impacting a wide range of human rights-related issues ranging from discrimination to supply chain due diligence. Given the increasing human rights obligations of companies and the intensifying discourse on AI and human rights, we shed light on the

[1] Este artigo foi originalmente publicado em inglês na revista *Business and Human Rights Journal,* 5 (janeiro de 2020), pp. 84-104. O texto original está disponível em: https://doi.org/10.1017/bhj.2019.28 .

[2] Orcid: https://orcid.org/0000-0001-7959-5980

[3] Orcid: https://orcid.org/0000-0002-3870-4789

[4] Orcid: https://orcid.org/ 0000-0001-9210-6651

responsibilities of corporate actors in terms of human rights standards in the context of developing and using AI. What implications do human rights obligations have for companies developing and using AI? In our article, we discuss firstly whether AI inherently conflicts with human rights and human autonomy. Next, we discuss how AI might be linked to the beneficence criterion of AI ethics and how AI might be applied in human rights-related areas. Finally, we elaborate on individual aspects of what it means to conform to human rights, addressing AI-specific problem areas.

Keywords: artificial intelligence, corporate ethics, data privacy, digitization, human rights

I - INTRODUÇÃO

A inteligência artificial (IA) está, de maneira crescente, conquistando nossa realidade e moldando a forma como as sociedades e suas instituições são mantidas, organizadas e controladas, abrangendo ferramentas de reconhecimento facial, veículos autônomos, mecanismos de busca, ferramentas de tradução e programas que preveem a evolução dos preços nas bolsas de valores. Em comparação com tecnologias convencionais, a IA ocupa posição de destaque em termos de sua capacidade de interpretação e reação a dados, que (para fins de IA) são documentados, gerados e armazenados em dispositivos eletrônicos; os dados começam a se comunicar uns com os outros e a gerar o que chamamos de *big data*. Nesse sentido, podemos apropriadamente descrever a IA como uma constelação de diferentes processos e tecnologias (KAYE, 2017; COWLS et. al., 2019), conduzindo a uma substituição incremental das ações humanas pelo processamento automatizado de dados. Ainda que a IA claramente ofereça grandes vantagens para a humanidade, a exemplo de ferramentas diagnósticas mais precisas e medidas aprimoradas para combater o crime e limitar o terrorismo, os críticos apontam para os riscos que podem acompanhar essa revolução tecnológica. A Carta Aberta sobre IA de 2015, assinada por grandes cientistas e empresários, gerou um intenso debate sobre como regular a IA e como evitar potenciais armadilhas decorrentes da má gestão dessa tecnologia (SPARKES, 2015). Nesse contexto, Stephen Hawking se referiu à IA como potencialmente o pior evento da história humana, apto a anunciar o fim da humanidade (CELLAN-JONES, 2014), enquanto outras profecias sobre a tecnologia relacionada à IA soam tão ameaçadoras quanto os avisos dados na obra 1984, de Orwell, ou no livro Admirável Mundo Novo, de Huxley.

As incertezas que acompanham este período de mudança tecnológica requerem um intenso debate sobre como orientar o desenvolvimento futuro da IA, bem como sua ética e

governança; esses debates suscitam novas questões sobre o desenho de estruturas éticas e legislação em todo o mundo (FLORIDI et. al., 2018; RUSSEL, 2015). Embora a IA certamente contribua para a realização de vários objetivos sociais e ambientais, como os Objetivos de Desenvolvimento Social da Organização das Nações Unidas (ONU), ainda existe o risco de conflito entre os fundamentos normativos de nossa civilização e o uso concreto da IA. Assim, legisladores e especialistas em ética em todo o mundo começaram a desenvolver normas e padrões legais para lidar com casos potenciais de uso indevido de IA e para regulamentar o assunto. Estes incluem, por exemplo, a Declaração de Montreal para IA Responsável; os Princípios de Asilomar para IA; os princípios da AI4People para ética em IA (FLORIDI et. al, 2018); os relatórios dos dois Grupos de Especialistas de Alto Nível sobre relatórios de IA, tratando sobre ética (HLEG, 2019) bem como sobre governança de IA (FLORIDI et. al, 2018; COWLS e FLORIDI, 2019); o Comitê de Inteligência Artificial da Câmara dos Lordes do Reino Unido; o Regulamento Geral de Proteção de Dados europeu – RGPD; e o Código de Ética Alemão para Direção Automatizada e Conectada; que envolvem aspectos importantes de questões éticas relacionadas à IA.

Afora o Relatório da ONU sobre Inteligência Artificial e suas implicações para os direitos humanos, ainda não foi articulada uma codificação legal ou ética adaptada à aplicação da IA no contexto dos direitos humanos. Os direitos humanos, no entanto, desempenham um papel essencial no contexto da governança da IA, visto que são considerados normas fundamentais da civilização ocidental e desempenham um papel cada vez maior, em geral, no direito internacional (SIMMA e PULKOWSKI, 2006).

Para além dos deveres éticos dos Estados e de organizações internacionais, como a ONU, na salvaguarda e proteção dos direitos humanos, o foco dos direitos humanos tem, de maneira mais proeminente após a formulação dos Princípios de Ruggie (ONU, 2011)[5], gravitado em direção à sua aplicação pelas empresas. Dadas as crescentes obrigações e deveres das empresas como responsáveis pela tutela dos direitos humanos (JÄGERS, 2000; WETTSTEIN, 2009; RUGGIE, 2013; MUCHLINSKY, 2002), este artigo lança luz sobre as responsabilidades dos atores corporativos para a aplicação e realização dos padrões de direitos

[5] Nota da Tradutora: os Princípios de Ruggie, endossados em 2011 pelo Conselho de Direitos Humanos da ONU (Resolução 17/4), estabelecem diretrizes orientadores para Estados e empresas no que tange à proteção de direitos humanos.

humanos no contexto da IA. Quais são as implicações das obrigações de direitos humanos para as empresas que desenvolvem e usam IA?

Em um sentido mais amplo, nosso artigo tem como objetivo conectar o discurso sobre a ética da IA ao discurso sobre as obrigações de direitos humanos dos tomadores de decisão corporativos. Em nossa opinião, ambos os discursos não representam pontos de vista concorrentes ou exclusivos, mas se complementam e enriquecem à medida que integram os domínios mais amplos da ética empresarial e da ética da tecnologia.

II. O QUE SÃO DIREITOS HUMANOS E POR QUE SÃO IMPORTANTES?

Abordar a IA a partir de uma perspectiva de direitos humanos requer uma breve descrição do conceito de direitos humanos. No pensamento ocidental, os direitos humanos são considerados a norma suprema da lei e constituem a base da maioria dos sistemas jurídicos. De acordo com a maior parte dos especialistas em direito internacional (SIMMA e PULKOWSKI, 2006; GARDBAUM, 2008), os direitos humanos não são meramente uma enumeração dos direitos individuais, mas formam um regime independente. O pilar integrante deste regime é uma antropologia baseada na autodeterminação e autonomia do ser humano (KANT, 1998).

De acordo com esse entendimento, os direitos humanos obrigam o Estado e outras organizações sociais a observar certos princípios e procedimentos no trato com subalternos; esses princípios abrangem, por exemplo, a adesão estrita ao princípio do Estado de Direito e o direito a um julgamento justo. Ao mesmo tempo, a filosofia dos direitos humanos compreende a liberdade como condição básica do ser humano, concluindo que restrições a essa liberdade devem servir ao bem comum, e não à vontade de um monarca ou tirano. Esse conceito corresponde em grande parte à noção defendida por Isaiah Berlin (1969), que definiu a liberdade como "a ausência de obstáculos às escolhas e atividades possíveis", e que contribuiu para a compreensão dos direitos humanos como "pretensões" *(claims)* que limitam o poder do Estado (HOHFELD, 1917).

Nessas circunstâncias, as intervenções na autonomia do indivíduo só são legítimas se forem baseadas no consentimento do indivíduo em questão ou se a liberdade de um indivíduo entrar em conflito com o interesse de outros. A transferência de bens ou a efetivação de um tratamento médico - uma intervenção na integridade inviolável do corpo - só são lícitas se gozarem do consentimento explícito do indivíduo, ressalvadas algumas isenções necessárias,

como as emergências. A principal exceção que permite restringir a liberdade de uma pessoa exige que tal restrição sirva para prevenir danos a terceiros. De acordo com o princípio do dano, que forma a base dos direitos humanos enquanto pretensões que explicitamente vinculam instituições e terceiros, "o único propósito pelo qual o poder pode ser legitimamente exercido sobre qualquer membro de uma comunidade civilizada, contra sua vontade, é prevenir danos a outros" (MILL, 1859).

Como resultado, as incursões do Estado, limitando a liberdade dos indivíduos, enfrentam restrições substanciais e só são legítimas em casos de colisão de normas. Desse modo, a regulação do trânsito, na medida em que constitui uma limitação à liberdade individual, serve para minimizar os acidentes de trânsito, em última análise decorrentes da tarefa do Estado de proteger a vida humana.

No entanto, de acordo com o princípio da proporcionalidade, a interferência do Estado deve ser proporcional ao dano evitado. Essa noção deriva da alta prioridade concedida à ideia de igualdade perante a lei, que decorre da ideia aristotélica de "justiça corretiva" (AMBROSI, 2007; POJMAN, 1996), implicando que o dano gerado por um regulamento ou por um ato do Estado deve ser proporcional ao dano evitado. Um aspecto importante deste termo é o conceito weberiano de *Augenmaß* (WEBER, 1919) - um senso de proporção ou de "bom senso" - que deve nortear as ações políticas e os legisladores (ZUCCA, 2007; HARBO, 2010; CRAIG e DE BRUCA, 2015). Prender uma criança por roubar uma maçã ou confiscar a carteira de motorista de uma pessoa após ela ter excedido o limite de velocidade em 5 km/h seriam exemplos de interferência desproporcional nas liberdades dos indivíduos.

No entanto, há casos em que o princípio da proporcionalidade não é aplicável, uma vez que não é possível derrogar determinados tipos de direitos humanos. De acordo com a corrente majoritária dos especialistas em direitos humanos, a própria natureza dos direitos humanos veda determinadas ações, como tortura, escravidão, estupro ou comportamento extremamente humilhante, ainda que sirvam a outros direitos fundamentais (ONU, 2004).

A partir dessa perspectiva, derivamos algumas consequências básicas de direitos humanos para a regulamentação da IA:

(a) Os direitos de um indivíduo só podem ser transferidos com o seu consentimento (princípio do consentimento).

(b) A única justificativa para o uso do poder contra a vontade de uma pessoa é a

prevenção do dano (princípio do dano).

(c) O uso da força deve ser proporcional à ameaça (princípio da proporcionalidade).

Os direitos humanos, entretanto, vão além de uma natureza puramente defensiva, para influenciar os objetivos das organizações sociais, como na forma de "objetivos estatais" ou imperativos (KUSZLER, 2007). O Estado, como autoridade normativa máxima, e as organizações internacionais enfrentam, assim, a tarefa de garantir a observância dos direitos humanos em bases materiais e econômicas. O Pacto Internacional sobre Direitos Sociais, Econômicos e Culturais, por exemplo, contém o princípio de implementação progressiva *(progressive realization)*, que insta os Estados a adotar "políticas e técnicas para alcançar um desenvolvimento econômico, social e cultural estável [...]" (ONU, 1966, Artigo 6.2). Por fim, a noção de direitos humanos e a ideia de que os seres humanos nascem livres e iguais requerem que a *participação* seja ancorada no processo político. Essa noção é, grosso modo, equivalente ao princípio lockeano de governo, sustentado pelo "consentimento dos governados" mencionado na Declaração de Independência dos Estados Unidos.

Devido ao papel crescente das empresas no direito internacional na esteira da globalização, John Ruggie (2013) leciona que as empresas se tornaram parte do discurso dos direitos humanos (ZERK, 2006:19). O Pacto Global da ONU, a Declaração da OIT sobre Princípios e Direitos Fundamentais no Trabalho, a Lei de Escravidão Moderna e Tráfico Humano do Reino Unido, bem como os Princípios Orientadores da ONU sobre Negócios e Direitos Humanos (*United Nations Guiding Principles* ou UNGPs) influenciaram amplamente esse desenvolvimento. Os UNGPs, de acordo com Ruggie, estabelecem que as empresas devem "identificar e avaliar qualquer impacto adverso real ou potencial sobre os direitos humanos com o qual possam estar envolvidas, seja por meio de suas próprias atividades, seja como resultado de relações comerciais". Portanto, chama-se a atenção para as responsabilidades das empresas em relação às mencionadas pretensões *(claims)* em um contexto internacional (ONU, 2011).

III. O QUE DISTINGUE A IA DE OUTRAS TECNOLOGIAS?

A discussão da relação entre direitos humanos e IA requer um exame das propriedades e peculiaridades da IA. O que torna a IA diferente de outras tecnologias, como veículos tradicionais, *smartphones* ou usinas nucleares? Por que, afinal, precisamos de uma ética sob

medida para IA? A definição mais próxima do nosso entendimento de IA é dada pelo Dicionário Merriam-Webster, que define inteligência artificial como a capacidade de uma máquina imitar o comportamento humano inteligente. A rigor, a palavra "inteligente" não se refere à máquina, mas sim ao fato de que se a tarefa da solução de IA tivesse sido resolvida por um ser humano, o modo de realizar a tarefa teria sido chamado de inteligente (KAYE, 2017). A comparação à inteligência humana refere-se, portanto, em primeira instância, ao resultado *(output)* de uma ação, e não ao *input* ou ao processo de tomada de decisão das máquinas (MCCARTHY, 1959). Quanto a esse ponto, a Carta Aberta sobre IA (RUSSEL et. al., 2015) se refere às noções estatísticas e econômicas de inteligência. Esse entendimento tem uma implicação importante para os direitos humanos, uma vez que a IA - não sendo uma entidade ontológica - não pode ser considerada um ator independente ou potencial perpetrador de violações de direitos humanos, pelo menos não ainda. Ao invés, as obrigações de conformidade com os direitos humanos relacionadas às soluções de IA permanecem no domínio da responsabilidade humana, e buscam vincular os Estados-nação, empresas ou organizações não governamentais (ONGs) que utilizem tais tecnologias.

Com base nessa conceituação de IA, as violações dos direitos humanos podem ter origem em diferentes impulsos e inclinações. Ao imitar o comportamento inteligente, a IA combina grandes quantidades de dados com processamento iterativo rápido e algoritmos inteligentes, permitindo que o *software* aprenda automaticamente a partir de padrões ou características dos dados. Nesse sentido, a IA utiliza redes neurais para evitar a necessidade de coleta e interpretação de dados, podendo, portanto, chegar a conclusões não previstas por humanos, uma vez que estes não participam da definição dos objetivos e dos *outputs* da IA (KAYE, 2017). Nesse sentido, a característica especial da IA é que alguns de seus processos são executados automaticamente e, nesses casos, os humanos não podem intervir diretamente; seus resultados também não podem ser previstos *ex ante*, resultando em consequências não previstas.

Essas características suscitam questões mais gerais quanto à sua conformidade com os direitos humanos, a saber, se a IA, na medida em que consiste na transferência da agência humana para uma máquina, acarreta um conflito intrínseco com a ideia de autodeterminação moral. Caso não exista um conflito intrínseco, permanecem dúvidas sobre as maneiras como

certas características e aspectos da IA podem afetar os direitos humanos. As perguntas a seguir lançam luz sobre alguns desses problemas:

- Existem cenários em que a IA produz um impacto positivo sobre os direitos humanos?
- Existem cenários em que os dados de entrada *(input)* de IA violam os direitos humanos?
- Existem cenários em que os resultados *(output)* de IA violam os direitos humanos?
- O uso de IA em domínios específicos viola os direitos humanos, especialmente no que tange aos direitos de participação?
- É possível usar a IA para violar ou restringir os direitos humanos?

IV. INTELIGÊNCIA ARTIFICIAL E DIREITOS HUMANOS - UM CONFLITO INTRÍNSECO?

Na primeira parte de nossa análise, abordamos a questão de saber se existe um conflito intrínseco entre o uso de IA e os direitos humanos. Isso requer um esclarecimento quanto ao termo "intrínseco" aqui utilizado. Em geral, podemos distinguir entre atos que conflitam de maneira intrínseca com os direitos humanos e atos que representam um conflito fortuito com os direitos humanos (BRANNON e JAWORSKI, 2015; MUELLER, 2018). Em nossa opinião, um ato é intrinsecamente contrário aos direitos humanos se constituir uma violação dos direitos humanos independentemente das circunstâncias. Um exemplo paradigmático é a escravidão, que coloca em xeque a natureza dos direitos humanos e da autodeterminação, independentemente de suas circunstâncias e causas exatas (cfr. Artigo 4, Pacto Internacional sobre Direitos Civis e Políticos). A construção de uma estrada pode ser um exemplo para ilustrar a diferença entre limites intrínsecos e fortuitos. Normalmente, não há nada de errado em construir estradas. Se, no entanto, o governo constrói a estrada com trabalho forçado ou pela expropriação de tribos indígenas, há uma clara violação dos direitos humanos. Casos desse tipo constituem violações de direitos humanos por motivos fortuitos, pois sua avaliação ética não depende da ação em si, mas de suas circunstâncias.

O uso de IA constitui uma violação intrínseca dos direitos humanos, justificando assim uma proibição universal? A razão pela qual abordamos esta questão um tanto teórica deve-se à

rejeição geral do uso de IA por alguns especialistas em ética e entidades religiosas, postulando a existência de um conflito intransponível entre a autodeterminação moral e o uso de IA. Uma linha de argumento frequente é que o uso da IA representa um conflito com a autonomia humana, porque mesmo decisões graves podem ser tomadas pela IA, entrando assim em conflito direto com o próprio significado dos direitos humanos e conduzindo à alienação (WOGU, 2018). Na declaração de princípios da Convenção de Inteligência Artificial da igreja Southern Baptist *("An Evangelical Statement of Principles")*, a igreja se posicionou contra a atribuição de um nível de identidade humana, valor, dignidade ou agência moral à IA (ERLC, 2019).

Para analisar se essa afirmação é válida, nos referimos a um dos exemplos mais controversos, a saber, o uso de IA para sopesar decisões de vida ou morte. Na direção autônoma, como em outras aplicações, ações como dirigir o veículo ou desacelerar, originalmente realizadas por humanos, são cada vez mais gerenciadas por processos mecânicos e automáticos. Além disso, as soluções de *software* podem intervir, em caso de acidente, para minimizar o número de vítimas. No caso de um acidente inevitável, se o carro tiver que escolher entre atingir um dos dois indivíduos que atravessam uma rua, o carro autônomo passa a ser a entidade responsável pela decisão de qual alvo atingir. Além das considerações gerais relativas à prestação de contas e responsabilidade, a permissibilidade da automação em tais casos é importante para a perspectiva de direitos humanos,[6] vez que se refere à questão de saber se a IA pode tomar decisões que envolvem a vida ou a morte de pessoas. Isso, por sua vez, se conecta ao debate mais amplo sobre a relação entre dignidade humana e automatização. Uma posição deontológica estrita indicaria que sopesar vidas não pode ser legítimo, com base no pressuposto de que entra em conflito com a ideia de dignidade humana, que estabelece que os seres humanos não devem ser tratados como objetos; ao passo que considerações utilitárias incitariam o programador a escolher a alternativa com menos vítimas.

Do nosso ponto de vista, entretanto, um argumento prático e outro teórico depõem contra a alegação de que o uso de IA em tais casos constitui uma violação intrínseca dos direitos humanos. Na prática, a decisão de um determinado motorista geralmente não resulta de um processo de pensamento bem elaborado, mas sim de uma reação inconsciente ou de pânico.

[6] A questão se vincula ao Princípio de Asilomar "Não-subversão e controle humano".

Portanto, é questionável se o ato realmente envolve um momento de ação humana. Depois de atingirem um certo nível de desenvolvimento, os processos sensoriais e mecânicos podem ser superiores às reações humanas, pois podem reagir muito mais rapidamente do que qualquer cérebro humano - mesmo em crise. Do ponto de vista teórico, o conceito rawlsiano de um véu de ignorância (RAWLS, 1971) oferece uma saída para o dilema, ao demonstrar que o princípio do consentimento pode se harmonizar com a maximização do direito à vida. Se imaginarmos o cenário concreto de um acidente inevitável que envolve a morte do motorista ou a morte de um grupo de pessoas (WALLACH e ALLEN, 2008) e, além disso, assumimos que todos os indivíduos desejam primeiro salvar suas próprias vidas; decorre que as partes envolvidas não seriam capazes de chegar a um consenso unânime. Devido ao alto valor da vida, também é questionável se um indivíduo estaria disposto a desistir de sua vida pelo bem dos outros, apesar dos achados empíricos (DI NUCCI, 2013). Obviamente, sem sacrificar o princípio do consentimento unânime, a perspectiva de se chegar a um acordo nesta situação concreta está fadada ao fracasso. Embora a aplicação de uma decisão baseada na vontade - hipotética - da maioria seja possível, de acordo com algumas teorias amplamente utilitárias, tal possibilidade permanece controversa (GOODALL, 2014; AWAD et al, 2018).

A única maneira de resolver essa situação de impasse é mudar o foco do caso individual para uma regra geral. Dada a premissa de que indivíduos anônimos não sabem sua posição com antecedência e devem concordar *ex ante* quanto a um procedimento (RAWLS, 1971) sobre como lidar com o uso de IA em situações inevitáveis, eles provavelmente concordarão com princípios abstratos e imparciais. O fato de os indivíduos não terem consciência de seu papel exato no cenário estabelece um ambiente propício à imparcialidade, garantindo "que ninguém tenha vantagens ou desvantagens na escolha dos princípios pelo resultado do acaso natural ou pela contingência das circunstâncias sociais" (RAWLS, 1971). Com base nisso, indivíduos razoáveis podem propor um regulamento segundo o qual um carro deve ser programado para minimizar o número de vítimas em situações de dilema inevitáveis, visto que a probabilidade geral de ser atropelado ou morto em um carro autônomo diminui com a redução do número de vítimas. A regra deve, portanto, implicar que o maior grupo de pessoas é sempre salvo, após ponderar as decisões, independentemente de sua condição de motoristas, pedestres ou outras características pessoais. Tal regra está de acordo com o princípio da dignidade humana, que afirma que todos os indivíduos envolvidos devem ter oportunidades iguais de vida. Elevar a

questão da programação de acidentes inevitáveis ao nível de participação social parece ser mais significativo, pois a formulação da regra tem impacto sobre todos os indivíduos que participam do tráfego rodoviário. A escolha exata dos mecanismos e a decisão sobre a aplicação da randomização nos casos em que o carro deve decidir entre dois grupos do mesmo tamanho é, em última análise, uma questão de consentimento social e democrático. Isso implica que as empresas são incapazes de formular suas próprias interpretações das estruturas de direitos humanos para tais instâncias e precisam se referir às estruturas jurídicas abrangentes ou às decisões dos tribunais constitucionais. Além disso, a programação deve distinguir entre situações que envolvem pessoas anônimas e situações em que as pessoas se conhecem para evitar danos extremos (veja-se a Regra 3). Portanto, a decisão final, talvez na forma de uma decisão parlamentar e finalmente na sua implementação pelas empresas, deve cumprir os seguintes critérios:

- **Regra 1:** A probabilidade de morrer em um acidente deve diminuir para todas as pessoas.
- **Regra 2:** Todos os indivíduos devem ter chances iguais de sobreviver.
- **Regra 3:** Situações extremamente difíceis para os indivíduos podem anular a Regra 2, com base no consentimento explícito das partes afetadas (por exemplo, a avó consente que o carro a atropele em vez de seu neto.)

O princípio da "meta-autonomia", que se refere à decisão de delegar decisões específicas às máquinas (FLORIDI et. al, 2018), não é um conceito muito novo, uma vez que já delegamos voluntariamente a liberdade a organizações e normas coletivas no sentido de um "consentimento dos governados". É possível traçar paralelos entre o supramencionado exemplo de direção autônoma e a noção de delegar liberdades pessoais a organizações sociais, a elaboração de um testamento, ou considerações quanto a danos colaterais em conflitos armados. Nesse sentido, o exemplo também leva a uma conclusão relacionada à relação geral entre a agência humana e a IA. A transferência da decisão para o veículo autônomo na situação concreta do sinistro não representa conflito com a agência humana, desde que seja com base em recurso a princípios gerais legitimados pelo consentimento democrático e assente em regras racionais e abstratas, as quais tenham sido estabelecidas em um discurso aberto.

V. BENEFICÊNCIA E DIREITOS HUMANOS

Nesta seção, examinamos a direção em que o desenvolvimento e a pesquisa de IA por empresas devem seguir a partir de uma perspectiva de direitos humanos. Pretendemos vincular o debate existente sobre o princípio da beneficência nas regulamentações da IA às implicações socioeconômicas dos direitos humanos. Embora os direitos humanos sejam classicamente definidos como pretensões *(claims)* face ao Estado, pretendemos lançar luz sobre as implicações mais amplas dos direitos humanos para a implementação das normas de direitos humanos *(human rights standards)*, bem como suas implicações para os fatores socioeconômicos que podem estar ligados aos direitos humanos (RAWLS, 1971; BANCO MUNDIAL, 2012).

A ideia de que o uso de tecnologias científicas deve servir ao bem maior decorre do critério de beneficência da IA, que está presente em diferentes marcos regulatórios e diretrizes éticas. Nesse sentido, os critérios de Pequim sobre Inteligência Artificial são bastante representativos quanto à opinião de que "a IA deve ser projetada e desenvolvida para promover o progresso da sociedade e da civilização humana [...]" (BAAI, 2019).

A ideia de beneficência, no sentido de que o desenvolvimento econômico e científico deve contribuir para o atingimento de objetivos normativos, se enquadra na estrutura dos Objetivos de Desenvolvimento Sustentável – ODS das Nações Unidas, que abrangem uma ampla gama de objetivos sociais e ambientais, como o de "acabar com a pobreza em todas as suas formas, em todos os lugares". Usar os ODS como um *proxy* para os direitos humanos pode ser um tanto problemático, pois há um debate em andamento atualmente sobre a relação exata entre as duas estruturas.[7] No entanto, argumentamos que alguns desses objetivos correspondem aos direitos socioeconômicos enumerados no Pacto Internacional sobre Direitos Econômicos, Sociais e Culturais, tais como o direito a um padrão de vida adequado, o direito à educação ou o direito à saúde (ONU, 2008b). A Agenda 2030 para o Desenvolvimento Sustentável apontou que os ODS das Nações Unidas "buscam realizar os direitos humanos de todos" e se referiu aos direitos socioeconômicos da Declaração dos Direitos Humanos das Nações Unidas (ONU, 2015). Na próxima seção, ilustramos, portanto, como as empresas podem vincular suas

[7] O debate também está relacionado ao papel geral dos direitos humanos e das obrigações econômicas correspondentes. Alguns postulam que os direitos humanos se desdobram em obrigações econômicas (por exemplo, Universal Rights Group [2017]); outros afirmam que os direitos humanos não resultam em uma dimensão econômica ou o fazem em menor grau (NOZICK, 1974).

estratégias de IA à realização dessas questões socialmente importantes e como integrar a realização dos direitos humanos consagrados em tratados internacionais em abordagens corporativas relacionadas à IA.

A. OBJETIVO DE DESENVOLVIMENTO SUSTENTÁVEL 1: ELIMINAÇÃO DA POBREZA

De acordo com uma publicação recente, a IA viabiliza tecnologias que desencadeiam o crescimento econômico e aumentam a produtividade da economia (VINUESA et. al., 2019; ACEMOGLU e RESTREPO, 2018). Nesse sentido, a IA parece ter consequências positivas diretas para a pobreza e para a prosperidade global, que são, segundo as Nações Unidas, os maiores desafios globais contra a humanidade (ONU, 2017). Além das implicações gerais da IA para o crescimento econômico, muitas das implicações positivas da IA são mais diretas. O uso de *drones* na agricultura, por exemplo, ajuda os agricultores a trabalhar, produzir e manter suas fazendas e gado de forma eficiente. O Stanford Poverty & Technology Lab tem feito pesquisas intensivas para encontrar soluções para os agricultores de baixa renda, tanto na agricultura como também em todos os outros aspectos que permitem tirar os humanos da pobreza. O Laboratório também usa IA e imagens para prever a pobreza, e essas previsões foram confirmadas com 81-99% de precisão (BENNINTON-CASTRO, 2017).

B. OBJETIVO DE DESENVOLVIMENTO SUSTENTÁVEL 5: IGUALDADE DE GÊNERO

A IA também pode ser usada para conduzir à igualdade de gênero e ao empoderamento de todas as mulheres e meninas (ONU, 2015). No Paquistão, o *chatbot* RAAJI conversa com mulheres sobre saúde reprodutiva feminina, higiene e segurança. A educação e a igualdade de gênero não são apenas um direito humano e uma meta dos ODS, mas também a principal ferramenta para valorizar outros tópicos relevantes para os direitos humanos, como a igualdade, a liberdade pessoal ou a dignidade humana. A empresa que viabiliza o *chatbot* tem parceria com a UNESCO para criar conteúdo veiculado nas áreas rurais do Paquistão.

Embora os dois casos descritos representem apenas alguns casos de IA benéfica, eles evidenciam que as empresas podem contribuir para o bem maior ao implementar soluções de IA. Isso se vincula ao objetivo normativo da beneficência no sentido de contribuir para os

objetivos econômicos, sociais e ambientais. Na perspectiva de longo prazo, o critério de beneficência derivado da governança da IA pode influenciar o entendimento geral quanto às responsabilidades corporativas no tocante aos direitos humanos, no sentido de que as empresas se tornam cada vez mais comprometidas com aprimoramentos éticos. Isso pode ser comparável às implicações para os Estados, que têm de elevar o nível de desenvolvimento do respectivo país. Até agora, as Diretrizes da ONU sobre Negócios e Direitos Humanos consideram as responsabilidades das empresas principalmente em sua dimensão de pretensões *(claims)*[8]. Portanto, a incorporação do critério de beneficência na ética da IA pode contribuir para uma interpretação dos direitos humanos como objetivos normativos para a tomada de decisões corporativas.

VI. VIOLAÇÃO FORTUITA DE DIREITOS HUMANOS POR INTELIGÊNCIA ARTIFICIAL

Na seção a seguir, discutiremos os casos em que o uso de IA pode entrar em conflito com os direitos humanos. Os casos aqui descritos são exemplificativos e não representam uma lista exaustiva.

Tabela 1: tipologia de Violações de Direitos Humanos relacionadas à IA

Situação	**Exemplos**
Situação I: os dados de entrada *(input)* de IA conflitam com os direitos humanos	• Uso de dados sem ou contra a vontade explícita dos clientes • Uso desproporcional de dados íntimos e pessoais de indivíduos por instituições públicas
Situação II: o resultado *(output)* da IA leva a violações involuntárias dos direitos humanos	• Discriminação ilegal em candidaturas a empregos com base na etnia • Discriminação ilícita de mulheres no sistema público de saúde
Situação III: o uso de IA em áreas específicas	• Violação do direito de opinião, devido ao uso excessivo de algoritmos nas redes

[8] Os Princípios de Ruggie falam explicitamente em 'proteger, respeitar e remediar', o que, em grande parte, confirma a natureza de direitos de defesa ou de pretensões dos direitos humanos.

conflita com os direitos humanos	sociais • Substituição de decisões democráticas por decisões de IA (robotocracia)
Situação IV: um violador de direitos humanos usa IA	• Uso de IA para monitorar os cidadãos que criticam o governo • Uso de IA para suprimir minorias étnicas e rastrear indivíduos

Em contraste com a seção acima, discutimos aqui não o que a IA deveria fazer, mas, sim, o que não deveria ser feito em termos de direitos humanos. Para facilitar a compreensão dos aspectos individuais do que significa estar em conformidade com os direitos humanos, nos orientamos para diversas situações, abordando áreas problemáticas específicas da IA. Os aspectos centrais são a conformidade com os direitos humanos do *input* e do *output* de soluções de IA, o tipo de uso e as intenções do ator (Tabela 1).

A. SITUAÇÃO I: OS DADOS DE ENTRADA *(INPUT)* DE IA CONFLITAM COM OS DIREITOS HUMANOS

A IA precisa processar dados para expandir suas capacidades e realizar certas tarefas. Compreender essa ligação estreita entre IA e dados é crucial para casos práticos, pois a coleta de dados pode entrar em conflito com o direito dos indivíduos à privacidade e com a sua autonomia de dados (DATEN ETHIK KOMMISSION, 2019). Do ponto de vista dos direitos humanos, o direito à privacidade pode ser considerado uma extensão da dignidade humana, o que foi confirmado por decisões judiciais (cfr. *Lawrence v. Texas*) e textos legais como a Declaração Universal dos Direitos Humanos e a Convenção Europeia sobre Direitos Humanos (artigo 8) (ETZIONI, 2007).

Diferentemente do que ocorre com a dignidade humana, a propriedade dos dados geralmente pode ser transferida a terceiros com base no consentimento, o que também se aplica à forma como os dados são usados (LESSIG, 2002). O Regulamento Geral de Proteção de Dados – RGPD estabelece que "sempre que o tratamento for realizado com base no consentimento do titular dos dados, o responsável pelo tratamento deverá poder demonstrar que o titular deu o seu consentimento à operação de tratamento dos dados". Essa passagem reflete,

sobretudo, o aumento do uso de dados pela IA, como ocorre no caso de veículos altamente automatizados e, especialmente, conectados: as interações entre os veículos aumentariam substancialmente a segurança no trânsito e - pelo menos, em princípio - provavelmente seriam objeto do consentimento de todas as partes envolvidas. Embora a proteção de dados no caso de veículos conectados sirva para aumentar a segurança do veículo, os interesses comerciais podem ser preponderantes no que tange ao acúmulo de dados em outras áreas. Isso não é problemático em si, desde que a transação seja baseada no consentimento de ambas as partes. No entanto, a realização objetiva do princípio do consentimento pode ser dependente do ambiente socioeconômico, que precisa ser abordado no nível regulatório ou por meio de padrões autorregulatórios para todo o setor (PAGALLO *et. al*, 2019).

De grande importância neste contexto é o princípio do consentimento informado, que foi integrado em muitos *frameworks* que lidam com IA (FUTURE OF LIFE INSTITUTE, 2017). A noção de consentimento informado implica que "medidas devem ser tomadas para garantir que os interessados nos sistemas de IA tenham consentimento informado suficiente sobre o impacto do sistema sobre seus direitos e interesses" (BAAI, 2019). Ao mesmo tempo, uma quantidade excessiva de dados no local errado pode levar a um resultado prejudicial, não pretendido pelos tomadores de decisões corporativas. A proporção de judeus mortos na Holanda durante o Holocausto foi relativamente alta porque a Câmara Municipal de Amsterdã realizou um censo populacional detalhado, que registrou estatisticamente a religião dos habitantes. O acesso a esses dados permitiu que a Gestapo alemã transferisse cidadãos judeus para campos de concentração (STELTZER, 1998). Para mitigar futuras violações dos direitos humanos, as empresas devem respeitar o princípio da minimização de dados, que consiste em não coletar mais informações pessoais do que as necessárias para uma finalidade específica[9]. Esse princípio pode ser integrado com o princípio da responsabilidade previdente *(foresighted responsibility)* (DATEN ETHIK KOMMISSION, 2019), que significa que as empresas precisam levar em consideração os efeitos de rede e as constantes mudanças nos arranjos entre atores. Desse modo, a devida diligência de uma empresa suscita questões relacionadas não apenas à origem dos dados, mas também ao seu destino final e a casos de uso futuros.

[9] Compare com o art. 5º do Regulamento Geral de Proteção de Dados da Europa.

A cooperação com as autoridades públicas, mediante o fornecimento ou recebimento de dados, representa talvez o desafio mais importante para as empresas no âmbito das violações de direitos humanos relacionadas ao *input* de dados. O principal motivo é que a relação entre o receptor e o provedor de dados é assimétrica e que o direito à privacidade pode ser derrogado no caso de conflitos de normas (REIGADA, 2012). O uso de dados pela polícia ou por unidades de investigação, por exemplo, requer o equilíbrio do direito à privacidade com o interesse do público em geral na investigação de infrações penais ou administrativas. No entanto, nem todos os fins justificam meios específicos, já que o uso de tais *inputs* de dados para crimes "menores", como consumo de drogas, evasão fiscal ou trabalho não declarado, tornaria obsoleta a noção de privacidade como um direito de defesa face ao Estado. O uso de tecnologias de vigilância pela China na província de Xinjiang aponta para o uso indevido de dados por autoridades estatais (UHRP, 2018). A orientação para o princípio da proporcionalidade e a aplicação estrita da necessidade podem, portanto, fornecer uma base importante para garantir a aplicação lícita da Inteligência Artificial, conduzindo às seguintes regras gerais para regular e autorregular o *input* de dados na IA:

- A transferência de dados pelas empresas deve estar alinhado com o consentimento das partes envolvidas e deve considerar a alteração das constelações de atores.
- Os dados só podem ser usados por soluções de IA com o consentimento das partes envolvidas, de modo a reduzir o dano a terceiros (aplicação do princípio do dano).
- A invasividade da solução de IA deve ser proporcional aos seus objetivos. O acesso ao *input* de dados deve se dar da maneira menos invasiva possível.

Tal perspectiva é importante para as empresas, visto que podem interagir com entidades públicas em infraestruturas críticas como vigilância, formação de perfis ou reconhecimento facial. A aplicação iBorderCTRL, que verifica a confiabilidade dos passageiros de voos em aeroportos europeus, pode ser um exemplo da cooperação entre entidades públicas e empresas de desenvolvimento de IA envolvendo dados pessoais críticos. Devido à proximidade com tópicos sensíveis aos direitos humanos, as empresas que operam em tais ambientes precisam estar atentas a possíveis violações aos direitos humanos e desenvolver seus próprios códigos de conduta em termos de uso de dados. Para avaliar a probabilidade de conflitos com os direitos humanos, podem ser traçados paralelos com outras ferramentas avançadas de investigação, como perfis de DNA (SIMONCELLI e WALLACE, 2006), o estabelecimento de bancos de

dados de DNA (WALLACE, 2006; WALLACE et. al, 2014) ou o uso de polígrafos pelos órgãos responsáveis pela aplicação da lei, que são parcialmente percebidos como violações excessivas e desproporcionais da privacidade se usados contra ou sem o consentimento ou conhecimento do suspeito (JOHNSTON, 2016).

A aplicação de IA em sistemas judiciais também é altamente crítica. Um limite dos instrumentos de interrogatório seria o princípio jurídico do *nemo tenetur se ipsum accusare*, segundo o qual ninguém pode ser forçado a acusar a si mesmo. O estatuto jurídico do chamado direito ao silêncio foi consagrado na Convenção Europeia dos Direitos do Homem (Artigo 6) e é relevante no direito processual penal de muitos países. De acordo com algumas jurisdições (Tribunal Federal de Justiça da Alemanha, 1954), o uso de polígrafos seria até qualificado como uma violação absoluta dos direitos humanos, indicando que o uso de tecnologias de IA para detectar a confiabilidade de uma pessoa (Tribunal Federal de Justiça, 2003), ao usar dados muito pessoais, enfrenta grandes barreiras. A comparação entre os efeitos da IA e técnicas mais convencionais pode ser relevante nesse ponto, uma vez que a IA não constitui uma novidade absoluta em termos de invasividade e considerando que comparações semelhantes entre tecnologias convencionais e modernas já foram feitas no contexto da regulação da guerra cibernética. Nessas situações, as empresas podem não apenas fazer referência à legislação nacional, mas também vincularem-se aos direitos humanos reconhecidos internacionalmente, devido à criticidade do *input* de dados utilizado.

B. SITUAÇÃO II: O RESULTADO *(OUTPUT)* DA IA CONDUZ A VIOLAÇÕES INVOLUNTÁRIAS DOS DIREITOS HUMANOS

Nesta seção, fazemos referência aos abusos de direitos humanos originados do desalinhamento entre os objetivos da máquina e a sua implementação (FLORIDI et. al., 2018). Como ocorre com toda tecnologia, consequências indesejadas para uma tecnologia específica podem ter resultados devastadores. O exemplo do *chatbot* da Microsoft, Tay, expõe as consequências potenciais das falhas de *design*. Originalmente projetado para imitar os padrões de linguagem de uma garota americana de 19 anos, o robô de bate-papo acabou elogiando Hitler e incitando o ódio (HUNT, 2016). Ademais, exemplos de erros resultantes de programação ou treinamento incorreto de IA que afetam os direitos humanos variam de riscos de segurança em veículos autônomos a problemas de discriminação em *softwares* de recrutamento de

trabalhadores. Muitos destes casos podem constituir violações não intencionais dos direitos fundamentais, mas antes falhas técnicas ou atos de negligência. Nesse sentido, os vieses *(biases)* podem constituir uma nova forma de violação dos direitos humanos, em que o perpetrador não tem interesse em violar os direitos humanos. Os impactos dessas falhas, vieses e erros sobre os direitos humanos, entretanto, não devem ser subestimados. Angwin *et al.* (2016), por exemplo, descobriram que as soluções de IA operadas pela polícia discriminam pessoas negras, enquanto a Amazon deixou de utilizar uma solução de IA que era tendenciosa contra mulheres em candidaturas a empregos de natureza técnica. Obermeyer et al. (2019) apontaram que o sistema de saúde dos EUA confiava em um algoritmo para orientar as decisões de saúde que foi afetado por um viés, levando à discriminação contra americanos negros.

Vieses que resultam na discriminação ilícita de indivíduos são os exemplos mais representativos de violações involuntárias dos direitos humanos. As razões pela discriminação acidental por soluções de IA variam, mas uma das principais fontes de falhas e violações de direitos humanos desse tipo é que a IA frequentemente é incapaz de diferenciar causalidade de correlação. Além disso, os problemas relativos à discriminação por IA muitas vezes se relacionam à forma como a "variável alvo" *(target variable)* e os "rótulos de classe" *(class labels)* são calibrados, como os dados de treinamento são rotulados e coletados e à seleção de recursos e *proxies*. Diferentemente de outras situações em que a IA e os direitos humanos entram em conflito, neste caso há a vantagem de que as empresas já começam a entrar no radar da lei, pois a maior parte dos países proíbe a discriminação com base em fatores como gênero e religião. Juridicamente, os arcabouços normativos voltados a combater a discriminação também consideram casos de discriminação não intencional, de modo que a maior parte das leis também se aplica ao uso de IA. O uso de IA em circunstâncias em que algumas formas de discriminação não são explicitamente proibidas pode exigir mudanças legais e códigos de conduta voluntários adaptados ao uso de IA (BORGESIUS, 2018). Com efeito, alguns códigos e diretrizes no campo da inteligência artificial, como os Princípios de Asilomar, já colocaram em foco a questão da não-discriminação. Como a inteligência artificial pode ter um grande impacto sobre a vida e sobre a propriedade, processos aptos a prevenir o viés e identificar problemas éticos tempestivamente desempenham um papel essencial. A mitigação de

preconceitos, portanto, endereça o risco de causar ou contribuir para abusos graves de direitos humanos como uma questão de *compliance* jurídico.[10]

A consequência para as empresas é que, no que tange à qualidade dos produtos de IA, as questões relacionadas aos consumidores e os direitos humanos precisam ser compreendidos de maneira interconectada; daí resulta que a simples probabilidade de que as decisões de IA sejam melhores do que as humanas não é suficiente para a implementação de IA compatível com os direitos humanos. Do ponto de vista regulatório, as soluções que tratam da aplicação de padrões mínimos precisam ser compatíveis com a estrutura de incentivos existente, já que os apelos morais nem sempre são fortes o suficiente para dissuadir as empresas de violar as normas. O verdadeiro problema por trás do uso de IA, portanto, pode ser a ausência de padrões de qualidade ou a negligência por parte do produtor. Do ponto de vista da ética empresarial, o precedente relevante seria o caso Ford Pinto da década de 1970. Nesse caso particular, a Ford deu prioridade à maximização de lucros sobre a segurança do produto e evitou investimentos adicionais em segurança, aceitando uma taxa de mortalidade mais elevada (STROBEL, 1980).

Para além dos padrões de qualidade relativos aos impactos sobre os direitos humanos e das estruturas contra a discriminação, a transparência desempenha papel importante para evitar que o *output* conduza a violações relevância de "processos transparentes de devida diligência em direitos humanos, que envolvam a identificação dos riscos aos direitos humanos associados aos seus sistemas de IA e a adoção de medidas eficazes para prevenir e/ou mitigar os danos causados por tais sistemas" (CONSELHO DA EUROPA, 2019). Esse raciocínio se estende à atuação de empresas privadas, que tratam de serviços de interesse relevante para as partes envolvidas, como, por exemplo, no setor de saúde. Mecanismos para fornecer mais transparência podem incluir relatórios de sustentabilidade, que já tratam dos impactos sobre os direitos humanos. A título de exemplo, as empresas podem relatar os critérios usados nos algoritmos dos mecanismos de candidatura a empregos e descrever os processos usados para reduzir os riscos de viés.

Transparência sobre a abordagem da gestão da IA, explicações quanto às soluções de IA e seus impactos potenciais sobre os direitos humanos, bem como os mecanismos de

[10] Veja-se os Princípios Orientadores da ONU sobre Negócios e Direitos Humanos.
Nota da tradutora: os princípios em questão estabelecem, em seu item 23 (c), que empresas devem tratar como uma questão de *compliance* jurídico o risco de causar ou contribuir para graves abusos de direitos humanos, onde quer que estejam operando.

remediação, desempenham papeis centrais para garantir que a IA esteja em conformidade com os direitos humanos. A intensidade dos mecanismos de *enforcement* desses princípios, na forma de regras de licenciamento e de auditoria, deve depender do impacto da solução de IA e de sua relevância para os direitos humanos.

C. SITUAÇÃO III: O USO DE IA EM ÁREAS ESPECÍFICAS CONFLITA COM OS DIREITOS HUMANOS

Nas seções anteriores, discutimos os impactos da IA sobre os direitos humanos a partir de uma perspectiva de pretensões *(claims)*. No entanto, uma perspectiva holística sobre o tema requer a inclusão de direitos que possibilitem a participação no processo de tomada de decisões políticas e sociais. A participação política do cidadão se expressa classicamente no direito de voto e no direito de expressar sua opinião. Ambos os conceitos podem estar ligados à noção de "participação democrática", que foi prevista como um princípio orientador da regulação de IA pela Declaração de Montreal para o Desenvolvimento Responsável da Inteligência Artificial (ARTIFICIAL INTELLIGENCE CLUSTER STEERING COMMITTEE QUEBEC, 2019), e ao "princípio da autonomia", previsto nos Princípios de Asilomar sobre IA. A fim de ilustrar a importância de áreas específicas para a discussão sobre violações de direitos humanos, particularmente no que tange a direitos de participação, será dado foco ao comportamento das empresas de IA no que se refere à liberdade de expressão, por se tratar da área mais contestada da ética em IA. Sua relevância foi destacada pelo Relatório da ONU "Promoção e proteção do direito à liberdade de opinião e expressão", que abordou o impacto da IA sobre a liberdade de expressão e opinião e destacou o papel da Internet como plataforma para formar e articular opiniões (KAYE, 2017).

Em geral, a influência da IA sobre a liberdade de opinião e expressão parece ter duas implicações para os tomadores de decisões corporativas. Por um lado, os vieses criados pela IA podem impactar a autodeterminação e a autonomia dos indivíduos para formar e desenvolver opiniões pessoais com base em informações factuais e variadas (BALKIN, 2018; BRKAN, 2019). Os impactos da IA podem ser particularmente fortes, se considerarmos o papel importante que as teorias do discurso atribuem ao pressuposto de que nenhum argumento relevante deve ser suprimido ou excluído pelos participantes (HABERMAS, 1992). Como resultado, a filtragem pela IA poderia alterar o rumo do discurso e suprimir partes do espectro

de opinião. A Declaração de Montreal argumentou, portanto, que salvaguardar a "democracia contra a manipulação de informações para fins políticos" constitui um dos maiores desafios éticos para os tomadores de decisão, o que implica que as empresas precisam "prevenir e mitigar" os efeitos negativos sobre o discurso por meio de avaliações de risco e da adoção de padrões de qualidade[11].

Por outro lado, o uso de IA para censurar comentários políticos específicos permanece altamente problemático. O debate atual gira em torno da questão de saber se a IA deve limitar o discurso de ódio (BBC, 2018). A primeira questão diz respeito ao aspecto da viabilidade técnica. A IA ainda não parece ser capaz de distinguir entre comentários apropriados e discurso de ódio. As linhas que separam expressões de opinião "ainda permitidas" e comentários de ódio podem ser difíceis de distinguir e dependentes do contexto: as soluções de IA requerem uma compreensão da ironia e de expressões específicas de determinado ambiente e cultura, tornando a padronização de decisões de casos individuais uma tarefa muito árida. A segunda questão é de natureza mais teórica. A transferência da censura de comentários para as empresas pode ser questionável não apenas em termos de liberdade de opinião, mas também quando se considera o princípio do Estado de Direito, uma vez que decisões anteriormente tomadas pelo Poder Judiciário são substituídas por algoritmos empregados por empresas privadas (LA RUE, 2011). Nessas circunstâncias, torna-se, portanto, imprescindível que as empresas e, especialmente, as redes sociais estabeleçam mecanismos de reparação que permitam aos indivíduos se protegerem em casos de injustiça. Ademais, o Relatório da ONU sobre a promoção e proteção do direito à liberdade de opinião e expressão referiu-se à responsabilidade das empresas de "publicar dados sobre remoções de conteúdo, [...] juntamente com estudos de caso e educação sobre formação de perfis comerciais e políticos" (KAYE, 2017). Devido ao fato de que o uso excessivo e a subutilização (FLORIDI et. al., 2018) de mecanismos de censura de comentários nas mídias sociais podem ser igualmente prejudiciais aos direitos humanos, as empresas podem ser obrigadas a enfatizar aspectos procedimentais, como a transparência perante terceiros, avaliações de risco e padrões operacionais, em maior extensão do que em outros casos de IA.

D. SITUAÇÃO IV: UM VIOLADOR DOS DIREITOS HUMANOS USA IA

[11] Vide os Princípios Orientadores da ONU *(United Nations Guiding Principles on Business and Human Rights)*.

Na seção final, lidamos com o caso de uso mais problemático, ou seja, o uso de IA para infringir direitos humanos. Esta forma de uso de IA é normalmente chamada de "IA maliciosa" e a regulamentação de IA tem dedicado grande atenção a esse aspecto, que pode até mesmo ser considerado o seu princípio mais antigo[12]. A diferença com relação aos casos anteriormente citados está na intenção de usar a IA como instrumento de violação de direitos humanos. Nesse sentido, a violação dos direitos humanos não é uma externalidade ou um ato de negligência, mas sim o objetivo da solução de IA.

Possíveis cenários desse tipo incluiriam o uso de algoritmos para discriminar membros de sindicatos em processos automatizados de remuneração e promoção, ou programas que - deliberadamente - favorecem certos grupos étnicos na distribuição de serviços sociais. Ambos os casos constituem atos de discriminação e retratam violações graves do direito ao tratamento igual e justo. Embora o Poder Legislativo possa estabelecer disposições legais e mecanismos de triagem para prevenir violações de direitos humanos sob sua jurisdição, violações de direitos humanos planejadas fora de seu escopo são muito mais difíceis de prevenir.

Desenvolvimentos recentes em alguns países geram preocupações de que a combinação de IA com *big data* possa fortalecer os mecanismos de vigilância de Estados que vivem à margem das leis internacionais, ou organizações terroristas . Um exemplo que ressalta a relevância prática da não maleficência é a repressão do governo chinês às minorias étnicas. Recentemente, a vigilância governamental se expandiu e foi combinada com o uso de IA para acessar bancos de dados de material biológico e genético, com vínculo com sistemas de reconhecimento facial (UHRP, 2018). As medidas do governo chinês também levantam questões relativas à legitimidade, uma vez que alguns desses aplicativos parecem ser usados em conexão com medidas de repressão a parcelas da oposição chinesa ou a minorias nacionais (KUO, 2019). De acordo com um relatório do New York Times em 2019 (MOZUR, 2019), *startups* chinesas criaram algoritmos para monitorar muçulmanos étnicos na província chinesa de Xinjiang, o que gerou críticas de várias ONGs e também de governos (MOGHERINI, 2018) e de organizações internacionais (ONU, 2018). A existência de sistemas de câmeras com o objetivo de aumentar o controle sobre os cidadãos, combinados com campos de detenção e ferramentas de reconhecimento facial, reforçam a urgência do assunto (ZENZ, 2019). Os

[12] Veja-se a Lei de Asimov: "Um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano sofra algum dano".

eventos que ocorrem em Xinjiang constituem uma violação dos direitos humanos, uma vez que as intervenções do Estado entram em conflito com as considerações de proporcionalidade e com o uso legítimo do poder do Estado. O fato de os membros de minorias étnicas serem geralmente colocados sob suspeita, as violações ao direito à integridade física e à igualdade de tratamento (como o artigo 18/19/21/29 da Lei de Contraterrorismo da República Popular da China, ou o "Regulamento sobre Desradicalização") e os pouco exigentes requisitos legais para a detenção de indivíduos parecem confirmar esta opinião (THUM, 2018).

Esses casos naturalmente geram repercussões de governança para empresas que fornecem tecnologias para regimes repressivos, porque a IA se qualifica como uma tecnologia de uso dual.

A natureza de uso dual se aplica explicitamente a *softwares* de reconhecimento facial e de voz, à coleção de dados biológicos e aos bancos de dados de policiamento preditivo. De acordo com Princípios Orientadores da ONU sobre Negócios e Direitos Humanos, as empresas precisam, portanto, considerar como essas tecnologias serão utilizadas pelo usuário final e se precisam estabelecer medidas de devida diligência específicas para o país, de modo a evitar casos de uso indevido. Isso também é importante do ponto de vista de *compliance*, uma vez que as violações de direitos humanos entranhadas na cadeia logística mais ampla de empresas já podem recair sob o escopo de leis com efeitos extraterritoriais, incluindo a Lei de Escravidão Moderna, do Reino Unido, ou a Lei Magnitzky, dos EUA.

Apesar disso, o caráter de uso dual da IA pode ser útil para que as empresas detectem violações dos direitos humanos. As tecnologias modernas têm apoiado decisivamente o trabalho de jornalistas, ativistas e acadêmicos em suas pesquisas sobre os eventos em andamento em Xinjiang e em outros lugares (THUM, 2018; ZENZ, 2019). Um caso concreto em que a IA teve um impacto positivo sobre os direitos humanos inclui o uso de *blockchain* na cadeia de suprimento de cobalto, para ajudar empresas e governos a avaliar se o material está envolvido em questões de direitos humanos (SULKOWSKI, 2019). Os relatórios também destacaram o papel da IA na detecção e no combate ao crime financeiro e lavagem de dinheiro, bem como na verificação da existência de sanções (ZIMILES e MUELLER, 2019).

Embora o uso de algoritmos e de tecnologias de *blockchain* possa ser central para rastrear violações de direitos humanos no futuro e aumentar a pressão sobre as empresas para cumprir os padrões de direitos humanos ao operarem no exterior, o uso dessas tecnologias pode

atingir certas limitações concretas. Os países com maior tamanho econômico e peso político são mais propensos a resistir aos mecanismos de sanção que lhes são impostos e têm mais capacidade para contornar as sanções impostas individualmente. A prevenção efetiva de tais ações é, portanto, não apenas uma questão de legislação específica da IA, mas também uma questão de política externa e de aplicação das normas internacionais de direitos humanos.

VII. CONCLUSÃO

Neste artigo, examinamos as responsabilidades dos atores corporativos para a aplicação e concretização dos padrões de direitos humanos no contexto da inteligência artificial. Em geral, descobrimos que o uso de IA - mesmo em decisões de vida ou morte - não entra, em princípio, em conflito com o princípio da autodeterminação moral. No entanto, o tratamento da IA precisa ser baseado em considerações mais amplas de direitos humanos. A aplicação geral da IA em situações irreversíveis e de escolhas difíceis, por exemplo, precisa obedecer a regras universais e abstratas, que podem evoluir a partir de um discurso aberto e democrático.

Em uma etapa posterior, delineamos os desafios para as empresas em relação às condutas de direitos humanos impostas pela inteligência artificial. De acordo com o critério de beneficência, que decorre da ética de IA, a IA tem como objetivo beneficiar os seres humanos e atingir o bem comum. Esta ideia pode estar ligada à realização progressiva dos direitos humanos, em linha com o Pacto Internacional sobre Direitos Econômicos Sociais e Culturais, bem como com a estrutura dos Objetivos de Desenvolvimento Social da ONU. Encontramos muitos casos que mostram que o uso de IA pode levar à realização dos direitos humanos, especialmente quando se trata de saúde, combate à pobreza e educação, e que as empresas podem desempenhar um papel positivo no oferecimento de soluções de IA que abordem questões relevantes de direitos humanos.

Em outros contextos, em que as ações das empresas podem entrar em conflito com os direitos humanos, torna-se necessário o estabelecimento de mecanismos de governança quanto aos casos de uso de IA, com vistas a prevenir as violações dos direitos humanos. No caso do *input* de dados, a aplicação de IA geralmente não é problemática, desde que todas as partes envolvidas consintam com as condições de seu uso e desde que o uso esteja em conformidade com os padrões éticos e legais. O princípio do consentimento, entretanto, pode ser questionado, dada a assimetria de poder entre a empresa e o consumidor. Isso se aplica especificamente à

interação entre empresas e autoridades públicas. Outra área de potenciais violações dos direitos humanos se encontra nos vieses que levam à discriminação ilícita. Aqui, as empresas precisam aderir a padrões de qualidade e elucidar, com transparência, a tomada de decisões corporativas e os riscos de violações dos direitos humanos que resultam do uso de IA. Além disso, o uso de IA é questionável em alguns casos e ambientes, especialmente quando se trata da participação dos cidadãos ou da transferência de poderes de decisão com consequências impactantes. Para prevenir as violações ao direito de opinião, as empresas precisam desenvolver medidas preventivas e construir uma infraestrutura interna voltada a remediar as violações dos direitos humanos. Por fim, a última questão relacionada à IA e aos direitos humanos diz respeito aos atores que usam a IA para violar os direitos humanos. Em contraste com os mecanismos anteriores de *enforcement*, as vantagens da IA são que os seus resultados são mais confiáveis e mais rápidos. Contudo, o *enforcement* de direitos humanos face a atores estatais representa um grande desafio a empresas que atuam internacionalmente, uma vez que o direito internacional tem progressivamente enfatizado as responsabilidades corporativas.

Ao avaliar esses aspectos da discussão sobre IA e direitos humanos, constatamos que os tomadores de decisões corporativas são confrontados com desafios complexos, que incluem a transparência, problemas de qualidade de dados e o gerenciamento da cadeia logística. O *input* e o *output* da IA, os usuários e a forma como as tecnologias são utilizadas ensejam possibilidades regulatórias completamente distintas e, portanto, requerem medidas adequadas à situação.

REFERÊNCIAS

ACEMOGLU, Daron; RESTREPO, Pascual. Artificial Intelligence, Automation and Work. **NBER Working Paper** No. 24196 (2018).

AMBROSI, Gerhard M. **Aristotle's Geometrical Model of Distributive Justice** (2007). Disponível em: https://www.uni-trier.de/ fileadmin / fb4 / prof / VWL / EWP / Publikationen / Ambrosi / Aristotle-4.pdf . Acesso em: 27 de novembro de 2019.

ANGWIN et al, Julia. 'Machine Bias', **ProPublica** (23 de maio de 2016). Disponível em: https://www.propublica.org/article/machine-bias- avaliações de risco em condenações criminais . Acesso em: 27 de novembro de 2019.

ARTIFICIAL INTELLIGENCE CLUSTER STEERING COMMITTEE QUEBEC. **Montréal Declaration for Responsible Development of Artificial Intelligence.** Disponível em: https://forumia.quebec/wp-content/uploads/sites/2/2018/12/News-release_Launch_Montreal_Declaration_AI-04_12_18.pdf. Acesso em: 27 de novembro de 2019.

AWAD, Edmond *et al.* The Moral Machine Experiment (2018) 563 **Nature**, 59.

BAAI - Beijing Academy of Artificial Intelligence. **Beijing AI Principles** (28 de maio de 2019). Disponível em: https://www.baai.ac.cn/news/beijing-ai-principles-en.html . Acesso em: 27 de novembro de 2019.

BALKIN, Jack M. 'Free Speech in the Algorithmic Society: Big Data, Private Governance, and New School Speech Regulation', **Yale Law School, Public Law Research Paper No. 615** (2018).

BANCO MUNDIAL. **Human Rights and Economics: Tensions and Positive Relationships** (2012). Disponível em: http://siteresources.worldbank.org/PROJECTS/Resources/40940-1331068268558/Report_Development_Fragility_Human_ Rights.pdf . Acesso em 27 de novembro de 2019.

BBC. 'Germany Starts Enforcing Hate Speech Law' (1 de janeiro de 2018). **BBC.** Disponível em: https://www.bbc.com/news/technology- 42510868 . Acesso em: 30 de agosto de 2019.

BENNINTON-CASTRO, Joseph. AI is a Game-Changer in the Fight Against Hunger and Poverty. Here's Why'. (12 de junho de 2017). **NBC News**. Disponível em: https://www.nbcnews.com/mach/tech/ai-game-changer-fight-against-hunger-poverty-here-s-why- ncna774696 . Acesso em: 27 de novembro de 2019.

BERLIN, Isaiah. 'Two Concepts of Liberty'. In: **Four Essays on Liberty**. Oxford: Oxford University Press, 1969.

BORGESIUS, Frederik Z. **Discrimination, Artificial Intelligence, and Algorithmic Decision-Making**. Conselho da Europa: Estrasburgo (2018). Disponível em: https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73 . Acesso em: 27 de novembro de 2019.

BRANNON, Jason; JAWORSKI, Peter M. **Markets without Limits: Moral Virtues and Commercial Interests**. Abingdon, UK: Routledge, 2015.

BRKAN, Maja. **Freedom of Expression and Artificial Intelligence: On Personalization, Disinformation and (Lack of) Horizontal Effect of the Charter** (17 de março de 2019). Disponível em: http://dx.doi.org/10.2139/ssrn.3354180 . Acesso em: 27 de novembro de 2019.

CELLAN-JONES, Rory. Stephen Hawking Warns Artificial Intelligence Could End Mankind'(2 December 2014). **BBC**. Disponível em: https:// www.bbc.com/news/technology-30290540 . Acesso em: 27 November 2019).

CONSELHO DA EUROPA. **Unboxing Artificial Intelligence: 10 steps to protect Human Rights.** Disponível em: https://rm.coe.int/ unboxing-artificial-intelligence-10-steps-to-protect-human-rights-reco / 1680946e64 p. 18 . Acesso em: 27 de novembro de 2019.

COWLS, Josh et al. **Designing AI for Social Good: Seven Essential Factors** (15 May 2019). Disponível em: https://dx.doi.org/10.2139/ ssrn.3388669 . Acesso em: 27 de novembro de 2019.

COWLS, Josh; FLORIDI, Luciano. **Prolegomena to a White Paper on an Ethical Framework for a Good AI Society** (2019). Disponível em: https://dx.doi.org/10.2139/ssrn.3198732 . Acesso em: 27 de novembro de 2019.

CRAIG, Paul; DE BRUCA, Grainne. **EU Law - Text, Cases and Materials**. Oxford, UK: Oxford University Press, 2015.

DATEN ETHIK KOMMISSION. **Opinion of the Data Ethics Commission** (outubro de 2019). Disponível em: https://www.bmjv.de/SharedDocs/Downloads/DE/Themen/Fokusthemen/Gutachten_DEK_EN.pdf?__blob=publicationFile&v=2. Acesso em: 27 de novembro de 2019.

DI NUCCI, Ezio, Self-Sacrifice and the Trolley Problem. (2013) 26 **Philosophical Psychology** 5, 662.

ERLC - The Ethics and Religious Liberty Commission of the Southern Baptist Convention. **Artificial Intelligence: An Evangelical Statement of Principles** (2019), https://erlc.com/resource-library/statements/artificial-intelligence-an-evangelical-statement-of-principles/ (acessado em 27 de novembro de 2019).

ETZIONI, Amitai 'Are New Technologies the Enemy of Privacy?' (2007) 20 **Knowledge, Technology, and Policy** 115.

FLORIDI, Luciano *et al*. AI4People – An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations (2018). 28 **Minds & Machines** 4, 689

FUTURE OF LIFE INSTITUTE. **Asilomar AI Principles** (2017). Disponível em: https://futureoflife.org/ai-principles/ . Acesso em: 27 de novembro de 2019.

GARDBAUM, Stephen. Human Rights as International Constitutional Rights (2008). 19 **European Journal of International Law** 4, 749.

DEUTSCHER BUNDESTAG. Documento 19/5544 do parlamento alemão (novembro de 2018). Disponível em: https://dserver.bundestag.de/btd/19/055/1905544.pdf. Acesso em: novembro de 2021.

GOODALL, Noah J, 'Machine Ethics and Automated Vehicles', In: Gereon Meyer and Sven Beiker (eds.), **Road Vehicle Automation** (Springer, 2014), 93

HABERMAS, Juergen, **The Theory of Communicative Action**. Boston, MA: Beacon Press, 1992.

HARBO, Tor-Inge, 'The Function of the Proportionality Principle in EU Law' (2010) 16 **European Law Journal** 2, 158;

HLEG - High Level Expert Group on Artificial Intelligence, '**Ethics Guidelines for Trustworthy AI**' (8 April 2019). Disponível em: https:// ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419 . Acesso em: 27 de novembro de 2019.

HOHFELD, Wesley. Fundamental Legal Conceptions as Applied in Judicial Reasoning (1917). 26 **Yale Law Journal** 8, 710

HUNT, Elle. 'Tay, Microsoft's AI Chatbot, Gets a Crash Course in Racism from Twitter', **The Guardian** (24 de março de 2016), https://www.theguardian.com/technology/2016/mar/24/tay-microsofts-ai-chatbot-gets-a-crash-course-in-racism- do twitter (acessado em 27 de novembro de 2019).

JÄGERS, Nicola, **Corporate Human Rights Obligations** (Cambridge: Intersentia, 2002).

JOHNSTON, Ed. 'Brain Scanning and Lie Detectors: The Implications for Fundamental Defense Rights' (2016) 22 **European Journal of Current Legal Issues** 2

KANT, Immanuel **Groundwork of the Metaphysics of Morals**. Cambridge: Cambridge University Press, 1998.

KAYE, David, '**Mandate of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression', Open Letter to Office of the High Commissioner for Human Rights** (1 June 2017). Disponível em: <https://www.ohchr.org/Documents/Issues/Opinion/Legislation/OL-DEU-1-2017.pdf> . Acesso em: 27 de novembro de 2019.

KUO, Lily, '"If You Enter a Camp, You Never Come Out": Inside China's War on Islam', **The Guardian** (11 de janeiro de 2019). Disponível em: <https://www.theguardian.com/world/2019/jan/11/if-you-enter-a-camp-you-never-come-out-inside-chinas-war-on-islam>. Acesso em: 27 de novembro de 2019.

KUSZLER, Patricia, 'Global Health and the Human Rights Imperative' (2007) 2 **Asian Journal of WTO & International Health Law and Policy** 1, 99.

LA RUE, Frank. **Relatório do Relator Especial sobre a Promoção e Proteção do Direito à Liberdade de Opinião e Expressão**. Conselho de Direitos Humanos da Assembleia Geral das Nações Unidas (16 de maio de 2011). Disponível em: <https://www2.ohchr.org/english/bodies/hrcouncil/docs/17session/A.HRC.17.27_en.pdf> . Acesso em 12 de setembro de 2019.

LESSIG, Lawrence, 'Privacy as Property' (2002). 69 **Social Research** 1, 247.

MCCARTHY, John. 'Programs with Common Sense', In: **Proceedings of the Teddington Conference on the Mechanization of Thought Processes** (1959), 75.

MILL, John S, **On Liberty.** Londres: Longman, Roberts & Green, 1859.

MOGHERINI, Federica, **Speech by HR/VP Mogherini at the Plenary Session of the European Parliament on the State of the EU–China Relations** (11 September 2018). Disponível em: <https://eeas.europa.eu/headquarters/headquarters-homepage/50337/ speech-hrvp-mogherini-plenary-session-european-parliament-state-eu-china-relations_en>. Acesso em: 26 de novembro de 2019.

MOZUR, Paul, 'One Month, 500,000 Face Scans: How China Is Using AI to Profile a Minority', **New York Times** (14 de abril de 2019). Disponível em: <https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial- >. Acesso em: 27 de novembro de 2019.

MUCHLINSKY, Peter, 'Implementing the New UN Corporate Human Rights Framework. Implications for Corporate Law, Governance, and Regulation' (2002) 22 **Business Ethics Quarterly** 1, 145–177.

MUELLER, Julian F, 'The Ethics of Commercial Human Smuggling' (2018) **European Journal of Political Theory,** 1–19.

NOZICK, Robert. Anarchy, State and Utopia (1974).

OBERMEYER, Ziad et al. Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations (2019). 366 **Science**, 447–453.

ONU - ORGANIZAÇÃO DAS NAÇÕES UNIDAS. Escritório do Alto Comissariado das Nações Unidas para os Direitos Humanos. **Pacto Internacional sobre Direitos Econômicos, Sociais e Culturais** (1966). Disponível em: <https://www.ohchr.org/en/professionalinterest/pages/cescr.aspx>. Acesso em: 27 de novembro de 2019.

_____. Comitê contra a Tortura. **'Concluding Observations on the Fourth Periodic Report of the United Kingdom',** UN Doc. CAT/C/CR/33/3 (25 November 2004), para 4(a)(i).

_____. Conselho de Direitos Humanos. **Report of the Special Representative of the Secretary-General on the issue of human rights and transnational corporations and other business enterprises, John Ruggie** (2008a). Disponível em: <https: //www.business-humanrights.org/sites/default/files/reports-and-materials/Ruggie-report-7-Apr-2008.pdf> . Acesso em: 27 de novembro de 2019.

_____. Escritório do Alto Comissariado das Nações Unidas para os Direitos Humanos, **The Right to Health. Fact Sheet No. 31** (2008b). Disponível em: <https://www.ohchr.org/ Documentos / Publicações / Factsheet31.pdf> . Acesso em: 15 de novembro de 2019.

_____. Escritório do Alto Comissariado das Nações Unidas para os Direitos Humanos**. United Nations Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework** (2011). Disponível em: <https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf>. Acesso em: 27 de novembro de 2019.

_____. **Agenda 2030 para o Desenvolvimento Sustentável** (2015). Disponível em: <https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E>. Acesso em: 12 de novembro de 2019.

_____. 'Task of eradicating poverty must be met 'with a sense of urgency', says deputy UN chief. **UN News Centre** (8 de maio de 2017). Disponível em: <https://www.un.org/sustainabledevelopment/blog/2017/05/task-of-eradicating-poverty-must-be-met-with-a-sense-of-urgency-says-deputy-un-chief/>. Acesso em: 2 de setembro de 2019.

_____. Escritório do Alto Comissariado das Nações Unidas para os Direitos Humanos. **Committee on the Elimination of Racial Discrimination reviews the report of China** (13 de agosto de 2018). Disponível em: <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=23452&LangID=E>. Acesso em: 27 de novembro de 2019.

PAGALLO et al. **On Good AI Governance: 14 Priority Actions, a SMART Model of Governance, and a Regulatory Toolbox** (2019). Disponível em: <https://www.eismd.eu/wp-content/uploads/2019/11/AI4Peoples-Report-on-Good-AI-Governance_compressed.pdf>. Acesso em: 27 de novembro de 2019.

POJMAN, Louis P. **Ethical Theory: Classical and Contemporary Readings**. Belmont, CA: Wadsworth Publishing, 1996.

RAWLS, John. **A Theory of Justice**. Cambridge: Belknap Press of Harvard University Press, 1971.

REIGADA, Antonio T. The Principle of Proportionality and the Fundamental Right to Personal Data Protection: The Biometric Data Processing (2012) 17 **Lex Electronica** 2.

RUGGIE, John Gerard. **Just Business: Multinational Corporations and Human Rights**. New York: WW Norton & Company, 2013.

RUSSELL, Stuart et al. Research Priorities for Robust and Beneficial Artificial Intelligence. (2015) 36 **Artificial Intelligence Magazine** 4, 105.

SANTORO, Michael A. **Profits and Principles: Global Capitalism and Human Rights in China.** Ithaca: Cornell University Press, 2000.

SIMMA, Bruno; PULKOWSKI, Dirk, 'Of Planets and the Universe: Self-Contained Regimes' (2006) 17 **European Journal of International Law** 3, 483.

SIMONCELLI, Tania; WALLACE, Helen, 'Expanding Databases, Declining Liberties' (2006) 19 **Genewatch: A Bulletin of the Committee for Responsible Genetics** 1, 3.

SPARKES, Matthey, 'Top Scientists Call for Caution over Artificial Intelligence', **The Telegraph** (13 January 2015). Disponível em: <https://www.telegraph.co.uk/technology/news/11342200/Top-scientists-call-for-caution-oer-artificial-intelligence.html>. Acesso em 27 de novembro de 2019.

STELTZER, William, 'Population Statistics, the Holocaust, and the Nuremberg Trials' (1998) 24 **Population and Development Review** 3, 511–552.

STROBEL, Lee P, **Reckless Homicide? Ford's Pinto Trial.** South Bend and Books, 1980.

SULKOWSKI, Adam J. 'Blockchain, Business Supply Chains, Sustainability, and Law: The Future of Governance, Legal Frameworks, and Lawyers?' (2019) **43 Delaware Journal of Corporate Law** 2, 303–345.

THUM, Rian 'China's Mass Internment Camps Have No Clear End in Sight', **Foreign Policy** (22 de agosto de 2018), Disponível em: <https://foreignpolicy.com/2018/08/22/chinas-mass-internment-camps-have-no-clear-end-in-sight/>. Acesso em: 27 de novembro de 2019.

TRIBUNAL FEDERAL DE JUSTIÇA DA ALEMANHA. Decisão 1954 BGH, 16.02.1954-1 StR 578/53

_____. Decisão 2003 BGH, v. 24.06.2003-VI ZR 327/02.

UHRP - Uyghur Human Rights Project. **China's Repression and Internment of Uyghurs: US Policy Responses**. Comitê de Relações Exteriores da Câmara - Subcomitê da Ásia e do Pacífico (26 de setembro de 2018). Disponível em: < https://docs.house.gov/ reuniões / FA / FA05 / 20180926/108718 / HHRG-115-FA05-Wstate-TurkelN-20180926.pdf> . Acesso em 27 de novembro de 2019.

UNIVERSAL RIGHTS GROUP. **Human Rights and the SDGs. Pursuing Synergies** (Dezembro de 2017). Disponível em: <https://www.universal-rights.org/wp-content/uploads/2017/12/RAPPORT_2017_HUMAN-RIGHTS-SDGS-PURSUING-SYNERGIES_03_12_2017_digital_use-2.pdf>. Acesso em: 12 de novembro de 2017.

VINUESA, Ricardo et al, 'The Role of Artificial Intelligence in Achieving the Sustainable Development Goals', **arXiv** (2019). Disponível em: <https://arxiv.org/ftp/arxiv/papers/1905/1905.00501.pdf> . Acesso em: 12 de setembro de 2019.

WALLACE, H. M. *et al*, 'Forensic DNA Databases - Ethical and Legal Standards: A Global Review' (2014) 4 **Egyptian Journal of Forensic Science** 3, 57.

WALLACE, Helen, 'The UK National DNA Database: Balancing Crime Detection, Human Rights and Privacy' (2006) 7 **EMBO Reports** 26;

WALLACH, Wendell; ALLEN, Colin**. Moral Machines: Teaching Robots Right From Wrong**. Oxford: Oxford University Press, 2008.

WEBER, Max. **Politik als Beruf', in: Geistige Arbeit als Beruf. Vier Vorträge vor dem Freistudentischen Bund**. Munique, Alemanha: Duncker & Humblot, 1919.

WETTSTEIN, Florian. **Multinational Corporations and Global Justice**. Bibliovault OAI Repository, the University of Chicago Press, 2009.

WOGU, I. A. P. et al. "Artificial intelligence, alienation and ontological problems of other minds: A critical investigation into the future of man and machines," **2017 International Conference on Computing Networking and Informatics (ICCNI),** 2017, pp. 1-10, doi: 10.1109/ICCNI.2017.8123792

ZENZ, Adrian, 'Thoroughly Reforming Them Towards a Healthy Heart Attitude: China's Political Re-Education Campaign in Xinjiang' (2019) 38 **Central Asian Survey** 1, 102.

ZERK, J. A. **Multinationals and Corporate Social Responsibility. Limitations and Opportunities in International Law**. Cambridge: Cambridge University Press, 2006.

ZIMILES, Ellen; MUELLER, Tim, 'How AI Is Transforming the Fight against Money Laundering', **Fórum Econômico Mundial** (17 de janeiro de 2019).

Disponível em: https://www.weforum.org/agenda/2019/01/how-ai-can-knock-the-starch-out-of-money- lavagem / . Acesso em: 27 de novembro de 2019.

ZUCCA, Lorenzo. **Constitutional Dilemmas: Conflicts of Fundamental Legal Rights in Europe and the USA**. Oxford, UK: Oxford University Press, 2007.

Sobre os autores:

Alexander Kriebitz
Pesquisador Associado da Cátedra de Ética Empresarial da Universidade Técnica de Munique, Munique, Alemanha.

Christoph Lütge/ *E-mail*: luetge(at)tum.de
Chair Holder e professor titular da Cátedra de Ética Empresarial, Universidade Técnica de Munique, Munique, Alemanha. Como professor da Universidade Técnica de Munique e diretor do Instituto de Ética em Inteligência Artificial, Christoph Lütge recebeu financiamento para projetos do Facebook Inc., Fujitsu K.K. e Huawei Technologies Co. Ltd, relacionados à pesquisa em inteligência artificial e ética.

Sobre a Tradutora:
Miriam Wimmer/ E-mail:miriam.wimmer@yahoo.com.br
Doutora em Comunicação pela UnB, mestre em Direito Público pela UERJ e professora do IDP.

Artigo Convidado.