

PRINCÍPIOS INTERNACIONAIS DE INTELIGÊNCIA ARTIFICIAL

Artificial intelligence international principles

Carlos Bruno Ferreira da Silva

<http://lattes.cnpq.br/1254024052318025>

Doutor em Direito Constitucional pela Universidad de Sevilla – Espanha.
Orientador Pedagógico da Escola Superior do Ministério Público
da União (ESMPU). Procurador da República (MPF).

RESUMO

Este artigo apresenta e analisa os princípios internacionais para o desenvolvimento e uso da inteligência artificial (IA), a partir da análise de diplomas internacionais sobre o assunto da Organização para a Cooperação e Desenvolvimento Econômico (OCDE) e do Conselho de Europa. A discussão se concentra em seis pontos principais: desenvolvimento humano, direitos humanos e democracia, transparência, segurança, responsabilização e cooperação. Inicialmente, é enfatizado que a inteligência artificial só tem sentido de existir se contribuir para a diminuição de desigualdades econômicas, respeitando o Estado de direito, nossos direitos reconhecidos e o regime de escolha popular. O artigo explora a importância fundamental da transparência e da explicabilidade para as pessoas afetadas pela IA, principalmente aquelas que interagem com ela voluntariamente ou que atuam em virtude de relações jurídicas. A robustez, integridade e segurança da IA são tratadas sob a perspectiva da necessidade de confiabilidade do sistema diante das adversidades apresentadas pelo contexto sociotécnico. Por fim, se estabelece a importância de meios para a responsabilização de ilícitos e da cooperação ampla de entes públicos e privados para a consecução de todos esses objetivos. Em suma, o documento discute a relevância já reconhecida no cenário internacional de princípios éticos e legais para o uso da IA, a fim de que a tecnologia seja uma ferramenta para a promoção do bem comum e não uma ameaça aos indivíduos.

PALAVRAS-CHAVE

Inteligência artificial; democracia; direitos humanos; transparência; responsabilidade.

ABSTRACT

This article outlines and analyses international principles for the development and use of artificial intelligence (AI), based on an analysis of international legislation on the subject from the Organisation for the Economic Co-operation and Development (OECD) and the Council of Europe. The discussion focuses on six main points: human development, human rights and democracy, transparency, safety, accountability and cooperation. Initially, it is emphasized that artificial intelligence only makes sense to exist if it contributes to reducing economic inequalities and respects the rule of law, our recognized rights and the regime of popular decision. The article explores the fundamental importance of transparency and explainability for those affected by AI, especially those who interact with it voluntarily or who act in a legal relationship. The robustness, integrity and safety of AI are addressed from the perspective

of the system's need for reliability in the face of the challenges presented by the sociotechnical context. Finally, it establishes the importance of means of holding wrongdoers accountable and the broad cooperation of public and private entities to achieve all these objectives. In short, the document discusses the relevance already recognized on the international arena of ethical and legal principles for the use of AI, so that technology can be a tool for promoting the common good and not a threat to individuals.

KEYWORDS

Artificial intelligence; democracy; human rights; transparency; accountability.

Recebido em: 24.07.24 – Aprovado em: 11.09.24

INTRODUÇÃO

O vertiginoso desenvolvimento da tecnologia da inteligência artificial vem sendo acompanhado por um intenso esforço dos legisladores nacionais em regulá-la. A busca dos parlamentos é conseguir normatizar, sem ainda ter a exata noção de todas as potencialidades, permitindo o desenvolvimento da ciência, mas já evitando perigos da subjugação humana pela máquina.

A esperança envolve acreditar que o avanço dos cérebros computadorizados permitirá que superemos grandes desafios que ainda enfrenta a humanidade. Carros autônomos diminuirão os acidentes de trânsito ao dirigir sem margem para o erro. Na medicina, equipamentos de análise serão capazes de, avaliando exames de pacientes de doenças disseminadas, criar remédios e executar cirurgias, trazendo a cura. Grandes problemas multifacetados e com excesso de variáveis para a mente humana, como a mudança climática ou a completa exploração do universo, serão desafios superáveis para o algoritmo com a capacidade adequada de processá-los.

Não se trata de um cenário de muitas décadas à frente, mas uma realidade que nos envolve sem sentirmos. Na Dinamarca, serviços de emergência utilizam leitores de voz que verificam se os que ligam não estão sofrendo parada cardíaca. Clínicas de radiologia na Áustria encontram tumores com mais facilidade inserindo os exames de raios X e outras análises clínicas em computadores em vez de entregar a médicos. Fazendeiros deixam máquinas monitorando seu rebanho e adaptando a alimentação e o calor às suas necessidades, sem intervenção humana (European Commission, 2018, p. 1).

Os Estados não agem somente por expectativa de um futuro melhor, mas também premiados pelo risco da obsolescência. Evidentemente, uma tecnologia tão promissora o é sendo ao mesmo tempo disruptiva. Quem ficar para trás nessa corrida pode ter como consequência a perda abrupta de produtividade de sua economia, a dependência de tecnologias externas para manter a qualidade de vida de suas populações e, ao fim, a modificação do seu *status* na disputa geopolítica global.

Ao mesmo tempo, a aposta no avanço não se dá sem a presença de riscos muito evidentes. Os sistemas de inteligência artificial são treinados através da análise de um número tendente ao infinito de detalhes de seres humanos e de seu ambiente. Dados são, assim, a matéria prima para se produzirem melhores sistemas de IA. Contudo, esse conhecimento acumulado facilmente se transmuda de meio de melhora da vida humana para modo de classificação, discriminação, desconsideração e dominação do indivíduo, seja por seus controladores, seja pelo atingimento da autonomia robótica.

A flexibilidade em endereçar esses temores sem prejuízos para as *startups* na matéria envolve a presença de princípios normativos que possam refletir valores e ética no campo de conhecimento e inspirar os atores executivos da regulação a decidir em frente aos problemas concretos que surjam.

A globalização do interesse em determinado assunto permite que olhemos a legislação que se visa produzir internamente não como um feito único, mas como uma experiência em que se podem comparar vantagens e desvantagens de outros estudos através do debate racional.

A expansão mundial do interesse na IA encontra também eco nos fóruns internacionais. E as normas internacionais que surgem são especialmente úteis ao consagrar consensos entre países e visões de mundo que terminam por valorizar aqueles pontos que já parecem indiscutíveis. A relevância de se observarem as normas fruto de discussões de vários países se torna ainda mais intensa quando, assim como na temática do meio ambiente, há uma sensação de que decisões de um território podem ser fundamentais para os demais, num verdadeiro panorama de preocupação mundial compartilhada (Sands, 1994, p. 294).

Diante deste contexto de transformação tecnológica sem precedentes, emerge uma questão fundamental para os campos jurídico e tecnológico: como estabelecer marcos regulatórios que permitam o avanço da inteligência artificial, preservando simultaneamente os direitos humanos e os valores democráticos? O presente estudo propõe-se a realizar uma análise crítica e sistemática dos princípios internacionais para desenvolvimento e uso da inteligência artificial, com ênfase nos documentos normativos produzidos pela Organização para Cooperação e Desenvolvimento Econômico (OCDE) e pelo Conselho da Europa. Busca-se, por meio de uma investigação documental comparativa, identificar os consensos e divergências nos marcos regulatórios internacionais sobre inteligência artificial.

A relevância desta pesquisa reside na necessidade de compreender cientificamente como diferentes atores globais estão construindo respostas jurídicas para uma tecnologia com potencial simultâneo de transformação e disrupção. A inteligência artificial não representa apenas uma inovação tecnológica, mas um possível redesenho das estruturas institucionais contemporâneas, com implicações profundas nas esferas social, econômica e política.

Do ponto de vista metodológico, o estudo caracteriza-se como uma pesquisa documental de natureza qualitativa, utilizando como fontes primárias a Recomendação sobre Inteligência Artificial da OCDE (2019/2023) e a Convenção-Quadro sobre Inteligência Artificial do Conselho da Europa (2024). Serão analisados comparativamente os princípios, as diretrizes e as abordagens presentes nesses documentos internacionais, buscando extrair os elementos comuns e as perspectivas compartilhadas sobre o desenvolvimento ético da inteligência artificial.

O artigo visa contribuir para o debate acadêmico e para a compreensão dos esforços internacionais de regulação, oferecendo uma análise crítica dos princípios emergentes que orientam o desenvolvimento da inteligência artificial no cenário global contemporâneo.

Ao problematizar os marcos regulatórios internacionais, pretende-se lançar luz sobre os desafios e possibilidades de uma abordagem ética e responsável da inteligência artificial, fundamentada no respeito aos direitos humanos e aos valores democráticos.

1 NORMATIVOS INTERNACIONAIS FRUTO DE DISCUSSÕES MULTINACIONAIS SOBRE INTELIGÊNCIA ARTIFICIAL

A Organização para Cooperação e Desenvolvimento Econômico é responsável pela elaboração do primeiro documento intergovernamental para estabelecer parâmetros na

regulação interna da inteligência artificial. A Recomendação sobre IA foi adotada em 22 de maio de 2019, sendo subscrita pelos 36 países-membros da organização e por não membros como Argentina, Brasil, Colômbia, Costa Rica, Peru e Romênia (Yeung, 2020, p. 27). Com o propósito expresso de garantir a confiabilidade dos sistemas de IA e assegurar o respeito aos direitos humanos e à democracia, foi a seguir acolhida como princípios do G20 em Inteligência Artificial, já na reunião da cúpula dos países mais ricos em Osaka, Japão, em junho de 2020. Essa Recomendação, para se manter compatível com as novidades tecnológicas dos últimos anos, em especial a IA generativa, foi revisada em 8 de novembro de 2023 e referendada na reunião ministerial de 3 de maio de 2024 (OECD, 2024).

O fato de a OCDE e o G20 emitirem observações não vinculantes não deve rebaixar o significado da referida recomendação. Sua característica de *soft law* internacional influencia a formação de normas obrigatórias, legitima autorregulações no mesmo sentido e dificulta a justificação de posicionamentos contrários. Serve também como possível auxílio de interpretação em conflitos de solução hermenêutica. Fundamentalmente, marca estágios de consenso entre países que servem frequentemente de marcos para próximos acordos multilaterais exigíveis (Boyle; Chinkin, 2007, p. 211).

Não menos relevante é a conclusão da Convenção-Quadro sobre Inteligência Artificial e Direitos Humanos, Democracia e Estado de Direito pelo Conselho da Europa, em 17 de maio de 2024 (CoE, 2024b). A celebração de acordos internacionais em matérias econômicas, culturais, científicas, legais e administrativas é um dos principais objetivos do Conselho desde seu estatuto de fundação, de 1949, estando definida essa intenção já no artigo 1.b. E, embora o artigo 1.c do mesmo estatuto indique que essa atuação não afetará o trabalho em outras organizações internacionais e nas Nações Unidas, o fato é que, por vezes, o fracasso na definição conjunta de documentos com escala mundial termina tornando o texto no Conselho da Europa o verdadeiro tratado multilateral abrangente sobre determinados assuntos, inclusive pela tendência de sempre abrir a adesão de não membros (Benoît-Rohmer; Klebes, 2004, p. 94)¹. Isso se vê, por exemplo, no tema da criminalidade informática, com a Convenção de Budapeste e seus protocolos, e na transferência de condenados, com a Convenção de Estrasburgo.

O texto da Convenção-Quadro sobre Inteligência Artificial começou a ser elaborado em 2019, por um Comitê *ad hoc* para Inteligência Artificial (CAHAI). Essa tarefa foi assumida em 2022 pelo Comitê de Inteligência Artificial (CAI), que propôs e negociou a versão final, aceita pelos 46 membros do Conselho, assim como pelos países observadores Canadá, Japão, México, Santa Sé e Estados Unidos. Também a União Europeia e países não membros participaram desse processo, especificamente Austrália, Argentina, Costa Rica, Israel, Peru e Uruguai. Houve também ampla participação de atores não estatais, com 68 entidades da sociedade civil e da indústria presentes nas discussões como observadores (CoE, 2024a, p. 1).

A relevância de estudar essas normas que já chegaram a ser consensuadas se torna mais clara ao verificarmos o panorama geral do direito internacional, em especial quando

1 Desde 1987, não há acordos "fechados" a não integrantes da organização.

referente a novas tecnologias. O cenário de implementação desse tipo de normas pelas organizações multilaterais tem sofrido com o incremento de rivalidades no plano comercial e com a obsolescência dos métodos diplomáticos tradicionais de construção lenta de unanimidades através da negociação em face da premência de regulações em grande escala territorial. A consequência tem sido frequentemente a inação ou a edição de regramentos que já se veem desatualizados quando chegam a se tornar realidade (Abhivardhan, 2023, p. 51). Nesse interim as grandes corporações agem sem quaisquer limites significativos, à margem da ação estatal coordenada.

Devem ser reconhecidos aqui os limites da autorregulação privada sem a guia do Direito a partir do(s) Estado(s). A ausência de uma visão estereotipada de corporações sem componentes humanos e que só visam ao lucro não significa deixar de compreender os limites e desafios de superar a concorrência num palco em que a evolução tecnológica ocorre em velocidade notável. A busca de cada vez maior engajamento de usuários tem como importante componente a sua monetização, ou seja, a transformação das suas interações em faturamento na forma da produção de anúncios feitos de modo direcionado aos seus interesses verificados. Esse modelo só alcança o auge de eficiência ao haver a extração máxima possível dos dados dos consumidores de conteúdo, permitindo sua categorização máxima. Ao mesmo tempo, deve-se garantir que sua atenção não seja desviada, o que implica não fornecer mensagens que podem desagradar quem utiliza o aplicativo, ainda que isso signifique reforçar ideias pré-concebidas e não estimular o contraditório e a reflexão (Abhivardhan, 2023, p. 274-276). Vê-se, assim, que o meio ótimo de se alcançarem os ganhos que permitem às empresas investirem mais em desenvolvimento se choca com o estilo melhor de abordagem de esferas de convivência de seres humanos numa sociedade democrática.

Portanto, a análise dos pontos em comum e mais destacados desses documentos internacionais permite verificar quais disposições sobre inteligência artificial já possuem um excepcional alto grau de concordância em cerca de 1/3 de países com percentual de participação na riqueza mundial ainda mais pronunciado.

A exposição comum dos dois diplomas se concentrará no que há de similar, consistindo eminentemente em princípios de inteligência artificial, normas de caráter mais amplo e abstrato que, sem perder aplicabilidade, servem também de inspiração para regras mais específicas e de interpretação geral dos estatutos e do que existe de direitos humanos específicos na inteligência artificial. São temáticas cuja união entre diversos países nas negociações internacionais pode ser explicada pela anuência na sua relevância que já existia na literatura especializada.

2 PRINCÍPIOS DE INTELIGÊNCIA ARTIFICIAL

2.1 UTILIZAÇÃO EM PROL DO DESENVOLVIMENTO SUSTENTÁVEL, BEM-ESTAR E CRESCIMENTO DAS SOCIEDADES E DO PLANETA

Esse é o princípio constante do item 1.1 da Recomendação da OCDE. Há dois grandes focos aqui: o crescimento da sociedade e a redução de desigualdades. Quanto a esse último,

verifica-se que a base de dados que serve para treinamento dos algoritmos de inteligência artificial pode ser contaminada pelos vieses de visão de mundo e julgamento que existem nos seres humanos que os alimentam, provocando resultados que refletem as discriminações existentes na sociedade (Fosch-Villaronga; Poulsen, 2022, p. 123). O fim aqui, portanto, é incentivar o contrário, que haja meios na regulação que estimulem maior diversidade e permitam, através da tecnologia, a maior inclusão de minorias sub-representadas nas esferas públicas e privadas. Essa ideia de evitar preconceitos na execução dos sistemas de IA e atingir resultados com maior justiça e equidade se reproduz também nos artigos 10.1 e 10.2 da Convenção-Quadro do Conselho da Europa.

Além disso, o ganho de produtividade no uso de sistemas de inteligência artificial deve servir para substituir as tarefas repetitivas executadas por indivíduos, liberando-os e auxiliando-os para se empenharem em atividades que requeiram maior grau de criatividade.

O meio ambiente pode se ver muito beneficiado pelas soluções trazidas a partir de algoritmos que podem analisar os múltiplos fatores que auxiliam na conservação das espécies (Vinuesa *et al.*, 2020, p. 5). Ao mesmo tempo, não se pode desconsiderar o elevado nível de energia e recursos naturais, em especial água, que são necessários na manutenção em funcionamento dos grandes processadores.

2.2 RESPEITO AO ESTADO DE DIREITO, AOS DIREITOS HUMANOS E VALORES DEMOCRÁTICOS

Princípio presente do item 1.2 da Recomendação da OCDE. Dentro da ideia central – uso da tecnologia sem violar a dignidade humana e os valores do império da lei e do regime representativo –, algumas questões mereceram destaque. O reconhecimento de que, em todas as grandes democracias mundiais, vem ocorrendo a utilização das redes sociais para fins de desinformação, com a manipulação da opinião pública para que a polarização política advinda permita a ocupação de espaços de poder pelos seus agentes (Vinuesa *et al.*, 2020, p. 6), exige que o tema não possa passar sem enfrentamento pelas instituições constituídas, ainda que com o cuidado de ser respeitada sempre a liberdade de expressão dos cidadãos.

A constatação de que as bases de dados da inteligência artificial demandam uma quantidade enorme de informações sobre indivíduos impõe a necessidade de se reafirmar a importância de que a privacidade e o direito à proteção de dados sejam integralmente respeitados, inclusive com garantias efetivas ao indivíduo quando houver violações. Isso está expresso no artigo 11 da Convenção-Quadro (CoE, 2024b).

Fundamentalmente, a garantia de que a tecnologia de IA esteja sempre sendo utilizada conforme os direitos humanos exige a constante supervisão privada e estatal durante todo seu ciclo de existência (desde o design até o encerramento) e a existência de meios efetivos de contestação por parte das pessoas afetadas. A certeza de que padrões éticos adequados estejam sendo respeitados quando há vigilância, monitoramento e armazenamento de dados pessoais por meio de processos automatizados, com uso elevado de raspagem de informações na rede, impõe a presença de meios para alterar, deletar e ser compensado por afirmações errôneas que surjam a partir das respostas mecanizadas (Vinuesa *et al.*, 2020, p. 5).

As plataformas também devem responder pelas consequências que têm para o debate público, a tarefa que exercem, com intuito de lucro, de moderação de redes sociais. Isso significa que devem exercer o controle de exclusão de conteúdos considerados ilícitos pela ordem jurídica e assim submeterem-se ao controle estatal (Elkin-Koren; Perel, 2020, p. 852-853). Mas, ao mesmo tempo, implica que têm o dever de implementar meios de essa atividade manter o acesso equânime à ágora tecnológica e estimular o exercício da liberdade de expressão, em conformidade com o artigo 5.2 da Convenção-Quadro (CoE, 2024b).

Deve ser sempre reconhecida a sensibilidade dessa atividade de excluir e de promover informação. Regimes reconhecidos como de supressão de amplitude de opiniões, como o iraniano, utilizam a IA para excluir sem intervenção humana mensagens consideradas nocivas para sua população e impulsionam *bots* que propagam de forma maciça aquelas narrativas interessantes ao poder político (Ünver, 2024, p. 34).

A inteligência artificial também não deve ser utilizada para afetar a separação de poderes, conforme o artigo 5.1 da Convenção-Quadro (CoE, 2024b). A possibilidade de o Poder Executivo ou de entes privados se aproveitarem de sistemas de informação para coagir Poderes Públicos a se submeterem às suas decisões tem especial relevância quando o alvo é o Poder Judiciário. A liberdade de pressões externas, que é indispensável à independência judicial, envolve regras que evitem o uso da inteligência artificial de forma a violar a autonomia do juiz e a igualdade de armas. O uso da tecnologia preditiva em julgamentos deve ser sempre no sentido do auxílio à conclusão do ser humano, com caráter científico e objetivo e, preferencialmente, corroborado por outras evidências. O uso de IA em tribunais não deve assumir a imposição de constrangimentos à atividade jurisdicional que maculem a real existência de um contraditório e da presunção de inocência (Greenstein, 2022, p. 314).

2.3 TRANSPARÊNCIA E EXPLICABILIDADE

A *transparência* de um sistema de IA está relacionada com o fato de ele não se comportar como uma "caixa preta", ou seja, que suas estruturas e comando estejam abertos à análise alheia. No entanto, essa característica não se apresenta como suficiente. A complexidade dos modelos exige do responsável um passo a mais, que sejam explicáveis de forma acurada, mas igualmente compreensível aos seres humanos que estejam em interação com ele de qualquer forma.

Essa diversidade de pessoas à qual se submete a *explicabilidade* significa que esse é um conceito que precisa ser adaptado constantemente ao nível de conhecimento de cada um dos grupos de interlocutores necessários (Angelov *et al.*, 2021, p. 4-5). Aqui, portanto, tratamos conjuntamente de uma série de *stakeholders* para quem deverão ser proporcionadas diferentes respostas sobre o funcionamento da IA que necessitarão de ajuste de maior ou menor simplificação, conforme a expertise que apresentem.

Esses princípios são fundamentais para o trabalho dos órgãos de fiscalização, que, conquanto devam procurar ter os profissionais mais habilitados em seus quadros, frequentemente estarão em estágio inferior aos empregados da indústria de informática. A transparência e a explicabilidade ainda são mais fundamentais para os indivíduos afetados por

interação voluntária ocasional ou por relações jurídicas com maior constância e estabilidade, como o contrato de trabalho. Nesses casos, o caráter leigo dos diretamente sujeitos à atuação da IA leva à maior importância da linguagem clara e acessível para pessoas com provável baixo nível de formação na matéria.

Ambos os princípios seriam inúteis sem meios administrativos, procedimentos extrajudiciais e garantias processuais para o acesso. É importante, assim, que o conhecimento externo dos usos e a disponibilização ampla dessas informações sejam buscados de forma ativa pelo responsável pelo sistema, juntamente com as formas disponíveis para buscar mais detalhes gerais ou específicos.

Fica evidente que, sem transparência e explicabilidade, boa parte das regras para controle estatal e social dos desenvolvimentos e usos da inteligência artificial terminariam na prática inaplicadas, o que justifica uma certa reiteração nos diplomas internacionais de disposições com esse conteúdo².

2.4 ROBUSTEZA, SEGURANÇA E INTEGRIDADE

Os três princípios indicados no item 1.4 da Recomendação da OCDE devem ser trabalhados em conjunto e se referem à confiabilidade que se exige da Inteligência Artificial. O sistema deve ser preparado para resistir a circunstâncias hostis, sejam elas fruto de erros nos dados utilizados, de ataques cibernéticos externos, ou de ocorrência de resultados equivocados ou inesperados.

A *robustez*a do sistema se mede inicialmente pela pertinência dos *outputs* em relação ao que se visava originalmente. Isso pode ocorrer tanto por execução errônea do algoritmo quanto por ausência ou inverdade na informação na base de dados (Chander *et al.*, 2024, p. 23). A *segurança*, conceito já presente no direito à proteção de dados, trata da estrutura que é criada para evitar danos na manipulação do sistema. Isto é, por um lado, os elementos de software e componentes de equipamentos e rede que impeçam acessos por *hackers*; por outro, a existência da possibilidade de supervisão e intervenção humana interna para corrigir falhas que se apresentem ou mesmo para parar ou destruir a IA em funcionamento com algum tipo de independência. Já a *integridade* se refere ao esforço, que é inicial, mas também contínuo, para assegurar que os dados pessoais utilizados contenham veracidade, ou seja, conformidade ao espectro da realidade que visam reproduzir. Além disso, a base de dados deve ter a amplitude necessária para não incidir em desvios de análise. Há conseqüentemente a necessidade de um treinamento com ênfase na qualidade e na quantidade da informação.

Essas características não se alcançam por simples acerto na concepção ou cuidado no recolhimento de dados, embora, evidentemente, esses fatores sejam essenciais. Os sistemas devem estar sendo sempre testados por adversários estabelecidos pelo próprio responsável, envolvendo os diferentes aspectos, de modo a que se possa medir e verificar o que ainda deve ser melhorado nos mecanismos de defesa (Chander *et al.*, 2024, p. 26).

2 Vide, nesse sentido, artigos 8, 14 e 15 da Convenção do Conselho de Europa e alíneas I a IV do artigo 1.3 da Recomendação da OCDE.

Tecnologias automatizadas para reconhecimento facial são especialmente sensíveis a esses princípios. A ausência de uma base adequada pode levar a “falsos positivos” que impactam de maneira desproporcional mulheres e minorias. E o ônus de provar a conformidade deve ser suportado por aquele que visa se aproveitar do instrumento. Nesse sentido, o caso *Bridges* no Reino Unido é um marco na jurisprudência sobre o tema. Embora algumas cidades, especialmente as norte-americanas, como São Francisco ou Berkeley, tenham suspenso a integração de câmeras de vigilância com IA de identificação algorítmica de pessoas com base nessas preocupações, foi da Corte de Apelações britânica o primeiro julgado a afirmar que cabia ao Estado, no caso a polícia de South Wales, demonstrar, diretamente ou via auditoria independente, que seu sistema não violava a igualdade entre os cidadãos (Zalnieriute, 2021, p. 288 e 296).

2.5 RESPONSABILIDADE (E RASTREABILIDADE)

As tecnologias que se utilizam de inteligência artificial só serão verdadeiramente úteis ao desenvolvimento humano se os atores que as desenvolvem e se aproveitam delas puderem arcar com as consequências de eventuais danos que causarem. Para tanto, o processo de design, desenvolvimento e funcionamento deve estar registrado de forma precisa e inteligível. A rastreabilidade é um pressuposto para a responsabilização e consiste na documentação que demonstra cronologicamente cada operação durante o ciclo de vida de um sistema de IA. São essas provas a serem obrigatoriamente e contemporaneamente produzidas em relação a cada ente que se aproveite da inteligência artificial, sejam criadores do conteúdo, sejam utilizadores, as quais servirão para definir as imputações em caso de dano a pessoas (Williams *et al.*, 2022, p. e7-7). Essa possibilidade de traçar o caminho é especialmente relevante para a IA generativa, já que a criação de mídia se dá a partir de elementos humanos prévios, com possibilidade de ausência de anuência daquele que vê suas características corpóreas ou produto intelectual aproveitados por outrem.

Saber quem é o responsável por determinado ato envolvendo a IA não se confunde com a imediata responsabilização, já que os parâmetros de gerenciamento do risco do negócio, se este foi lícito ou não, ainda dependem da avaliação jurídica na forma da regulação de cada país no âmbito administrativo, trabalhista, penal e/ou cível, com destaque para os direitos da propriedade artística e intelectual. O imprescindível é que violações a direitos humanos por modelos com qualquer tipo de autonomia também sejam em tese passíveis de provocar sancionamentos em relação às pessoas físicas e jurídicas que se aproveitaram daquela tecnologia.

Essa questão se torna também especialmente sensível no cenário eleitoral. A manipulação por meio da inteligência artificial visando a desinformação do eleitor vem ocorrendo em países tão diversos como Bangladesh, Turquia, Brasil ou Argentina (Ünver, 2024, p. 15) e impõe medidas céleres, sob pena de macular a expressão da vontade popular. O desafio aqui é mais intenso em face da exiguidade do período de campanha, da imprescindibilidade de evitar injustiças e da possibilidade de efeitos duradouros na mente do votante, mesmo com ação efetiva das autoridades.

Além da indicação geral da rastreabilidade e da responsabilidade como inerentes ao regime da IA no item 1.5 das Recomendações da OCDE, a Convenção do CE trata expressamente da responsabilização civil e indenização no seu artigo 9.

2.6 COOPERAÇÃO

Muita embora não diretamente referida como princípio, ambos os normativos valorizam a cooperação público-privada e entre países como meio indispensável para termos uma regulação efetiva em escala global (item 2.5 da Recomendação da OCDE e artigo 25 da Convenção do CE). As iniciativas locais, ainda que úteis, têm uma eficácia naturalmente limitada neste tema, em que a tecnologia pode ser desenvolvida num espaço e rapidamente expandida para outro através da internet. As políticas menos cuidadosas quanto a riscos de um determinado Estado podem, portanto, ter rapidamente consequências sobre os demais.

E, quando se fala na importância de entes privados também participarem desse esforço conjunto, ainda que a referência envolva certamente as grandes empresas de tecnologia, não somente nesse sentido deve ser entendida a alusão, mas também se estende a institutos de pesquisa, acadêmicos, sociedade civil em geral e indivíduos. A inerente competição do sistema capitalista de desenvolver melhores sistemas que seus concorrentes não acarreta que os princípios aqui descritos não devam ser perseguidos por quaisquer atores que atuem na IA. Os meios serem distintos e se buscar a liderança não significa que os valores não possam ser comuns em benefício da humanidade (Ala-Pietilä; Smuha, 2021, p. 241 e 242).

A tão buscada confiabilidade de sistemas de IA, ao final uma sensação que deve estar presente a nível individual, exige, assim como em outros mercados, como o aéreo e o bancário, que exista um coletivo ambiental de criação e desenvolvimento que atenda a padrões e regras semelhantes mundialmente para serem encaradas como aceitáveis. A verificabilidade em cada local depende da anuência em conteúdos substantivos e procedimentais, assim como a existência de fiscalização cruzada pelos países parceiros, em nível regional, setorial ou global, do atendimento na prática. Ao final, esse modelo permite assegurar que haja uma maximização de benefícios para a humanidade, combinada com minimização de riscos (Ala-Pietilä; Smuha, 2021, p. 245, 247 e 252).

CONCLUSÃO

Há uma natural e bem-vinda corrida dos países no sentido de regular a inteligência artificial em seus territórios. A busca visa evitar que uma tecnologia com enorme potencial de mudar os modos de vida de nossas sociedades em muito pouco tempo esteja em funcionamento sem as salvaguardas necessárias e o controle estatal e social já em marcha. No entanto, a fúria legislativa pode levar também a um mosaico caótico de regras e sem a maturação necessária antes da sua edição.

Por isso, merecem destaque as (poucas) hipóteses em que países se reuniram e conseguiram expor propostas para a normatização comum. A superação das inerentes rivalidades

econômicas na vanguarda da inovação é resumida em propostas relativamente curtas, mas que expõem unanimidades já possíveis.

Ainda que haja similitude de participantes no contexto da OCDE e do Conselho da Europa, estes não são comunidades internacionais idênticas e representam parcela significativa do cenário de ponta da pesquisa em IA e do PIB mundial. Por isso, e sendo os normativos internacionais, vinculantes ou não, já existentes, merecem o apontamento de seus princípios comuns.

Além disso, quando observamos as proximidades em conteúdo, vemos consagradas questões de real alcance comungado. Parece certo que a inteligência artificial pode ser um elemento para o florescimento de sociedades mais justas, ricas e igualitárias e que não pode ser uma ameaça aos direitos humanos, ao Estado de direito e ao regime democrático.

Sistemas computadorizados já não são uma novidade para as gerações hoje presentes no nosso planeta. A importância de que tenham seus meandros claramente explicitados para reguladores e afetados, que funcionem sem falhas internas ou externas e que permitam o conhecimento daqueles a quem devem ser imputados os danos que causarem, é de fácil compreensão. A complexidade com a qual ainda se apresenta a IA neste momento histórico explica somente um maior adensamento desses conceitos, para que abarquem todos os aspectos já identificáveis.

Todos esses objetivos serão inalcançáveis sem a junção de países e entes privados envolvidos em prol da sua consecução. Ao lado da busca de cada um de alcançar os sistemas mais reconhecidos como melhores, deve estar um substrato comum de respeito a esse índice compartilhado, sob pena de fracassarmos todos. Essa junção imprescindível além do espaço físico de países deve ser constante e crescente, até permitir uma certeza geral de que podemos acreditar todos no potencial modificador positivo da automatização de funções antes eminentemente exercidas pelos seres humanos.

REFERÊNCIAS

ABHIVARDHAN, A. **Artificial intelligence ethics and international law**. 2. ed. Delhi: BPB Publications, 2023.

ALA-PIETILÄ, Pekka; SMUHA, Nathalie A. A framework for global cooperation on artificial intelligence and its governance. In: BRAUNSCHWEIG, Bertrand; GHALLAB, Malik (org.). **Reflections on artificial intelligence for humanity**. Cham: Springer International Publishing, 2021. (Lecture Notes in Computer Science). v. 12600. p. 237-265. DOI: <https://tinyurl.com/ymy3z32j>.

ANGELOV, Plamen P.; SOARES, Eduardo A.; JIANG, Ricardo; ARNOLD, Nicolau; ATKINSON, Peter M. Explainable artificial intelligence: an analytical review. **WIREs Data Mining and Knowledge Discovery**, v. 11, n. 5, p. e1424, 2021.

BENOÎT-ROHMER, Florence; KLEBES, Heinrich. **Council of Europe law towards a pan-European legal area**. Strasbourg: Conseil de l'Europe, 2004.

BOYLE, Alan E.; CHINKIN, Christine. **The making of international law**. Oxford, New York: Oxford University Press, 2007.

CHANDER, Bhanu; JOHN, Chinju; WARRIER, Lekha; GOPALAKRISHNAN, Kumaravelan. Toward trustworthy artificial intelligence (TAI) in the context of explainability and robustness. **ACM Computing Surveys**, jun. 2024. [Online]. DOI: <https://doi.org/10.1145/3675392>. Acesso em: 15 jul. 2024.

CoE – COUNCIL OF EUROPE. **Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law**: Council of Europe Treaty Series - No. [225]. Vilnius: Council of Europe, 2024a.

CoE – COUNCIL OF EUROPE. **Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law**. 2024b. Disponível em: <https://tinyurl.com/5xdv2k9s>. Acesso em: 20 maio 2024.

ELKIN-KOREN, Niva; PEREL, Maayan. Separation of functions for ai: restraining speech regulation by online platforms. **Lewis & Clark Law Review**, v. 24, 2020. DOI: <http://dx.doi.org/10.2139/ssrn.3439261>.

EUROPEAN COMMISSION. **Artificial Intelligence for Europe**. Brussels: European Commission, 2018.

FOSCH-VILLARONGA, Eduard; POULSEN, Adam. Diversity and inclusion in artificial intelligence. In: CUSTERS, Bart; FOSCH-VILLARONGA, Eduard (org.). **Law and artificial intelligence**. The Hague: T.M.C. Asser Press, 2022. (Information Technology and Law Series). v. 35, p. 109-134. Disponível em: <https://tinyurl.com/2mycnhzm>. Acesso em: 21 jun. 2024.

GREENSTEIN, Stanley. Preserving the rule of law in the era of artificial intelligence (AI). **Artificial Intelligence and Law**, v. 30, n. 3, p. 291-323, 2022. DOI: <https://link.springer.com/article/10.1007/s10506-021-09294-4>.

OECD. **Recommendation of the Council on Artificial Intelligence**, OECD/LEGAL/0449. 2024.

SANDS, Philippe. The “greening” of international law: emerging principles and rules. **Indiana Journal of Global Legal Studies**, v. 1, n. 2, 1994. [Online]. Disponível em: <https://tinyurl.com/34cah3ku>. Acesso em: 21 jun. 2024.

ÜNVER, H. A. **Artificial intelligence (AI) and human rights**: using AI as a weapon of repression and its impact on human rights. Brussels: European Union, 2024.

VINUESA, Ricardo; AZIZPOUR, Hossein; LEITE, Iolanda; BALAÃO, Madalena; DIGNUM, Virgínia; DOMISCH, Sami; FELLÄNDER, Anna; LANGHANS, Simone Daniela; TEGMARK, Max; NERINI, Francisco Fuso. The role of artificial intelligence in achieving the Sustainable Development Goals. **Nature Communications**, v. 11, n. 1, p. 233, 2020.

WILLIAMS, Rebeca; CLOETE, Ricardo; COBBE, Jennifer; COTTRILL, Caitlin; EDWARDS, Pedro; MARKOVIC, Milão; NAJA, Imã; RYAN, Frances; SINGH, Jatinder; PANG, Wei. From transparency to accountability of intelligent systems: Moving beyond aspirations. **Data & Policy**, Cambridge, v. 4, p. e7, 2022. DOI: <https://doi.org/10.1017/dap.2021.37>.

YEUNG, Karen. Recommendation of the Council on Artificial Intelligence (OECD). **International Legal Materials**, v. 59, n. 1, p. 27-34, 2020. [Online].

ZALNIERIUTE, M. Burning Bridges: The automated facial recognition technology and public space surveillance in the modern state. **Science and Technology Law Review**, v. 22, n. 2, p. 284-307, 2021.