

INTELIGÊNCIA ARTIFICIAL
CENTRADA EM PESSOAS: PROTEÇÃO
DE DADOS E DIREITOS FUNDAMENTAIS
NO CENÁRIO GLOBAL

*PEOPLE-CENTERED ARTIFICIAL
INTELLIGENCE: DATA PROTECTION
AND FUNDAMENTAL RIGHTS IN THE
GLOBAL CONTEXT*

INTELIGÊNCIA ARTIFICIAL CENTRADA EM PESSOAS: PROTEÇÃO DE DADOS E DIREITOS FUNDAMENTAIS NO CENÁRIO GLOBAL¹

PEOPLE-CENTERED ARTIFICIAL INTELLIGENCE: DATA PROTECTION AND FUNDAMENTAL RIGHTS IN THE GLOBAL CONTEXT

Stefani Juliana Vogel²

RESUMO

Este artigo exploratório investiga a importância dos princípios éticos para o desenvolvimento e uso de sistemas de inteligência artificial (IA), com foco na confiança e no respeito aos direitos fundamentais, especialmente na proteção de dados pessoais. Utilizando método dedutivo e revisão bibliográfica, o estudo apresenta as principais construções teóricas sobre IA centrada no ser humano. São abordados os desafios éticos e os riscos inerentes à IA sugerindo caminhos para uma IA ética e centrada em pessoas, essencial para a democracia e o debate público. Conclui destacando uma nova perspectiva para enfrentar os desafios contemporâneos em proteção de dados e privacidade.

Palavras-chave: inteligência artificial; direitos humanos; ética; equidade; justiça.

1 INTRODUÇÃO

As tecnologias baseadas em inteligência artificial, doravante denominada IA, têm o potencial de alterar profundamente a forma como as pessoas se veem e interagem com o mundo. Aplicações que vão desde assistentes virtuais até algoritmos preditivos são capazes de moldar comportamentos, influenciar decisões e criar relações de poder. Nesse sentido, é imperativo que a discussão sobre IA não se limite a seus benefícios técnicos, mas abarque os múltiplos impactos sociais. A confiança nestes sistemas é uma condição prévia para a aceitação da IA, dado o impacto transversal que ela possui (Comissão Europeia, 2020).

¹ Data de Recebimento: 28/10/2024. Data de Aceite: 05/11/2024.

² Chefe de Gabinete na Câmara dos Deputados e coordenadora do Grupo de Trabalho Mulheres no Direito Digital e nas Novas Tecnologias do Instituto Empoderar. Foi presidente suplente do Conselho Nacional de Proteção de Dados Pessoais (CNPD). É mestre em Alta Dirección en Seguridad Internacional pelo Centro Universitario de la Guardia Civil e Universidade Carlos III da Espanha. E-mail: dpvogel@gmail.com. Endereço para acessar currículo lattes: <http://lattes.cnpq.br/0205978399239505>. ORCID: 000-0003-4239-7248.

Como afirma Steven Pinker (2018, p. 479), o progresso da civilização está ancorado na centralidade da razão, ciência e humanismo, com o ser humano no centro do desenvolvimento científico e tecnológico. Esse enfoque no ser humano é, portanto, não apenas desejável, mas essencial para assegurar que o progresso tecnológico traga benefícios reais à sociedade.

O surgimento de marcos legais dedicados a estabelecer diretrizes sobre o uso e o desenvolvimento da IA tem pautado preocupações relacionadas à privacidade, ética e a proteção de dados pessoais. Nos últimos anos, o conceito de IA focada em pessoas ou *Human-Centered AI* ganhou relevância, ao apresentar uma abordagem que coloca os seres humanos no centro do desenvolvimento e uso da IA e, ao mesmo tempo garantindo que seus direitos e sua dignidade sejam priorizados.

Este enfoque ocorre em meio a uma regulação global crescente, como o *AI Act* europeu, aprovado pelo Conselho da União Europeia em maio de 2024 e o Projeto de Lei 2338/2023³ em discussão no Brasil. Ambos buscam estabelecer normas de governança robustas, impondo obrigações proporcionais ao potencial de risco e impacto de diferentes sistemas de IA, promovendo, assim, o seu uso ético e responsável. Esses marcos regulatórios refletem uma preocupação global em estruturar uma abordagem de IA baseada em princípios e comprometida com a segurança e os direitos fundamentais, alinhando a inovação tecnológica aos valores sociais.

Dado o caráter invasivo de muitas tecnologias associadas a IA e o seu potencial de alterar o comportamento e identidade do ser humano, o objetivo do presente trabalho é analisar os contornos ético-jurídicos que emergem dos avanços tecnológicos e o seu impacto sobre os direitos humanos. Para o desenvolvimento desta pesquisa, realizou-se pesquisa bibliográfica na doutrina nacional e estrangeira, por meio de materiais impressos e digitais, como livros e revistas científicas, com a finalidade de abordar temas centrais como a IA centrada no ser humano e as suas implicações éticas e jurídicas.

2 IA FOCADA EM PESSOAS: CONCEITO E PRINCÍPIOS

A interdependência entre ética, política e direito exige que as normas que venham a regular a IA não sejam apenas ferramentas técnicas, mas também reflitam os valores ético-políticos fundamentais de nossa sociedade. O conceito de IA centrada em pessoas parte da premissa de que o desenvolvimento e uso de sistemas de IA devem ter como

³ No momento do fechamento do presente trabalho, o PL 2338, de 2023, uma iniciativa do senador Rodrigo Pacheco, a partir do trabalho de uma comissão de juristas se encontrava na Comissão Temporária Interna sobre Inteligência Artificial no Brasil (CTIA) com o relator. O projeto visa regulamentar o uso de sistemas de IA no país, abordando questões como a transparência, proteção de dados e responsabilidade no uso dessas tecnologias. Para mais detalhes, acesse: <https://www25.senado.leg.br/web/atividade/materias/-/materia/157233>. Acesso em: 24 out. 2024.

objetivo fundamental o bem-estar humano, garantindo que as tecnologias trabalhem a serviço das pessoas e não o contrário. Dada sua natureza como tecnologia de propósito geral e o seu potencial impacto social, tanto sob o ponto de vista individual, como coletivo, como são os sistemas de seleção de candidatos para ingresso em empregos e estabelecimentos de ensino e os destinados a avaliar o acesso a serviços e benefícios públicos, torna-se essencial adotar uma abordagem ética para a IA.

Diversos princípios globais foram estabelecidos para guiar o desenvolvimento sustentável e equilibrado da IA. Em uma análise das principais iniciativas internacionais, esses princípios se organizam em seis dimensões: Equidade (*Fairness*); Confiabilidade e Segurança (*Reliability & Safety*); Impacto Social (*Social Impact*); Responsabilidade (*Accountability*); Privacidade & Segurança (*Privacy & Security*); e Transparência (*Transparency*) (Burle e Cortiz, 2019).

Os Princípios de Asilomar, por exemplo, elaborados pelo *Future of Life Institute* na *Beneficial AI Conference* (Asilomar, 2017), reúnem diretrizes focadas em maximizar os benefícios e mitigar riscos da IA. Entre os seus 23 princípios, destaca-se a “IA centrada no ser humano” como objetivo central da pesquisa, orientando para que a tecnologia seja projetada para o benefício humano e o bem-estar social (Floridi & Cowls, 2019).

Em “Ética e Valores”, o princípio da privacidade também é fundamental, assegurando que as pessoas mantenham controle sobre seus dados. O respeito à privacidade e à proteção de dados pessoais, a partir do tratamento de dados realizado de forma transparente e responsável, fortalece a confiança nos sistemas de IA. Em “Questões de Longo Prazo”, há um chamado para supervisão humana contínua, a fim de evitar que tecnologias autônomas ajam de forma independente e irresponsável. Assim, a IA centrada em pessoas orienta o desenvolvimento de sistemas que distribuam benefícios entre todos e assim minimizem desigualdades.

Além disso, as “Questões de Longo Prazo” reforçam uma perspectiva centrada no ser humano, ao estabelecer que a IA deve ser desenvolvida com um alinhamento de valores que esteja em conformidade com os direitos humanos e os interesses da humanidade como um todo. Isso inclui a supervisão humana contínua dos sistemas de IA para evitar que essas tecnologias atuem de forma autônoma e sem responsabilidade. Em um cenário onde sistemas autônomos podem tomar decisões importantes sem a participação direta de pessoas, é essencial criar diretrizes éticas e jurídicas que assegurem a responsabilidade e a transparência dessas ações.

O compartilhamento dos benefícios gerados pela IA entre todos os seres humanos é outro princípio fundamental, pois propõe que os avanços tecnológicos não resultem em exclusão ou aumento de desigualdades, mas sim em um impacto positivo global. Em suma, os Princípios de Asilomar constituem um importante referencial focado no ser

humano, ao focar em construir um futuro onde a IA contribua para o bem-estar coletivo e em respeito as normas éticas e os direitos de cada indivíduo.

O Regime sobre IA, robótica e tecnologias conexas aprovado pelo Parlamento Europeu em 2020 reforça o foco antropocêntrico e antropogênico dessas tecnologias, priorizando a dignidade, os direitos e a autonomia humana. A resolução recomenda supervisão e transparência em decisões automatizadas, protegendo contra discriminação e assegurando a segurança dos dados. Da mesma forma, o *Statement on Artificial Intelligence, Robotics, and Autonomous Systems* do *European Group on Ethics in Science and New Technologies* (EGE, 2018) enfatiza a necessidade de um quadro ético e legal reconhecido internacionalmente, alinhado aos valores morais e aos objetivos socioeconômicos da Europa.

A centralidade do ser humano fica evidenciada no conjunto de princípios éticos fundamentais que devem guiar o desenvolvimento e uso dessas tecnologias. Entre os princípios destacados, a dignidade humana é mencionada como fundamental, sendo descrita como a necessidade de respeitar o valor inerente de cada pessoa. Isso implica inferir que as tecnologias autônomas não devem violar a dignidade humana e que deve haver limites legais sobre como as pessoas são classificadas e tratadas por algoritmos e sistemas autônomos.

Outro destaque é a Organização para a Cooperação e Desenvolvimento Econômico (OCDE), que desde 2016 discute princípios e recomendações sobre IA. Em 2019, a OCDE publicou dez princípios para a IA, focados na proteção dos direitos humanos, autonomia e bem-estar, com a adesão de 47 países, incluindo a União Europeia. Em 2024, o princípio de “Crescimento inclusivo, desenvolvimento sustentável e bem-estar” foi atualizado para destacar o papel da IA na promoção dos Objetivos de Desenvolvimento Sustentável (ODS). Também se destaca o princípio de “Respeito pelo Estado de Direito, direitos humanos e valores democráticos”, que enfatiza o desenvolvimento de uma IA responsável, garantindo supervisão e segurança para prevenir riscos e assegurar justiça e transparência. O foco central é garantir que os valores humanos prevaleçam e que a IA seja desenvolvida e utilizada para promover o bem-estar, respeitando a dignidade e a autonomia dos indivíduos. A implementação de mecanismos de supervisão e salvaguardas são previstos como instrumentos-chave para atingir esse objetivo, prevenindo riscos e garantindo que os sistemas de IA sejam utilizados de forma segura, justa e transparente.

O *White Paper on Artificial Intelligence* (Comissão Europeia, 2020) estabelece a confiança como base para o quadro regulatório europeu, ao passo que as *Ethics Guidelines for Trustworthy AI* (União Europeia, 2018) promovem quatro princípios éticos: respeito à autonomia humana, prevenção de danos, justiça e explicabilidade. Esses princípios reforçam que os sistemas de IA devem empoderar os usuários, permitindo que compreendam e controlem suas interações com a tecnologia.

Mais recentemente, o relatório “Governing AI for Humanity” (2024) propôs um apelo para colaboração entre governos, empresas e sociedade civil para a criação de um ambiente seguro e ético para a IA. Diante do déficit das regulações, normas e instituições capazes de gerir a IA, propôs-se uma governança global que proteja os direitos humanos e maximize o bem-estar.

Nota-se que a IA centrada no ser humano atua como um meta princípio que conforma todos os demais. Esta abordagem prioriza valores e necessidades humanas no desenvolvimento, implementação e uso de sistemas de inteligência artificial. Por meio dela, busca-se garantir que a tecnologia não apenas atenda a objetivos técnicos ou de eficiência, mas que também respeite e promova a dignidade, os direitos e o bem-estar dos indivíduos e das comunidades.

3 PROTEÇÃO DE DADOS PESSOAIS: O PILAR DA REGULAÇÃO GLOBAL

A proteção de dados pessoais é um dos pilares fundamentais para o desenvolvimento de uma inteligência artificial que respeite os direitos humanos, como preconizam diversos marcos regulatórios globais. Legisladores em diferentes jurisdições, como o Brasil e a União Europeia, têm promulgado normas que visam garantir que o tratamento de dados pessoais seja conduzido de forma adequada e ética. A Lei Geral de Proteção de Dados Pessoais (LGPD) no Brasil e o Regulamento Geral de Proteção de Dados (GDPR) na União Europeia são exemplos que impõem padrões de privacidade e transparência necessários para o desenvolvimento responsável da IA.

A coleta e o uso de dados pessoais sem o consentimento explícito e informado dos titulares levantam graves preocupações quanto à violação de direitos fundamentais, notadamente o direito à privacidade. A utilização oculta ou inadequada de tais dados pelas empresas ou instituições não apenas constitui uma infração aos preceitos legais estabelecidos, mas também configura uma quebra de confiança entre os indivíduos e as entidades que fazem uso desses dados. A ausência de transparência nas práticas de coleta e tratamento de informações pessoais pode desencadear uma crise de confiança, afetando diretamente a legitimidade dessas instituições perante o público.

Quando os titulares de dados não têm ciência de como suas informações são coletadas e utilizadas, ou quando não possuem controle sobre esses dados, o impacto jurídico e social é evidente. A sensação de vulnerabilidade resultante dessa opacidade pode levar à resistência no uso de serviços digitais e a uma demanda crescente por regulamentações mais severas em matéria de proteção de dados pessoais. Dessa forma, a falta generalizada de conformidade com as normativas vigentes não só expõe as instituições a sanções legais, mas também compromete a confiança social para o uso de novas tecnologias no ambiente digital.

Com efeito, recentemente a Autoridade Nacional de Proteção de Dados Pessoais (ANPD) determinou a suspensão cautelar do tratamento de dados pessoais realizado pela empresa Meta para treinamento do seu sistema de IA⁴. A suspensão foi baseada na política de privacidade da Meta, que apresentava indícios de tratamento inadequado de dados pessoais, incluindo falta de informações claras e limitações excessivas ao exercício dos direitos dos titulares, especialmente de crianças e adolescentes.

Segundo o Voto nº 11/2024/DIR-MW/CD condutor, aprovado pelo Conselho Diretor em Circuito Deliberativo, as informações compartilhadas pelos titulares nas plataformas geridas pela Meta geralmente têm como finalidade a interação com amigos, comunidade ou empresas de interesse. Muitos titulares de dados não esperavam que informações compartilhadas para interação social fossem usadas para o treinamento de sistemas de IA. Para que o tratamento de dados com base no legítimo interesse seja válido, ele precisa estar fundamentado em situações concretas e bem definidas, e não em especulações. Assim, a decisão da ANPD ressalta a necessidade de que as expectativas legítimas dos titulares sejam observadas e reforça a importância de mecanismos que permitam o exercício de seus direitos, incluindo a possibilidade de oposição ao tratamento de dados e sua eliminação, quando necessário (ANPD, 2024).

Regulamentos como o GDPR estabelecem requisitos rigorosos para o tratamento de dados pessoais, incluindo princípios como a minimização de dados e a exigência de consentimento explícito. No contexto da IA, o GDPR é especialmente relevante, ao regular, por exemplo, as decisões automatizadas que afetam diretamente os indivíduos, assegurando o direito à revisão humana e à transparência nos processos. No Brasil, a LGPD impõe que sistemas de IA respeitem os princípios de necessidade, finalidade e segurança, protegendo os direitos dos titulares e promovendo a autodeterminação informativa. Tanto a LGPD, quanto o GDPR representam marcos significativos na proteção de direitos dos cidadãos especialmente em um contexto de crescente interação e expansão do uso de IA na vida cotidiana.

Nos Estados Unidos, o “*Blueprint for an AI Bill of Rights*” se apresenta como um documento estratégico desenvolvido pelo governo dos Estados Unidos que delinea princípios e diretrizes para a governança IA (Hine e Floridi, 2023). O objetivo do *Blueprint* é garantir que o desenvolvimento e a implementação de tecnologias de IA respeitem os direitos e a dignidade dos indivíduos, promovendo uma abordagem ética e centrada nas pessoas.

Para tanto, o documento estabelece cinco princípios que devem guiar a regulação da IA, dentre eles a proteção contra práticas abusivas de dados e agência sobre o uso de

⁴ Mais informações disponíveis em: <https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-determina-suspensao-cautelar-do-tratamento-de-dados-pessoais-para-treinamento-da-ia-da-meta>. Acesso em: 23 out.2024.

dados pessoais. De acordo com este princípio, os indivíduos devem ter controle sobre como seus dados são coletados e utilizados, com proteção contra abusos no tratamento de informações pessoais. Isso implica em garantir que as pessoas possam dar consentimento significativo para a coleta de dados e ter o direito de corrigir informações pessoais errôneas

O *Blueprint* é uma tentativa de moldar a governança da IA nos Estados Unidos, promovendo uma abordagem que prioriza os direitos humanos e a ética no uso de tecnologias emergentes. Contudo, não se trata de uma norma com efeitos vinculativos.

As normas de proteção de dados, como a LGPD e o GDPR, representam conquistas notáveis na proteção dos direitos dos cidadãos e no estabelecimento de diretrizes para o uso ético da tecnologia. No entanto, essas regulamentações foram concebidas antes do surgimento da IA generativa e de sistemas avançados de análise preditiva, cujas características e potenciais implicações desafiam as capacidades regulatórias originalmente previstas. Com o avanço dessas tecnologias, torna-se cada vez mais necessário que os marcos regulatórios sejam revisados e atualizados para garantir que continuem adequados à proteção dos direitos individuais.

À medida que a IA evolui, é essencial que os legisladores realizem revisões periódicas das normas para adaptá-las a novos cenários tecnológicos, assegurando que o desenvolvimento da IA ocorra de maneira ética, segura e alinhada com os valores democráticos. Essa atualização contínua é fundamental para promover uma governança centrada nos direitos humanos, colocando a dignidade e a segurança das pessoas no centro das inovações tecnológicas.

4 DESAFIOS E RISCOS: VIESES E DISCRIMINAÇÃO

Apesar dos avanços regulatórios e do foco crescente em uma IA centrada nas pessoas, emergem riscos substanciais, especialmente relacionados a vieses e discriminação algorítmica. Klaus Schwab, fundador do Fórum Econômico Mundial, descreve o período atual como a quarta revolução industrial, marcada pela fusão dos domínios físico, digital e biológico em uma escala sem precedentes. Embora muitas inovações baseadas em IA estejam ainda em estágios iniciais, elas já têm transformado profundamente todos os setores, alterando modelos de negócios, padrões de consumo e modos de comunicação e trabalho (Schwab, 2018, p. 13-14). A quarta revolução industrial, embora traga benefícios significativos, também introduz riscos de que sistemas de IA sem auditoria adequada ou dados representativos reforcem preconceitos e gerem decisões desiguais, impactando negativamente grupos vulneráveis.

A presença de preconceitos nos sistemas de IA é amplamente documentada. O *Blue-*

print for an AI Bill of Rights (2022) ilustra casos em que algoritmos impediram o acesso de pessoas mais velhas a benefícios de saúde ou limitaram o acesso de pessoas negras a transplantes de órgãos ao pressupor um menor risco de doenças renais.

Da mesma forma, no capítulo sobre “Desafios e Riscos: Vieses e Discriminação”, do documentário *Coded Bias* (2020), dirigido por Shalini Kantayya, há um exemplo contundente sobre os perigos dos vieses presentes em tecnologias de inteligência artificial. O episódio investiga como algoritmos de reconhecimento facial, amplamente utilizados em diversas aplicações, perpetuam vieses racistas e machistas. A pesquisadora Joy Buolamwini, do Instituto de Tecnologia de Massachusetts (MIT), é uma das protagonistas da obra e nela destaca as falhas e injustiças que descobriu ao conduzir uma pesquisa no MIT Media Lab.

Em uma de suas experiências, Buolamwini relata como, ao posicionar seu rosto em frente a um dispositivo de IA, o sistema de reconhecimento facial falhou em identificá-la. Apenas quando cobriu seu rosto com uma máscara branca o dispositivo foi capaz de “enxergá-la”. Esse incidente revelou o viés racial embutido nos algoritmos, que tendem a ser mais precisos no reconhecimento de rostos brancos, enquanto têm dificuldades com rostos de pessoas negras ou de mulheres.

A pesquisa de Buolamwini expôs o problema crítico de que os algoritmos de IA, muitas vezes desenvolvidos por equipes com pouca ou nenhuma diversidade, refletem os preconceitos da sociedade em que foram criados, amplificando desigualdades existentes e injustiças estruturais. Tendo em vista que os homens ainda dominam os campos da Ciência, Tecnologia, Engenharia e Matemática, apelidadas de STEM, é preciso que mais iniciativas e políticas públicas sejam criadas para evitar que o aumento da demanda por habilidades técnicas não venha a exacerbar desigualdades de gênero.

Para combater esses efeitos, é necessário que haja um foco contínuo em transparência e explicabilidade, capazes de garantir que os algoritmos sejam auditados por equipes interdisciplinares que incluam especialistas em ética, direitos humanos e proteção de dados. Os usuários devem ser capazes de entender os processos que levam a uma decisão. Além de saber como as decisões são feitas, os usuários devem ser capazes de compreender as razões por trás dessas decisões. Isso é especialmente importante em contextos em que as decisões da IA podem ter um impacto significativo na vida das pessoas, como em áreas de saúde, finanças ou justiça. (Comissão Europeia, 2018, p.17).

Embora possa parecer uma solução simples remover informações sensíveis, como gênero, dos conjuntos de dados para evitar discriminação, essa abordagem é complexa e apresenta desafios. Proxies e correlações subjacentes podem perpetuar preconceitos sociais, mesmo sem dados explícitos de gênero. Por exemplo, atributos como nível educacional ou comportamentos correlacionados ao gênero podem ser utilizados de forma

indireta, levando a modelos de aprendizado de máquina que ainda refletem desigualdades de gênero (Barocas; Hardt; Narayanan, 2024, p. 21). Mesmo sem a informação explícita de gênero, outros atributos podem atuar como substitutos, refletindo desigualdades sociais. Por exemplo, períodos de ausência no histórico profissional, que podem estar relacionados à licença-maternidade ou determinados padrões de palavras usadas em descrições de trabalho, acabam servindo como proxies para o gênero. Assim, o sistema pode priorizar candidatos cujos perfis se assemelham a padrões historicamente masculinos, enquanto rejeita ou desvaloriza candidaturas de mulheres, perpetuando um viés implícito.

Ou seja, para garantir decisões justas e equitativas, é fundamental considerar esses fatores e adotar uma abordagem crítica e abrangente na análise de dados e no desenvolvimento de modelos.

Uma abordagem centrada no ser humano visa, portanto, assegurar que os direitos e a dignidade dos indivíduos sejam respeitados. Isso é fundamental para evitar abusos, discriminação e violação de privacidade que podem surgir do uso inadequado de tecnologias baseadas em IA.

5 IA E DEMOCRACIA: IMPACTOS NAS ELEIÇÕES E NO DEBATE PÚBLICO

A democracia tem por objetivo dar maior liberdade a todos e para todos, na medida em que trabalha a favor da justiça e propicia, em sua aplicação, a construção do estado Democrático de Direito.

A partir da relação indissociável entre ética, política e direito, as normas jurídicas que tratam do tema da IA devem refletir e ser instrumentos para a concretização dos valores ético-políticos existentes em nossa cultura.

Em uma sociedade plural, multicultural e complexa, composta por indivíduos com diferentes projetos de vida, habilidades e oportunidades, John Rawls, em sua obra “Uma Teoria da Justiça” (1971), defende que a justiça como equidade deve ser o princípio que governa as estruturas de uma sociedade democrática. Segundo Rawls, a ética é a base para a concepção de justiça, sendo operacionalizada pela política, cujo papel é construir e manter estruturas sociais justas. O direito, por sua vez, institui e consolida essas estruturas, garantindo que os princípios de justiça e equidade sejam aplicados de forma consistente. Assim, a ética define os valores que orientam a política, enquanto o direito assegura a implementação desses valores por meio de um sistema normativo.

Esses ensinamentos são particularmente importantes no contexto atual, onde a IA exerce uma influência crescente sobre as democracias modernas. O uso de IA em

campanhas eleitorais, por exemplo, tem levantado preocupações sobre a manipulação de dados pessoais e a criação de estratégias de *microtargeting*, que podem distorcer o debate público e comprometer a liberdade de escolha dos eleitores.

O caso emblemático da *Cambridge Analytica*, que envolveu o uso indevido de dados pessoais para influenciar eleitores, ilustra os perigos de uma IA que não respeita os princípios de transparência e proteção de dados. Para garantir a integridade do processo democrático, é essencial que haja uma supervisão mais rígida sobre o uso de IA nas campanhas e que os eleitores sejam devidamente informados sobre o uso de seus dados. Normas centradas no ser humano, que promovam a transparência e o respeito aos direitos fundamentais, são necessárias para que o uso de IA em contextos eleitorais não subverta o processo democrático através do abuso do poder de comunicação e da redução do poder de escolha dos eleitores.

Além disso, a democracia se sustenta sobre proteções constitucionais explícitas, como a liberdade de expressão e de imprensa, que também limitam o poder estatal e devem ser protegidas dos efeitos prejudiciais da IA nas redes sociais. Os algoritmos que amplificam certas vozes e mensagens contribuem para a criação de “câmaras de eco”, um fenômeno em que os indivíduos são expostos predominantemente a conteúdos que reforçam suas crenças pré-existentes. Esse fenômeno é exacerbado pela IA, que utiliza dados comportamentais para fornecer conteúdo altamente personalizado, o que intensifica a polarização política e, conseqüentemente, degrada o debate público.

Cass Sunstein, em *Republic: Divided Democracy in the Age of Social Media* (2017), explora como as “câmaras de eco” nas redes sociais resultam em uma esfera pública fragmentada. Sunstein aponta que, ao se isolarem em bolhas ideológicas, os cidadãos perdem a oportunidade de serem desafiados por visões opostas, o que enfraquece a vitalidade do debate democrático e ameaça o diálogo aberto. Como a democracia prospera com a troca de ideias e o confronto de opiniões divergentes, a IA pode, ao restringir essa diversidade de perspectivas, contribuir para a erosão da própria essência democrática.

Nota-se que se por um lado a IA oferece inúmeras oportunidades para otimizar processos e promover a eficiência, por outro, o seu uso na democracia deve ser cuidadosamente regulamentado para evitar a erosão das liberdades individuais e garantir que a equidade e a justiça sejam preservadas no coração das sociedades democráticas. Em última análise, o uso da IA deve fortalecer os princípios fundamentais da democracia e não ser uma ferramenta que os enfraqueça, garantindo que os processos eleitorais e o debate público permaneçam genuinamente inclusivos, justos e representativos da diversidade de ideias que caracteriza uma sociedade democrática.

6 CAMINHOS PARA UMA IA ÉTICA E CENTRADA EM PESSOAS

Os mais pessimistas costumam associar o avanço da inteligência artificial a impactos negativos para os seres humanos, como a redução de oportunidades de emprego e o aumento de problemas de saúde mental, como a depressão. Além disso, surgem preocupações quanto a potenciais violações aos direitos da personalidade devido ao tratamento intensivo de dados que esses sistemas demandam. No entanto, é necessário reavaliar essas percepções predefinidas, pois cabe ao próprio ser humano e à sociedade moldar o desenvolvimento e as aplicações da tecnologia, a fim de assegurar que a IA atue de forma ética e beneficie todos os setores sociais (Soares e Medina, 2020).

Para alcançar esta ética, é essencial distinguir “*fairness*” (justiça) e “*equity*” (equidade) para garantir que as tecnologias de IA não apenas funcionem de maneira imparcial, mas também considerem as circunstâncias e necessidades individuais dos diferentes grupos afetados. De acordo com Barocas, Hardt e Narayanan (2024, p.24), os algoritmos de aprendizado de máquina devem ser projetados levando em consideração as limitações e oportunidades no que diz respeito à justiça e equidade. A justiça refere-se ao tratamento igual para todos os usuários, com base em normas universais. A equidade, por sua vez, reconhece que diferentes indivíduos ou grupos possuem necessidades distintas e que, para alcançar resultados verdadeiramente justos, a IA deve adaptar-se para oferecer recursos adequados a cada contexto. Em algoritmos de aprendizado de máquina, essa distinção é sutil, mas crucial. Um sistema de IA justo pode garantir que todos os grupos recebam o mesmo tratamento em termos de decisões ou resultados, enquanto um sistema equitativo busca mitigar desigualdades, ajustando-se às particularidades de cada indivíduo ou grupo.

No contexto de uma IA centrada em pessoas, essa distinção é essencial para assegurar que as tecnologias não apenas sejam imparciais, mas também considerem as circunstâncias específicas dos grupos que afetam. Barocas, Hardt e Narayanan (2024) argumentam que os algoritmos devem ser projetados com uma abordagem cuidadosa da justiça e da equidade. Para uma IA realmente focada em pessoas, é preciso identificar quando as disparidades nos sistemas de aprendizado de máquina são prejudiciais ou injustificadas, ou seja, se refletem preconceitos históricos e sociais. Esse processo envolve uma análise criteriosa dos dados usados para treinar os modelos e das decisões geradas por esses modelos.

Ao identificar essas disparidades problemáticas, o próximo passo é desenvolver intervenções que possam mitigá-las. O objetivo final é garantir que as decisões automatizadas sejam justas e equitativas, evitando que grupos marginalizados sejam desproporcionalmente afetados por sistemas supostamente neutros. Isso exige um compromisso

com a justiça social e a responsabilidade ética na concepção e implementação dessas tecnologias (Barocas; Hardt; Narayanan, 2024, p. 31).

Outro caminho não excludente é uma abordagem multidisciplinar, que inclua a participação de especialistas em direitos humanos, juristas, engenheiros e cientistas de dados. Alguns caminhos possíveis incluem auditorias contínuas dos algoritmos para identificar e mitigar vieses, promover a transparência em todos os processos de tratamento de dados, garantindo que os usuários saibam como seus dados estão sendo utilizados, implementar mecanismos de revisão humana nas decisões automatizadas e investir imediatamente em educação e conscientização sobre os direitos dos titulares de dados e as implicações do uso de IA.

Além de iniciativas voluntárias, o estabelecimento de marcos regulatórios é essencial para guiar o desenvolvimento e a implementação de tecnologias de inteligência artificial. Esses marcos devem incluir diretrizes claras sobre a ética no uso da IA, assegurando que os direitos humanos sejam priorizados em todas as fases do ciclo de vida da tecnologia. Regulamentações que exijam a realização de avaliações de impacto sobre direitos humanos antes da implementação de sistemas de IA podem ajudar a identificar e mitigar riscos potenciais, promovendo um ambiente mais seguro e responsável para o uso dessas ferramentas.

Um exemplo de avanço nessa área é a proposta do Projeto de Lei nº 2338, de 2022, que tramita no Senado Federal e é fortemente inspirada no *AI Act* da União Europeia. Esse projeto visa estabelecer um quadro regulatório específico para a IA no Brasil, classificando os sistemas em diferentes níveis de risco (baixo, médio e alto) e exigindo que sistemas de alto risco, como aqueles usados em decisões automatizadas em processos de recrutamento, de acesso à instituições de ensino e saúde, maiores cautelas como auditorias rigorosas de conformidade e participação social em processos de governança. Dessa forma, busca-se garantir que o uso dessas tecnologias ocorra com segurança e responsabilidade, preservando os direitos e a dignidade dos indivíduos.

7 CONCLUSÃO

O desenvolvimento de uma inteligência artificial focada em pessoas representa um desafio multifacetado, que requer um equilíbrio cuidadoso entre a inovação tecnológica e a proteção dos direitos fundamentais. Como abordado ao longo deste artigo, a IA possui um impacto social significativo que, se não for gerido com responsabilidade, pode perpetuar vieses, gerar discriminação e comprometer a integridade dos processos democráticos. Nesse contexto, a proteção de dados pessoais

desponta como um dos pilares centrais para a construção de uma IA ética, respeitosa dos direitos humanos e orientada para o bem-estar coletivo.

A crescente regulamentação da IA, como demonstram o GDPR, a LGPD, o AI Act e o Projeto de Lei 2338 no Brasil, é uma resposta essencial aos desafios éticos e jurídicos que essas tecnologias impõem. Esses marcos regulatórios estabelecem diretrizes que buscam assegurar que o uso da IA esteja alinhado com os princípios de liberdade, privacidade e dignidade humana, promovendo uma governança transparente e centrada nos direitos fundamentais. No entanto, como visto, essas normas devem acompanhar o ritmo acelerado das inovações, especialmente diante da IA generativa e dos sistemas automatizados que influenciam esferas vitais como o mercado de trabalho, a saúde, a educação e o debate público.

A aplicação da IA em contextos sensíveis, como os processos eleitorais, evidencia os riscos de manipulação e possível subversão da democracia. A transparência, justiça e equidade - conceitos essenciais para uma IA ética - ajudam a prevenir abusos de poder, assegurando que as tecnologias não enfraqueçam a liberdade de escolha e a confiança pública. Nesse contexto, a distinção entre “*fairness*” (justiça) e “*equity*” (equidade) é indispensável para que a IA não apenas trate todos de forma imparcial, mas também considere as necessidades e circunstâncias específicas dos grupos afetados, promovendo uma verdadeira igualdade de oportunidades.

Para construir uma IA ética e centrada em pessoas, é necessário um esforço multidisciplinar que envolva engenheiros, juristas, especialistas em ética, direitos humanos e cientistas de dados. Isso inclui a realização de auditorias contínuas, a implementação de mecanismos de revisão humana e a promoção de uma maior conscientização sobre os direitos de privacidade e as implicações éticas do uso de IA. Esses esforços devem ser acompanhados de marcos regulatórios robustos que exijam avaliações de impacto, principalmente em sistemas de alto risco, como os que envolvem a administração da justiça, a segurança pública e a assistência à saúde.

Em última análise, uma IA focada em pessoas não é apenas uma meta desejável, mas uma necessidade para garantir que o progresso tecnológico não ocorra às custas dos direitos e liberdades individuais. A construção de uma IA ética e inclusiva, que respeite a dignidade e as legítimas expectativas das pessoas, é o caminho para assegurar que essa tecnologia contribua positivamente para a sociedade. Ao reconhecer e mitigar os riscos, a sociedade pode maximizar os benefícios da IA, promovendo uma evolução tecnológica que fortaleça, ao invés de fragilizar, os valores democráticos e os direitos humanos.

PEOPLE-CENTERED ARTIFICIAL INTELLIGENCE: DATA PROTECTION AND FUNDAMENTAL RIGHTS IN THE GLOBAL CONTEXT

ABSTRACT

This exploratory article investigates the importance of ethical principles for the development and use of artificial intelligence (AI) systems, focusing on trust and respect for fundamental rights, particularly in data protection. Using a deductive method and literature review, the study presents key theoretical frameworks on human-centered AI. It addresses ethical challenges and inherent risks in AI, suggesting pathways for an ethical, human-centered AI that is essential for democracy and public debate. The article concludes by highlighting a new perspective to address contemporary challenges in data protection and privacy.

Keywords: artificial intelligence; human rights; ethics; equity; fairness.

REFERÊNCIAS

AGÊNCIA DOS DIREITOS FUNDAMENTAIS DA UNIÃO EUROPEIA. **Inteligência artificial e direitos fundamentais:** síntese. Luxemburgo: Publicações do Gabinete da União Europeia, 2021. Disponível em: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2021-artificial-intelligence-summary_pt.pdf. Acesso em: 19 out. 2024. ISBN 978-92-9461-234-2. DOI: 10.2811/990836.

ARÃO, Cristian. Por trás da inteligência artificial: uma análise das bases epistemológicas do aprendizado de máquina. **Trans/Form/Ação**, São Paulo, v. 47, n. 3, 2024. Disponível em: <https://doi.org/10.1590/0101-3173.2024.v47.n3.e02400163>. Acesso em: 14 out. 2024.

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. **Voto nº 11/2024/DIR-MW/CD: Medida preventiva para evitar dano grave e irreparável ou de difícil reparação.** Brasília: ANPD, 2024. Disponível em: <https://www.gov.br/anpd/pt-br/as-suntos/noticias/anpd-determina-suspensao-cautelar-do-tratamento-de-dados-pessoais-para-treinamento-da-ia-da-meta>. Acesso em: 15 out. 2024.

BAROCAS, Solon; HARDT, Moritz; NARAYANAN, Arvind. **Fairness and Machine Learning: Limitations and Opportunities.** Disponível em: <https://fairmlbook.org/>. Acesso em: 15 out. 2024.

BURLE, Caroline. CORTIZ, Diogo. **Mapeamento de princípios de inteligência artificial.** São Paulo: Comitê Gestor da Internet no Brasil, 2019. Disponível em: <https://>

www.nic.br/publicacao/mapeamento-de-principios-de-inteligencia-artificial/. Acesso em: 21 out. 2024.

COMISSÃO EUROPEIA. **Ethics guidelines for trustworthy AI**. 2018. Disponível em: <https://digital-strategy.ec.europa.eu/pt/library/ethics-guidelines-trustworthy-ai>. Acesso em: 24 out.2024.

COMISSÃO EUROPEIA. **Livro Branco sobre a inteligência artificial — Uma abordagem europeia virada para a excelência e a confiança**. Bruxelas: Comissão Europeia, 2020. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:52020DC0065>. Acesso em: 19 out. 2024.

EUROPEAN GROUP ON ETHICS IN SCIENCE AND NEW TECHNOLOGIES. **Statement on artificial intelligence, robotics and ‘autonomous’ systems**. Luxembourg: Publications Office of the European Union, 2018. Disponível em: <https://doi.org/10.2777/531856>. Acesso em: 16 out.2024.

FLORIDI, Luciano.; COWLS, Josh. A unified framework of five principles for AI in society. **Harvard Data Science Review**, v. 1, n. 1, 2019. Disponível em: <https://doi.org/10.1162/99608f92.8cd550d1>. Acesso em: 14 out. 2024.

FUTURE OF LIFE INSTITUTE. **Asilomar AI Principles**. Disponível em: <https://futureoflife.org/2017/08/11/ai-principles/>. Acesso em: 08 out. 2024.

HIGH-LEVEL ADVISORY BODY ON ARTIFICIAL INTELLIGENCE. **Governing AI for humanity: final report**. Setembro de 2024. Disponível em: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2021-artificial-intelligence-summary_pt.pdf. Acesso em: 19 out.2024.

HINE, Emmie; FLORIDI, Luciano. **The Blueprint for an AI Bill of Rights: In Search of Enaction, at Risk of Inaction**. Minds and Machines, 2023. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4279449. Acesso em: 19 out.2024.

KANTAYYA, Shalini. **Coded Bias**. Direção de Shalini Kantayya. Produção de 7th Empire Media, 2020. Disponível em: <https://www.netflix.com>. Acesso em: 14 out. 2024.

ORGANIZAÇÃO PARA A COOPERAÇÃO E DESENVOLVIMENTO ECONÔMICO (OCDE). **Principles on AI**. Disponível em: <https://www.oecd.org/going-digital/ai/principles/>. Acesso em: 21 out. 2024.

PARLAMENTO EUROPEU. **Resolução do Parlamento Europeu de 20 de outubro de 2020 que contém recomendações à Comissão sobre o regime relativo aos aspectos éticos da inteligência artificial, da robótica e das tecnologias conexas**. 2020/2012(INL). Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_PT.html. Acesso em: 16 out. 2024.

PINKER, Steven. **O novo iluminismo: em defesa da razão, da ciência e do humanismo**. Tradução de Laura Teixeira Mota e Pedro Maia Soares. - 1a ed.- São Paulo:

Companhia das Letras, 2018.

RAWLS, J. **Uma teoria da justiça**. 3. ed. São Paulo: Martins Fontes, 2008.

SCHWAB, Klaus. **A quarta revolução industrial**. São Paulo: Edipro, 2018.

SOARES, Marcelo Negri; MEDINA, Valéria Julião Silva. Os impactos da inteligência artificial no direito da personalidade do jurisdicionado. **Revista de Direito Brasileira, Florianópolis**, v. 26, n. 10, p. 277-291, mai./ago. 2020.

SUNSTEIN, Cass R. **Republic: Divided Democracy in the Age of Social Media**. Princeton: Princeton University Press, 2017.