INTELIGÊNCIA ARTIFICIAL – UMA QUESTÃO ÉTICA

Marcelo Moreno Gomes Lisboa[1]

RESUMO

A tecnologia da inteligência artificial está mudando o mundo em vários aspectos, na saúde por exemplo, com melhores diagnósticos e controle de doenças, nas cirurgias. Nas indústrias pode gerar desemprego em algumas funções, mas, criando outras, promovendo um aumento da produtividade, corte de custos e melhoria na qualidade dos produtos. Enfim, a IA, vem causando impacto em todos os campos da nossa vida, mudando o modo como trabalhamos, nos comunicamos e até mesmo descansamos e nos divertimos. Justamente por isso é que se faz necessário tratá-la além do aspecto tecnológico, como uma questão ética. Além disso, como todo produto humano que é, a IA tem exposto a humanidade em situações de tensão ética muito complexas e delicadas em tanto em razão de suas vantagens como também em razão de suas imperfeições. Os erros cometidos por inteligências artificiais em todo o mundo têm gerado uma série de problemas éticos e jurídicos que demandam novas e velhas soluções. Esse é basicamente o escopo deste artigo; provocar uma reflexão sobre os dilemas éticos da IA e suas soluções.

Palavras-chave: Inteligência Artificial; Ética; Direito.

ABSTRACT

Artificial intelligence (AI) is transforming the world across multiple domains. In healthcare, for instance, it contributes to more accurate diagnoses, improved disease management, and enhanced surgical procedures. In industrial settings, while AI may lead to job displacement in certain roles, it simultaneously fosters the creation of new functions, boosts productivity, reduces operational costs, and elevates product quality. AI's influence extends to virtually every facet of human life, reshaping how we work, communicate, rest, and engage in leisure. Precisely because of its pervasive impact, it is essential to approach AI not solely as a technological advancement, but also as an ethical concern. As a human creation, AI inevitably reflects human values and imperfections, often placing society

[1] Professor de Direito do Trabalho e de Direito Empresarial da Faculdade Asa de Brumadinho e Mestre em Direito pela Pontifícia Universidade Católica de Minas Gerais (2006).

in ethically complex and delicate situations—arising both from its benefits and its limitations. Errors committed by AI systems worldwide have triggered a host of ethical and legal challenges, demanding both innovative and traditional solutions. This article aims to provoke thoughtful consideration of these dilemmas and explore potential pathways toward responsible and ethical integration of AI into society.

Keywords: Artificial Intelligence; Ethics; Law.

INTRODUÇÃO

Há tempos atrás, a inteligência artificial (IA) foi retratada como uma tecnologia do futuro, sobretudo, em filmes de ficção científica. No presente, os seus usos estão sendo viabilizados em nosso cotidiano. A plataforma *Waze*, por exemplo, ao indicar o melhor caminho para o destino, adapta-se às mudanças de trânsito e ao trajeto de forma autônoma, sem que o usuário precise inserir informações ou programar qualquer detalhe. Esta utiliza IA para fornecer essas direções de maneira eficiente e em tempo real. A essa aplicação, podemos incluir o *Google Maps*, a *Alexa* e o *ChatGPT*, que ilustram como a IA está integrada no nosso dia a dia, tornando-se uma ferramenta essencial para diversas atividades (Duran, 2020). Diante disso, é cada vez mais necessário compreender o funcionamento e as utilidades dessa tecnologia.

Na saúde, a IA está sendo apropriada no diagnóstico de doenças, em exames de imagem e até nas cirurgias robóticas pelos médicos. Na indústria e na agricultura, vem facilitado a gestão e a tomada de decisões, além de impulsionar o desenvolvimento de sistemas de robôs e veículos autônomos. No setor militar, os drones de combate, reconhecimento e vigilância não tripuláveis, presentes nas atuais guerras, são o símbolo da aplicação da tecnologia. Em suma, a implementação da IA tem transformado esses setores, proporcionando maior precisão, agilidade e segurança (Heath, 2019).

No campo da educação, a IA tem contribuído significativamente, oferecendo recursos como a correção automática de provas, tutores virtuais e a criação de conteúdo personalizados. Nos bancos, é utilizada na análise de créditos, otimizando o processo de avaliação de risco. No setor jurídico, ela tem sido aplicada em processos licitatórios, com supervisão dos tribunais de contas e prefeituras. Aliás, é verificada na triagem de processos no Judiciário, utilizando modelos de linguagem para interpretar textos jurídicos, gerar resumos automáticos e classificar informações, facilitando a tomada de decisão (Santos; Almeida, 2021). Diante desses exemplos, nota-se que a aplicação da IA na sociedade tem se expandido rapidamente, com uma tendência de crescimento contínuo e diversificação

de suas funcionalidades. Por isso, torna-se importante a abertura de um "parêntesis" para destrinchar brevemente – a partir da bibliografia – a tecnologia.

"DESTRINCHANDO" A IA

De maneira breve, a IA abrangeria sistemas computacionais projetados para simular algumas capacidades da inteligência humana, como o aprendizado, o raciocínio e a resolução de problemas. Ela pode ser dividida limitada e generativa. A primeira é instruída para realizar tarefas específicas, como assistentes pessoais virtuais (como *Siri*, *Alexa*, *ChatGPT*) ou sistemas de recomendação. Essas aplicações são eficientes dentro de suas funções pré-estabelecidas. Mas, não têm capacidade de adaptação a novas tarefas além do seu escopo de programação. Em contraste, a segunda ainda está na fase de testes e tem o objetivo de replicar a inteligência humana em um nível mais amplo. Quando consolidada, ela seria capaz de aprender, raciocinar e se adaptar de forma independente, comparáveis às humanas (Russell; Norving, 2017; Duran, 2020).

Cabe aos algoritmos a função em colaborar na assimilação da inteligência humana na IA. Eles consistem em uma série de instruções matemáticas e lógicas que permitem que o sistema execute suas funções, como o reconhecimento de padrões, a análise de dados e a tomada de decisões. Essas instruções podem ser ajustadas e treinadas de forma que a IA aprenda com os dados que recebe. Quando interagimos com essa última, fornecemos um prompt, que é uma solicitação ou tarefa que desejamos que seja executada. Este pode ser uma pergunta, um comando ou qualquer outra instrução que oriente a tecnologia a realizar uma ação específica. Exemplos disso encontramos no caso de perguntar algo ao *ChatGPT* ou pedir para a *Siri* tocar uma música (Cormen *et. al*, 2011; Russell; Norving, 2017; Duran, 2020; Córdova, 2025).

No campo da IA, o aprendizado de máquina (*machine learning*) é uma área essencial que permite que as máquinas compreendam e interpretem dados humanos sem serem explicitamente programadas para isso. O aprendizado de máquina envolve o uso de algoritmos que aprendem a partir de dados, ajustando-se automaticamente à medida que recebem mais informações. Dentro do aprendizado de máquina, existem várias abordagens, como o aprendizado supervisionado, onde um ser humano organiza e insere dados rotulados para que o sistema aprenda a partir de exemplos. Por exemplo, um modelo pode ser treinado para reconhecer imagens de gatos a partir de um conjunto de fotos rotuladas. Outro tipo é o aprendizado não supervisionado. Nesta modalidade, a IA identifica padrões e correlações nos dados sem a intervenção humana, buscando agrupamentos e tendências por conta própria (Russell; Norving, 2017; Almeida; Santos, 2018; Duran, 2020).

Outra forma de aprendizado no contexto da IA é o aprendizado semi-supervisionado, agregando dados rotulados e não rotulados para treinar modelos de aprendizado de máquina. Esse tipo de abordagem é utilizado quando há uma grande quantidade de dados disponíveis. Todavia, apenas uma parte é rotulada. O modelo mobiliza a parte rotulada para aprender e a parte não rotulada para encontrar padrões e melhorar seu desempenho. Ademais, o aprendizado por reforço é um método em que a IA aprende a partir de tentativas e erros, recebendo recompensas ou punições com base em suas ações. Essa tática é comumente usada em jogos ou sistemas que exigem decisões sequenciais, como veículos e robôs autônomos (Russell; Norving, 2017; Almeida; Santos, 2018; Duran, 2020).

Inspiradas no funcionamento do cérebro humano, as redes neurais são outro componente essencial da IA. Elas consistem em camadas de "neurônios artificiais" que processam informações de maneira interconectada, permitindo a identificação de padrões complexos. Essas são particularmente eficazes em tarefas como classificação, reconhecimento de padrões, previsão e tomada de decisão. O uso em reconhecimento de voz ou reconhecimento facial, em que o sistema é treinado para identificar características específicas a partir de grandes quantidades de dados, é um exemplo de seu emprego. (Russell; Norving, 2017; Duran, 2020).

Além disso, o processamento de linguagem natural (PLN) é uma subárea da IA que se concentra em permitir que as máquinas compreendam e gerem linguagem humana de maneira eficaz. Isso envolve a interpretação de textos ou falas em idiomas, como o português ou o inglês, e a conversação de forma fluida e coerente. Por outro lado, a visão computacional é uma área da tecnologia que permite que os computadores interpretem imagens e vídeos, o que é fundamental para equipamentos como câmeras de segurança, carros autônomos e reconhecimento de imagens em geral. Enfim, a IA pode combinar várias dessas tecnologias, como em robôs que utilizam redes neurais, PLN e visão computacional para realizar tarefas complexas de interação, decisão e aprendizado autônomo (Russell; Norving, 2017; Almeida; Santos, 2018; Duran, 2020).

CASOS CONCRETOS DE PROBLEMAS ÉTICOS NA UTILIZAÇÃO DA IA

Como qualquer produto humano, a IA não está isenta de falhas e imperfeições. Ela está sujeita a erros de programação e falhas operacionais, que podem ocorrer devido a falhas nos algoritmos ou nas bases de dados usadas para treinar os sistemas. Esses erros podem comprometer a eficácia das funções da tecnologia e, em alguns casos, resultar em resultados inesperados ou indesejáveis (Duran, 2020; Córdova, 2025). Além disso, por não possuir sentimentos ou empatia, ela depende de um treinamento cuidadoso para garantir que execute suas tarefas de acordo com as normas éticas

e jurídicas estabelecidas. O desafio de fazer com que a tecnologia atenda a esses requisitos éticos é conhecido como problema do viés (Córdova, 2025).

O viés em sistemas de IA é um problema importante, pois muitas vezes esses acabam reproduzindo preconceitos sociais, como discriminação de gênero, racial, xenofobia, entre outros. Isso ocorre porque, ao serem treinados com dados históricos, que podem carregar esses preconceitos, a tecnologia acaba replicando padrões pré-existentes. Esses vieses podem levar a decisões injustas e à perpetuação de estereótipos negativos. Situações disso são levantados por Paulo Córdova (2025):

> Um exemplo conhecido e muito representativo sobre esse problema é o caso do TAY, uma *chatbot* desenvolvida em 2016 pela *Microsoft*, como uma experiência em inteligência artificial para entender melhor como os jovens se comunicam on-line. Ela foi programada para simular um adolescente estadunidense, utilizando linguagem coloquial e memes populares. O objetivo era que ela interagisse de maneira natural com os usuários da rede que na época era chamada de *Twitter*, aprendendo com cada interação para fornecer respostas mais precisas e envolventes. O problema surgiu devido à maneira como ela foi projetada para aprender. Sem filtros adequados para moderar o conteúdo que absorvia, TAY começou a repetir e amplificar as mensagens tóxicas e ofensivas que recebia dos usuários (Córdova, 2025, p. 134).

Além desse caso, outros exemplos de discriminação ocasionados pelo uso da IA ocorreram em diferentes partes do mundo. No Reino Unido, sistemas utilizados para concessão de créditos imobiliários limitavam ou negavam crédito a pessoas de áreas marginalizadas, sem justificativa plausível além da localização. Na Itália, uma seguradora passou a ajustar os preços dos seguros de carros com base na nacionalidade do comprador, cobrando mais caro de pessoas originárias de países considerados subdesenvolvidos. Esses casos refletem como a IA, sem a devida atenção a questões éticas, pode reforçar práticas discriminatórias que afetam as oportunidades das pessoas com base em sua origem ou classe social (Córdova, 2025).

Ainda segundo Córdova (2025), outros exemplos de falhas éticas em IA incluem o Google e a Nikon. Em 2015, o Google lançou um aplicativo de fotos que, devido a falhas no sistema de reconhecimento de imagens, acabou identificando erroneamente pessoas negras como gorilas. Da mesma forma, a Nikon, fabricante de câmeras fotográficas, também enfrentou críticas quando suas câmeras digitais começaram a classificar pessoas de origem asiática como "sempre piscando" nas fotos. Esses erros expõem a fragilidade dos sistemas de IA, que, sem a devida programação e testes, podem gerar resultados discriminatórios ou imprecisos.

Outro exemplo impactante mencionado por Córdova (2025) ocorreu nos Estados Unidos: a atuação de um software que colaborava a decidir sobre a prisão preventiva de réus com base na probabilidade de cometerem crimes no futuro. O sistema falhou ao associar a probabilidade de reincidência ao perfil racial, sugerindo que pessoas negras tinham mais chances de cometer crimes do que pessoas brancas. Esse caso exemplifica como a IA pode reproduzir preconceitos existentes na sociedade. Além disso, empresas que usavam a tecnologia para recrutamento, como no caso do *ChatGPT*, demonstraram viés de gênero ao selecionar candidatos para cargos de liderança, discriminando mulheres com base na suposição de que elas seriam menos competentes do que os homens. Esses incidentes deixam claro que, para evitar problemas éticos, a IA deve ser constantemente monitorada e aprimorada para garantir que atenda aos padrões de justiça e igualdade.

A QUESTÃO ÉTICA NA IA: APONTAMENTOS FILOSÓFICOS E TECNOLÓGICOS

A ética da virtude, originada na Grécia Antiga, é essencialmente associada a Aristóteles, embora tenha sido introduzida pelo seu mestre, Platão. Segundo essa abordagem, a pessoa molda seu caráter com base em virtudes que são cultivadas por meio de hábitos ao longo de sua vida. Essas virtudes são baseadas nos valores morais dos gregos, sendo adquiridas pela repetição e prática constante, até que se tornem parte integrante da personalidade do indivíduo. A principal virtude é a justiça, pois considera a existência do outro, pois ela não existiria para si mesmo. Aristóteles acredita que, por meio da prática das virtudes, estas se tornam naturais no comportamento da pessoa, que passa a agir de acordo com os princípios éticos de forma espontânea. Porém, para que a virtude seja aplicada adequadamente nas várias situações cotidianas, é necessária a *fronezis*, ou sabedoria prática, que orienta a pessoa a aplicar a virtude com equilíbrio e discernimento. Infere-se que as virtudes devem ser aplicadas de maneira equilibrada, como no caso da coragem, que deve estar entre a temeridade excessiva e o medo paralisante (Aristóteles, 2015).

Uma abordagem contemporânea da perspectiva aristotélica é a ética exemplarista, proposta pela filósofa estadunidense Linda Zegzebski (2004). A pensadora sugere que a ação moral deve ser baseada no exemplo de modelos exemplares de boa moralidade. Em vez de se concentrar apenas em regras ou normas abstratas, a ética exemplarista foca em observar e emular as ações de indivíduos virtuosos como forma de aprender a agir moralmente. Para Zegzebski (2004), a moralidade é vista como uma prática de imitação de exemplos de vida virtuosa. Tal acepção a conecta diretamente com a percepção aristotélica de que o caráter moral é formado por meio da prática e do hábito. Essa abordagem oferece uma maneira moderna de aplicar os princípios da ética da virtude, enfatizando o papel do exemplo na formação do caráter moral dos indivíduos (Zegzebski, 2004).

Surgida no século XVIII, a ética utilitarista, capitaneada pelos filósofos Jeremy Bentham e John Stuart Mill, tem como ideia central que o agir moral deve ser pensado como aquele que tenha maior utilidade ou maior impacto na felicidade geral. Sintetizando ainda mais, a ação moral boa é aquela cujas consequências impactam mais positivamente na felicidade do maior número de pessoas possível. O utilitarismo é baseado na premissa de que a moralidade deve ser orientada para o bem-estar coletivo, buscando maximizar a felicidade e minimizar o sofrimento, com a análise das consequências das ações sendo o principal critério para determinar sua moralidade (Bentham; Mill, 1984).

No mesmo século, a ética formulada por Immanuel Kant busca estabelecer uma base racional, universal e sólida para a ação moral. Para isso, o filósofo formulou o conceito do imperativo categórico, que é uma forma da razão pensar a conduta moral, estabelecendo que as ações devem ser guiadas por máximas que possam ser aplicadas universalmente. Em outras palavras, as ações devem ser realizadas com a intenção de que possam ser seguidas por todos, sem exceções, em qualquer situação, garantindo que a moralidade seja universal e imparcial (Kant, 2003).

Outro conceito essencial na ética kantiana é o de fins, que pode ser divido em empíricos, que são contingentes e ligados ao desejo ou interesse pessoal, e os absolutos, que são permanentes e relacionados à razão prática pura, isto é, ao agir moral. Em *Metafísica dos Costumes*, ressalta-se que o agir moral deve sempre respeitar a humanidade como um fim em si mesmo, e nunca como um meio para alcançar outro objetivo. Isso significa que o ser humano, como sujeito racional, deve ser tratado com dignidade e nunca instrumentalizado para alcançar um fim utilitário. A moralidade não estaria relacionada a resultados ou a benefícios pessoais, mas ao cumprimento do dever moral, que deve ser orientado pela razão e não por interesses egoístas. Por exemplo, a ação de doação de recursos aos pobres com o objetivo de obter uma recompensa divina não é, nessa ética, moralmente correta. O motivo de sua ação não é baseado no dever, mas em um interesse pessoal, impedindo que a ação seja considerada verdadeiramente moral (Kant, 2003).

No tocante a dimensão tecnológica, os agentes morais desenvolvidos no contexto da IA visam a incorporação de princípios éticos em softwares projetados para tomar decisões com base em valores morais. Entre esses, podem ser citados o *Moral IDM*, o *Jeremy* e o *Ethoscool*. O *Moral IDM* é um software que compreende a linguagem natural humana e utiliza módulos que contêm princípios morais para avaliar as consequências de suas ações. O *Jeremy*, por sua vez, adota a ética utilitarista, avaliando a correção de suas ações com base no prazer ou felicidade gerados e no número de pessoas afetadas. Por fim, o *Ethoscool* é utilizado em ambientes educacionais para ensinar ética,

funcionando como um tutor virtual para os estudantes, promovendo a compreensão e aplicação dos princípios morais no cotidiano (UNESCO, 2021).

Ao se tratar de IA e ética é importante ponderar sobre a questão da responsabilidade civil de sua utilização. Quem deve ser responsabilizado pelos prejuízos causados por sistemas de IA? Deve ser o programador, o usuário ou a empresa proprietária do sistema? Essa questão se torna particularmente relevante à medida que a tecnologia assume atribuições preponderantes. Diversas organizações internacionais, como a UNESCO, a ONU, a OCDE e a União Europeia, têm se dedicado a estabelecer princípios para guiar o desenvolvimento e aplicação da IA, com um foco especial na justiça e na não discriminação. Tais princípios buscam evitar que a tecnologia amplifique preconceitos, como discriminação racial, de gênero ou de classe social. No contexto brasileiro, a Constituição Federal (CRFB 1988) apoia esses princípios ao garantir a igualdade e a não discriminação em seu art. 5º e ao estabelecer o compromisso com o bem-estar de todos no art. 3º (ONU, 2024).

Em relação à transparência e explicabilidade da tecnologia, a ética exige que os sistemas sejam compreensíveis, permitindo que usuários e autoridades entendam seu funcionamento e a origem das decisões tomadas. Esse princípio encontra respaldo na CRFB 1988, que garante o direito ao acesso à informação no art. 5º, inciso XIV. Além disso, o Código Civil e o Código de Defesa do Consumidor estabelecem requisitos para a transparência nos contratos e relações comerciais, incluindo os que envolvem a IA. Um exemplo disso seria um algoritmo de recomendação de crédito. Este deveria indicar claramente os critérios usados para aceitar ou negar um empréstimo, garantindo que os usuários compreendam como suas informações estão sendo processadas (Brasil, 1988).

A segurança e confiabilidade dos sistemas de IA são aspectos centrais da ética, especialmente em setores críticos, como o da saúde e transporte. A CRFB 1988, no art. 170, prevê a proteção ao consumidor como princípio da ordem econômica, exigindo que as empresas adotem normas que garantam a segurança dos produtos e serviços (Brasil, 1988). Isso inclui sistemas digitais, como os carros autônomos, que precisam ter mecanismos de redundância para evitar acidentes em caso de falhas técnicas. Se um carro autônomo sofrer uma falha no sistema de direção, ele deve ser capaz de acionar um sistema de backup para garantir a segurança dos passageiros e pedestres. Assim, a confiabilidade dos sistemas de IA deve ser assegurada para minimizar riscos e proteger a sociedade de falhas tecnológicas (OCDE, 2021).

No que diz respeito à privacidade e proteção de dados pessoais, a CRFB 1988, no art. 5º, inciso X, garante a inviolabilidade da intimidade e da vida privada, e a Lei Geral de Proteção de Dados (LGPD) regula o uso de dados pessoais no Brasil, exigindo consentimento explícito e medidas de

segurança contra abusos. Um exemplo claro dessa questão é o uso de assistentes virtuais, como a *Alexa* ou o *Google Assistant*, que devem respeitar a privacidade dos usuários, evitando a coleta de dados sem autorização. A coleta de informações pessoais para melhorar a experiência do usuário deve ser feita de forma transparente e com o consentimento adequado (Brasil, 1998; 2018).

Outro aspecto relevante é a responsabilidade e prestação de contas no uso da IA. Deve ser claro quem é responsável pelas decisões tomadas por sistemas digitais, especialmente quando estas resultam em danos. Em um exemplo prático, o uso de dispositivos de reconhecimento facial pela polícia pode gerar erros na identificação, redundando em injustiças. Caso haja um equívoco de identificação digital de um indivíduo, levando-o a detenção, é imprescindível a previsão de um mecanismo claro para revisar a decisão e responsabilizar as partes envolvidas. Isso também inclui a responsabilidade das empresas e desenvolvedores de IA, que devem garantir que os sistemas possam ser auditados e que os erros possam ser corrigidos de forma justa e transparente (UNESCO, 2021).

Por fim, a IA deve ser desenvolvida com o objetivo de promover o bem-estar humano e o benefício social. A CRFB 1988, em seu art. 3º, determina que a República do Brasil deve promover o desenvolvimento nacional, a erradicação da pobreza e a redução das desigualdades sociais (Brasil, 1988). A tecnologia deve ser usada para fins que beneficiem a sociedade como um todo. Exemplo disso é encontrado no uso de tecnologias digitais na medicina para diagnósticos precoce de doenças, prevenindo o seu agravamento. Dessa forma, a IA pode contribuir para o desenvolvimento sustentável, a educação e a inclusão social

CONSIDERAÇÕES FINAIS

Como foi demonstrado, a IA é revolucionária, sem precedentes na história da nossa civilização, causando um impacto social poderoso e amplo. Ela gera um aumento na produtividade industrial, melhorando a qualidade dos produtos e estimulando a inovação. Na saúde, facilita a pesquisa e a produção de novos medicamentos, além de proporcionar melhores diagnósticos e tratamentos de doenças, o que impacta positivamente no aumento da expectativa de vida da população. Na educação, amplia as fronteiras da ciência, ajudando em novas descobertas e oferecendo melhores ferramentas de aprendizado. Contudo, como toda tecnologia humana, ela está sujeita a falhas, com o potencial de provocar problemas graves para a sociedade. Além das falhas tecnológicas em si, há também os problemas gerados pelo seu mau uso, como a produção de armas químicas mais letais, drones autônomos com poder de matar pessoas e destruir cidades inteiras, além do perigo de seu uso na gestão de armamentos nucleares. É por isso que se faz urgente a criação de uma regulação ética e jurídica robusta sobre a IA, não apenas para os desenvolvedores e operadores, mas também

com princípios ético-jurídicos incorporados diretamente nos softwares. Nesse sentido, os princípios das éticas aristotélica, utilitarista e kantiana fornecem poderosas ferramentas para que a IA possa "pensar eticamente".

Na ética aristotélica, temos as virtudes e a *fronezis* (sabedoria prática), que nos fornecem um modelo para formar, por assim dizer, "o caráter interno" da IA. Expor a IA ao exercício das virtudes em várias situações da vida é uma boa abordagem de treinamento. Kant nos oferece a ideia de que o ser humano deve ser tratado sempre como um fim, e nunca como meio, assim como a lógica do imperativo categórico, que serve como uma boa base para auxiliar a IA na tomada de decisões. O utilitarismo de Jeremy Bentham e John Stuart Mill coloca a questão da utilidade da ação ética, tendo como parâmetro a felicidade geral, a qual pode ser mensurada matematicamente, facilitando a elaboração de algoritmos éticos. Além da ética, os preceitos jurídicos elaborados pela UNESCO e pela Carta dos Direitos Humanos das Nações Unidas, que foram inseridos nas leis específicas de cada país, criando um direito regulatório da IA com força coercitiva, nos fornecem uma base sólida para a proteção da humanidade contra os problemas causados pela nova tecnologia da IA.

Porém, entendemos que o poder da ética cristã não pode ser relegado, sobretudo no que diz respeito ao perigo da superinteligência, o qual, segundo alguns cientistas, será o último estágio da evolução da inteligência artificial, com o poder de destruir a humanidade. Na minha visão, a ética cristã pode esclarecer com mais profundidade a importância da humanidade na constituição e evolução do universo, ao colocá-la como parâmetro de um cosmos que pensa a si mesmo, sendo o amor a força que norteia o todo e que só pode ser expresso de forma plena por meio do elemento humano.

REFERÊNCIAS

ALMEIDA, José Manuel; Santos, Mário Miguel. *Introdução ao aprendizado de máquina*. São Paulo: Editora LTC, 2018.

ARISTÓTELES. *Ética a Nicômaco*. São Paulo: Companhia das Letras, 2015.

BENTHAM, Jeremy; Mill, John Stuart. *Uma Introdução aos princípios da moral e da legislação*. São Paulo: Abril, 1984.

BRASIL. Constituição da República Federativa do Brasil. 1988. Disponível em: *http://www.planalto.gov.br/ccivil_03/constituicao/constitu.* Acesso em: 05 ago. 2025.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais. Diário Oficial da União, Brasília, DF, 14 ago. 2018. Disponível em: *http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.* Acesso em: 5 ago. 2025.

COMPARATO, Fábio Konder. *Ética*: Direito, Moral e Religião no Mundo Moderno. São Paulo: Companhia das Letras, 2006.

CÓRDOVA, Paulo Roberto. *Inteligência Artificial*. Entre o fascínio e o medo. São Paulo: Contexto, 2025.

CORMEN, Thomas H.; Leiserson, Charles E.; Rivest, Ronald L.; Stein, Clifford. *Algoritmos*: Teoria e Prática. Rio de Janeiro: Elsevier, 2011.

DURAN, Carlos *et al*. *Inteligência artificial*: conceitos, aplicações e desafios. São Paulo: Senac, 2020.

HEATH, James. *Inteligência artificial aplicada à medicina e outras áreas*. São Paulo: Atlas, 2019.

KANT, Immanuel. *A Metafísica dos Costumes*. São Paulo: Edipro, 2003.

KANT, Immanuel. *Crítica da Razão Prática*. São Paulo: Ed. UNESP, 1999.

MERLE, Jean Christophe; Gomes, Alexandre Travessoni. *A Moral e o Direito em Kant*: Ensaios Analíticos. Belo Horizonte: Mandamentos, 2007.

ONU. *Princípios Globais das Nações Unidas para a Integridade da Informação*. 2024. Disponível em: *https://brasil.un.org/sites/default/files/202407/ONU_PrincipiosGlobais_IntegridadeDaInformacao_20240624.pdf.* Acesso em: 5 ago. 2025.

OCDE. *Artificial Intelligence:* The Next Digital Frontier?. 2021. Disponível em: *https://www.oecd.org/going-digital/ai/.* Acesso em: 5 ago. 2025.

RUSSELL, Stuart; Norvig, Peter. *Inteligência Artificial*: Estruturas e Estratégias para Solução Complexa de Problemas. São Paulo: Pearson Education, 2017.

SANTOS, Lúcia Maria; Almeida, Felipe Rodrigues. *Inteligência artificial e suas implicações no direito e na educação*. Rio de Janeiro: Editora FGV, 2021.

UNESCO. Ethics of Artificial Intelligence. 2021. Disponível em: *https://www.unesco.org/en/artificial-intelligence/recommendation-ethics.* Acesso em: 5 ago. 2025.

ZEGZEBSKI, Laura. *Exemplarist Moral Theory*. Oxford: Clarendon Press, 2004.