

UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE CIÊNCIAS HUMANAS
DEPARTAMENTO DE FILOSOFIA

NATHÁLIA DE SOUSA LIMA

A Técnica, o Sujeito e a Máquina: Fundamentos Ético-Filosóficos da Inteligência Artificial

BRASÍLIA
2025

NATHÁLIA DE SOUSA LIMA

A Técnica, o Sujeito e a Máquina: Fundamentos Ético-Filosóficos da Inteligência Artificial

Monografia apresentada ao Departamento de Filosofia da Universidade de Brasília, como requisito para obtenção do título de Bacharel/Licenciada em Filosofia.

Orientador: Prof. Dr. Philippe Claude Thierry Lacour

BANCA EXAMINADORA

Prof. Dr. Philippe Claude Thierry Lacour

Prof. Dr. André Leclerc

BRASÍLIA
2025

DEDICATÓRIA

Dedico este trabalho a todos aqueles que, direta ou indiretamente, convivem com as transformações trazidas pela inteligência artificial em nosso cotidiano. Que esta pesquisa, ainda que simples, possa despertar olhares mais atentos para os desafios éticos que surgem em meio às tecnologias que moldam nossas relações e a sociedade em que vivemos.

AGRADECIMENTOS

Primeiramente, agradeço imensamente a todos os professores e professoras com os quais tive a oportunidade de aprender nesses quatro anos de graduação, que finalizo agora em julho de 2025. De forma especial, agradeço ao meu orientador, Philippe Claude Thierry Lacour, por toda a paciência, apoio e dedicação desde o pré-projeto de monografia até a defesa.

Agradeço ao professor Herivelto Souza pela oportunidade de participar do Programa de Residência Pedagógica (PRP), experiência que foi de extrema importância para minha formação e para a vivência da docência.

Meus agradecimentos também aos professores André Leclerc, Rogério Basali, Priscila Rufinoni, Wanderson Flor e Hilan Nissior, que me inspiraram pela dedicação e compromisso em cada disciplina. As leituras, os debates e as reflexões que propuseram marcaram profundamente minha formação e visão de mundo.

Nos agradecimentos pessoais, meu coração se volta primeiramente para os meus pais, Ubiraci e Terezinha, por todo amor e apoio incondicional em cada decisão que tomei. Vocês sempre foram minha base e meu porto seguro. Se cheguei até aqui, foi porque me ensinaram a nunca desistir dos meus sonhos.

À minha tia Jandira, minha gratidão eterna por ser exemplo de força, caráter e por sempre estar ao meu lado nos momentos mais difíceis e também nos mais felizes.

Aos meus amigos e amigas, obrigada por compartilharem tantos momentos dessa caminhada. Em especial, o Luciano, que foi um verdadeiro pilar em todo o período deste trabalho. Vocês me acolheram, me deram forças quando pensei em desistir e acreditaram em mim mesmo quando eu duvidava. Cada palavra de incentivo e cada gesto de apoio de vocês foram fundamentais para que eu conseguisse concluir esta etapa tão importante da minha vida.

Por fim, agradeço a Deus pela vida, por me dar forças nos momentos mais desafiadores e por me mostrar, diariamente, que vale a pena lutar pelos nossos objetivos. Sou grata pela oportunidade de aprender, crescer e viver experiências que me moldaram não apenas como profissional, mas como ser humano. Que eu possa continuar a trilhar meus caminhos com amor, propósito e esperança.

A maior parte do
mal no mundo é
cometida por
pessoas que nunca
decidem ser boas ou
más.

Hannah Arendt

RESUMO

A presente monografia visa investigar as questões éticas envolvidas no uso da Inteligência Artificial (IA), especialmente quando as ações baseadas em IA podem não estar alinhadas com os valores éticos humanos tradicionais, tais como equidade, justiça e respeito à privacidade. Ao explorar essas problemáticas, a análise busca não apenas elucidar os dilemas éticos, mas também fornece diretrizes para uma implementação mais ética e responsável dessas tecnologias em nossa sociedade. Este trabalho investiga como as tecnologias de IA, ao mesmo tempo que oferecem soluções para problemas complexos, também podem perpetuar ou até intensificar discriminações e injustiças, desafiando a noção ética de 'fazer o bem'. Sob a ótica da filosofia aristotélica e dos autores Dora Kaufman, Mark Coeckelbergh e Martha Nussbaum, este estudo analisa como a crescente autonomia da IA desafia concepções éticas e jurídicas, e busca responder o problema de pesquisa: Até que ponto a inteligência artificial pode aderir aos princípios éticos humanos, considerando sua crescente capacidade de tomar decisões autônomas que afetam aspectos críticos da vida em sociedade.

Palavras-chave: inteligência artificial, ética, autonomia, dilemas éticos, sociedade

ABSTRACT

This monograph aims to investigate the ethical issues involved in the use of Artificial Intelligence (AI), especially when AI-based actions may not align with traditional human ethical values, such as equity, justice, and respect for privacy. By exploring these problems, this analysis seeks not only to elucidate ethical dilemmas but also to provide guidelines for a more ethical and responsible implementation of these technologies in our society. This work investigates how AI technologies, while offering solutions to complex problems, can also perpetuate or even intensify discrimination and injustices, challenging the ethical notion of 'doing good.' From the perspective of Aristotelian philosophy and authors Dora Kaufman, Mark Coeckelbergh, and Martha Nussbaum, this study analyzes how the increasing autonomy of AI challenges ethical and legal conceptions, and seeks to answer the research question: To what extent can artificial intelligence adhere to human ethical principles, considering its growing capacity for autonomous decision-making that affects critical aspects of life in society.

Keywords: artificial intelligence, ethics, autonomy, ethical dilemmas, society

SUMÁRIO

INTRODUÇÃO.....09
1-Capítulo I: Fundamentos Éticos e Filosóficos.....13
1.1 A Ética de Aristóteles.....16
1.2 O que é um Viés Algorítmico.....18

2-Capítulo II – Inteligência Artificial na Atualidade.....23
2.1 O que é a Inteligência Artificial Atualmente?.....25
2.2 Ética e Viés Algorítmico: Os Algoritmos Podem Ser Éticos?.....27
2.3 Superinteligência e Transumanismo: Possibilidades de uma IA Geral.....30
2.4 A Importância de Discutir Problemas Éticos e Sociais.....34

3-Capítulo III– Dilemas Éticos na Prática.....36
3.1 IA e Carros Autoguiados: O Problema do Bonde e a Responsabilidade Moral.....37
3.2 IA e Direitos Autorais: Criação, Cópia e os Limites Éticos.....42
4-CONSIDERAÇÕES FINAIS.....49
5-REFERÊNCIAS.....52

INTRODUÇÃO

A análise da Inteligência Artificial (IA) através do campo da ética surge como uma necessidade contemporânea diante do crescente impacto dessas tecnologias em diversas esferas da sociedade. A análise da Inteligência Artificial (IA) sob a perspectiva ética configura-se como uma exigência atual, diante do impacto crescente dessas tecnologias nas mais diversas esferas da vida social. Discutir os aspectos éticos relacionados à Inteligência Artificial não se limita à avaliação normativa de seus usos ou impactos sociais visíveis. Trata-se, sobretudo, de reconhecer que a IA carrega, em sua própria estrutura e concepção, decisões morais, epistemológicas e políticas previamente inscritas em seus códigos e arquiteturas. Assim, a ética não apenas tangencia a utilização de IA — ela está materialmente presente na forma como essas tecnologias são projetadas, implementadas e integradas ao tecido social. Em outras palavras, pensar eticamente a IA é interrogar os valores que orientam suas finalidades, os critérios de racionalidade que sustentam sua lógica operacional e as formas de poder que ela reproduz ou naturaliza. Abordar esse tema é essencial para garantir que o avanço tecnológico esteja, de fato, alinhado com valores humanos — não no sentido abstrato de uma ética universal descontextualizada, mas sim considerando os múltiplos sistemas de valores que disputam espaço na sociedade contemporânea. Isso implica refletir sobre quais vidas são priorizadas, quais visões de mundo são embutidas nos algoritmos, e quais normativas realmente asseguram uma justiça substantiva no uso dessas tecnologias. O desenvolvimento e a aplicação da IA devem, portanto, ser orientados por princípios éticos que não se limitem à obediência formal à lei, mas que reconheçam os limites, riscos e assimetrias de poder implicados em sua materialidade.

Além disso, a aplicação de uma perspectiva ética pode contribuir para identificar e mitigar vieses algorítmicos: como os preconceitos presentes nos dados podem influenciar as decisões automatizadas; incentivar um desenvolvimento sustentável e responsável: orientando a criação de sistemas que respeitem normas e direitos fundamentais e orientar políticas públicas e regulações: que acompanhem o ritmo acelerado da inovação tecnológica.

A seleção deste objeto de estudo reflete não apenas a relevância do debate contemporâneo sobre Inteligência Artificial, mas também a urgência de uma reflexão crítica acerca de seus impactos éticos. Ao inserir-se nos debates filosóficos que

tangenciam o tema, esta pesquisa busca compreender as implicações normativas, ontológicas e políticas que emergem da relação entre tecnologia e sociedade. Destacam-se, ainda, como fatores determinantes para o despertar do interesse investigativo: a relevância da discussão ética no ambiente tecnológico atual, o impacto social e cultural da IA e a necessidade de compreender como as escolhas de desenvolvimento podem refletir e reforçar determinadas visões de mundo e os debates cotidianos sobre o papel das redes sociais e da IA na vida contemporânea, bem como de suas representações em filmes e obras de ficção científica. Esse interesse foi consolidado após a participação, há aproximadamente um ano, em um minicurso promovido na Universidade de Brasília (UnB), intitulado "Inteligência Artificial: do Golem ao Chat GPT"[1], ministrado pelo professor Christian Arão. A experiência proporcionou uma base crítica e instigante para compreender as diversas camadas conceituais, históricas e sociais que envolvem o desenvolvimento da IA.

Busca-se apresentar argumentos, perspectivas e enquadramentos teóricos que revelem as contribuições da filosofia à análise da Inteligência Artificial, de modo a possibilitar o contraste crítico com discursos que tendem à naturalização ou à glorificação acrítica da tecnologia. Nesse sentido, a reflexão filosófica pode contribuir para a análise crítica dos avanços em inteligência artificial. Além disso, o trabalho busca estimular uma reflexão mais ampla sobre os benefícios e riscos associados ao uso da IA, bem como questionar os discursos sensacionalistas que reforçam uma imagem distorcida e alarmista da chamada "superinteligência", muitas vezes sustentada por narrativas cinematográficas, desinformação e especulações não fundamentadas. Assim, o objetivo é fomentar uma abordagem ética e crítica sobre o papel da inteligência artificial na sociedade contemporânea.

Ressalta-se ainda que a ética aplicada à inteligência artificial não se limita a uma análise teórica, mas envolve também práticas concretas no desenvolvimento e na implementação de algoritmos. A forma como sistemas são treinados, os dados que os alimentam, os objetivos que os orientam — tudo isso deve ser avaliado à luz de princípios éticos. Assim, torna-se urgente indagar: de que maneira a ética pode influenciar o

[1] ARÃO, Cristian (PPGFIL). *Minicurso com Dr. Cristian Arão –* Inteligência Artificial: do Golem ao Chat GPT Parte 1. [S. l.]: YouTube, 6 e 7 jul. 2023. Disponível em:https://youtu.be/dUctDHSTgkE. Acesso em: 29 jul. 2025.

processo de criação da IA? Que diretrizes éticas devem ser adotadas por cientistas e engenheiros da computação?

Entre os problemas emergentes, destaca-se o viés algorítmico, que se manifesta quando sistemas automatizados perpetuam ou ampliam desigualdades sociais, culturais e econômicas. Casos concretos de discriminação algorítmica em áreas como segurança pública, crédito bancário e recrutamento profissional demonstram que os algoritmos não são neutros. Pelo contrário, refletem valores, intencionalidades e estruturas sociais existentes. Assim, é necessário refletir, apresentar contradições e promover alterações para que esses sistemas não reproduzam desigualdades, opressões sobre os efeitos desses vieses e os mecanismos éticos de controle e correção possíveis.

Desta forma, o presente trabalho divide-se em três capítulos, organizados da seguinte maneira: O primeiro capítulo busca apresentar, a partir de uma abordagem ética, possíveis formas de analisar criticamente a Inteligência Artificial. Nesse contexto, a discussão se aprofunda ao considerar a possibilidade de que sistemas artificiais possam agir moralmente. É plausível afirmar que algoritmos podem ser éticos? E, se for possível incorporar valores humanos em sistemas automatizados, como isso afetaria a autonomia decisória dessas máquinas?

O segundo capítulo tem como objetivo desmistificar a ideia de neutralidade algorítmica, discutindo os vieses, as estruturas de poder e os discursos que sustentam o desenvolvimento tecnológico contemporâneo. As perguntas sobre a ética dos algoritmos ganham ainda mais peso quando se projeta o futuro da IA para além das inteligências fracas ou restritas, considerando a possibilidade de uma inteligência artificial generativa (IAG)[2] ou mesmo de uma superinteligência.

. Por fim, o último capítulo propõe-se a analisar manifestações concretas da IA na realidade social, com ênfase em suas implicações éticas, jurídicas e políticas. Nesse horizonte, perspectivas como o transumanismo e as especulações sobre a dominância das máquinas abrem espaço para novas discussões sobre liberdade, responsabilidade e os limites dos seres humanos. A filosofia, portanto, não é mero adorno reflexivo, mas torna-

[2] IAG geralmente significa Inteligência Artificial que se iguala (ou supera) os humanos em uma série de tarefas.

se uma instância crítica e inadiável, capaz de tensionar as promessas tecnológicas que, sob o verniz do progresso, podem mascarar formas mais sutis de dominação.

CAPÍTULO I

A ÉTICA APLICADA À INTELIGÊNCIA ARTIFICIAL

A ética, como ramo da filosofia, investiga os princípios que orientam o comportamento humano, definindo o que é moralmente correto ou incorreto. Ao longo da história, diferentes correntes filosóficas propuseram teorias para explicar e normatizar a conduta humana, buscando responder à pergunta: "como devemos viver?" Essas reflexões não se limitam ao campo teórico, tendo profundas implicações na sociedade e nas interações cotidianas, especialmente quando novas tecnologias, como a inteligência artificial (IA), se tornam parte integrante da vida moderna.

A ética aplicada, especialmente em relação à IA, surge como uma necessidade urgente na contemporaneidade. Sistemas de IA são cada vez mais utilizados para auxiliar em decisões que afetam diretamente as vidas humanas, como no campo da saúde, no direito e até mesmo em relações interpessoais mediadas pela tecnologia. Diante disso, a reflexão ética não pode ser deixada de lado, pois envolve questões como privacidade, responsabilidade, justiça, transparência e o impacto das ações automatizadas na sociedade.

Sob essa perspectiva, a ética aristotélica tem se mostrado relevante ao fornecer um arcabouço para a análise e implementação de sistemas de IA, motivo pelo qual adotou-se, como referencial teórico, a perspectiva da ética da virtude. Aristóteles, em sua obra *Ética a Nicômaco,* propõe que a virtude é o caminho para a *eudaimonia*, ou a realização plena, e que a moralidade está ligada à capacidade de deliberação e escolha racional.

Aplicar a ética aristotélica à IA significa questionar como sistemas autônomos podem ser programados para deliberar e agir de forma moral. Em uma sociedade onde as máquinas já executam funções complexas, é crucial que essas tecnologias sejam guiadas por princípios que levem em consideração o bem-estar humano e a justiça. Embora a IA não tenha capacidades de raciocínio, o uso de aprendizado de máquina e algoritmos sofisticados sugere a possibilidade de aproximá-las, ao menos parcialmente, de um modelo de tomada de decisão moralmente orientado.

Assim, a ética aristotélica oferece uma perspectiva relevante para a análise de sistemas de IA levantando questões sobre como podemos garantir que essas máquinas ajam em conformidade com os princípios da virtude e do bem comum, evitando decisões puramente instrumentais ou desumanizadas.

Este capítulo se aprofundará em dois pilares para entender os desafios éticos na era da Inteligência Artificial (IA) da seguinte maneira: O primeiro tópico vai explorar a ética de Aristóteles, revisitando suas origens e conceitos fundamentais como ethos, moralis e a busca pela Eudaimonia - a realização plena ou felicidade alcançada através da virtude. A sua obra *Política* abordava o tema numa época em que o termo "inteligência artificial" sequer havia sido imaginado, mas Aristóteles se preocupava com um tempo em que a automação eliminaria a necessidade do trabalho humano e como isso alteraria a dinâmica social.

Utilizando como alicerce teórico as obras de Martha Nussbaum, incluindo Fragilidade da Bondade(2009),Cultivando da Humanidade (2014) e Criar Capacidades: Proposta para o Desenvolvimento Humano(2012), este trabalho traça uma jornada pelo entendimento da formação humana no contexto das *capabilities*[3] sob o pensamento de Martha Nussbaum. Sua releitura da ética eudaimônica ampliam essa discussão, questionando se o avanço da IA estar alinhado com o bem-estar social e a formação de uma cidadania global, ou se, ao invés disso, privilegia um crescimento puramente econômico que pode exacerbar desigualdades.

Em seguida, o segundo tópico abordará sobre o viés algorítmico, um fenômeno intrínseco aos sistemas de IA que pode gerar resultados injustos e discriminatórios. Investiga-se como esses vieses surgem, muitas vezes a partir de dados históricos enviesados, e como se manifestam em diversas aplicações, como contratação de pessoal e reconhecimento facial. Pretende-se analisar também as diferentes categorias de viés, desde problemas nos dados de entrada (como seleção e desatualização) até questões no funcionamento interno do algoritmo que podem confundir correlação com causalidade.

[3] É importante notar que o conceito de capabilities é substancialmente mais abrangente em comparação com o termo capacidade (capacity), que é utilizado por Nussbaum principalmente em seus argumentos relacionados a formação e à educação. Portanto, permanece o termo original.

1.1 A ÉTICA DE ARISTÓTELES

Para falar do conceito de ética é preciso voltar a origem da palavra, portanto, etimologicamente, a palavra "ética" deriva do termo grego *ethos*, que significa o conjunto de costumes, hábitos e valores de uma determinada sociedade ou cultura. Os romanos traduziram para o termo latino *mos, moris* (mantendo o significado de *ethos*), dos quais provém *moralis,* que deu origem ao termo moral em português. Segundo o estudo de Santos (2021) a ética, desde a antiguidade grega, diz respeito à reflexão sobre a vida prática, ou seja, sobre a ação boa, justa. A ação ética é fruto de uma escolha refletida, pensada, deliberada, que pressupõe uma justificativa.

Apesar do significado ser similar ao de ética, a moral designa mais o conjunto de hábitos e costumes de uma sociedade. A moral muda constantemente, pois os hábitos sociais são renovados periodicamente conforme o local em que são observados. Esclarecidos esses aspectos preliminares, podemos, agora, iniciar, propriamente, a discussão do problema do bem supremo em Aristóteles. Logo no início da Ética a Nicômacos, Aristóteles dá o tom geral de toda sua ética. Toda arte e toda investigação e igualmente toda ação e toda escolha tendem a algum bem. Toda ação visa alguma coisa e de sua tendência a produzir esta coisa ela tira seu valor. A ética aristotélica é nitidamente teleológica (visa um objetivo substancial e não apenas uma forma) e eudemonista (o objetivo é a felicidade), fundada nos juízes morais do homem bom e virtuoso. Aristóteles interpreta a ação humana segundo a categoria de meio e fim. O fim ao qual tende uma ação particular não pode ser senão um meio em vista de um fim ulterior, mas é necessário que tenha um limite para a sequência. Cada ação deve ter um fim último que tenha um valor nele mesmo, e, conclui Aristóteles, sem hesitação, o fim último de todas as coisas deve ser o mesmo.

O termo *Eudaimonia*, muitas vezes traduzido como felicidade, não se refere a um estado de prazer fugaz ou hedonismo, mas sim a um estado de realização plena ou de caminho onde a excelência moral e intelectual é alcançada. Para Aristóteles, a *Eudaimonia* não é um fim em si mesma, mas sim o resultado natural de viver uma vida virtuosa em favor de uma construção ética coletiva, a saber, a pólis. Em um contexto aristotélico, a IA não poderia ser considerada um agente moral em si. Contudo, o desenvolvimento e o uso da IA são fruto das decisões humanas, e, portanto, aqueles que criam, programam e utilizam sistemas de IA têm responsabilidade ética sobre seus efeitos.

Isso sugere que a ética aristotélica poderia ser aplicada aos desenvolvedores de IA, que devem buscar a virtude ao construir tecnologias que promovam o bem comum.

Nussbaum, destacada filósofa política contemporânea, baseia sua abordagem ética na tradição aristotélica, dando enfoque ao conceito Eudaimonia que diz respeito à realização da excelência moral e ao pleno desenvolvimento das capacidades humanas. Mas a sua concepção sobre ética não se limita à esfera individual: ela promove uma cidadania global, destacando a importância da formação cidadã para influenciar e fortalecer as democracias modernas. Com sua análise, é possível explorar como a IA, cada vez mais relacionada com a vida humana, circunscreve na perspectiva de cultivo da humanidade e impacta não apenas na definição da democracia, mas também na formação humano democrática, inerentemente vinculada na construção ética aplicada ao agir humano diante de tais realidades.[4]

Desta maneira, ao aplicar a ética eudaimônica sob a releitura de Nussbaum em relação a IA, levanta questionamentos não somente sobre o avanço acelerado destes sistemas nas sociedades, mas sobretudo o risco em que se coloca a qualidade de vida e a ética do bem-estar social, como adverte Nussbaum; pois privilegia-se um modelo centrado exclusivamente no crescimento econômico:"[...] esqueça a desigualdade distributiva e social, esqueça a qualidade das relações raciais e de gênero, esqueça o aperfeiçoamento de outros aspectos da qualidade de vida do ser humano" (NUSSBAUM, 2019, p.14). Isto significa que, não há como discutir a Ética Aplicada à IA sem considerar as implicações das sociedades contemporâneas cada vez mais desiguais e instáveis democraticamente. Destaca-se, aqui, a contribuição da autora na ética eudaimônica, o que se atenta à formação individual e ao fazer coletivo.

Nesse processo de discussão da construção ética do ser humano, faz-se necessário a abordagem de Nussbaum para entender que, a partir da produção de sistemas padronizadores e centralizados ao resultado econômico e não a modelos de sistemas que gerem o desenvolvimento humano, a qualidade de vida e a ética do bem-estar social, estamos gerando risco as democracias e à formação ética. Posto isto, orienta Nussbaum: "o cidadão do mundo deve aprender a desenvolver compreensão e empatia para as culturas lejanas e para as minorias étnicas, raciais, religiosas que estão dentro de sua

[4] G. CARREIRA JUNIOR. Ética aplicada à inteligência artificial: interfaces com a teoria de Martha Nussbaum. *Revista Instante*, v. 5, n. 2, p. 160-176, 2023.

própria cultura. Além disso, você deve desenvolver sua compreensão da história e da variedade das ideias humanas sobre gênero e sexualidade" (NUSSBAUM, 2005,p. 98, tradução própria). Cabe reforçar que esses princípios são trazidos para observar o crescimento das IAs atuais e como estão previstas no modelo que não expande as possiblidades. Desse modo, aética aplicada ao agir humano em relação à inteligência artificial, sob a perspectiva da teoria de Martha Nussbaum, está profundamente enraizada na busca pela análise da formação moral e do desenvolvimento humano. A partir de Nussbaum, questionamos: será que a Ética Aplicada não deve ser uma orientação prática a fim de que moldemos nossas ações e decisões em um mundo cada vez mais permeado pela IA?

Essa abordagem ética, baseada na Eudaimonia que Martha Nussbaum enfatiza, está na importância do desenvolvimento humano em vista da compreensão de felicidade. Ela adverte contra a ênfase excessiva na busca pelo lucro econômico em detrimento do bem-estar social e da qualidade de vida. Nussbaum aponta a necessidade de empatia e compreensão para culturas diversas e as minorias, bem como a importância de reconhecer a diversidade de ideias. Isso levanta a questão de como as atuais inteligências artificiais estão sendo desenvolvidas em relação a esses princípios éticos e como as pessoas devem orientar suas ações e decisões na era da IA, com base em uma ética que promova valores humanos fundamentais, ancorada na teoria das *capabilities*.

1.2 **O QUE É UM VIÉS ALGORÍTIMO**

Viés algorítmico refere-se a tendências sistemáticas e não intencionais em sistemas e algoritmos que podem criar resultados injustos, como privilegiar um grupo de pessoas em detrimento de outros. Esse viés muitas vezes decorre dos dados usados para treinar esses sistemas, que podem refletir preconceitos históricos ou sociais existentes na sociedade. Quando os algoritmos são aplicados em situações reais, como na seleção de candidatos para entrevistas de emprego, no reconhecimento facial, na concessão de crédito, entre outros, eles podem perpetuar ou até ampliar essas disparidades se não forem devidamente ajustados para neutralizar tais vieses.

Por exemplo, se um algoritmo de contratação é treinado com dados históricos de uma empresa que tinha uma tendência a contratar mais homens do que mulheres, ele pode aprender erroneamente que ser homem é um indicador de um possível empregado bem-sucedido, resultando em uma preferência injusta por candidatos masculinos.

O algoritmo é uma construção matemática antiga, utilizada para processar dados que produzem evidências e motivam ações específicas em direção à solução de um determinado problema, destacam Rosetti e Angeluci (2021)[5]. Sua invenção ligada aos sistemas computacionais foi consolidada com estudos, como o da Máquina de Turing, fundantes da ciência da computação. Uma receita de um bolo, por exemplo, é um clássico exemplo metafórico de aplicação algorítmica: a correta execução de seus componentes em determinada sequência e medidas (uso dos ingredientes farinha, ovos e leite e seu modo de preparo passo-a-passo) conduzem ao objetivo final (o bolo pronto).

Um algoritmo não é apenas uma mera sequência de passos necessário para a realização de determinada tarefa. Sua implementação abrange diversos atores, autômatos ou humanos, tomadas de decisão por lógica e repetição de ações. Os autores explicam que pelo conceito atualmente está ligado à construção da IA, alguns estudos como (TADDEO; FLORIDI, 2018; MCCARTHY et al., 2006; SAMUEL, 1960; YANG et al., 2018)[6] apontam que os algoritmos são algo além do que um pacote de modelos computacionais que dão base a uma arquitetura tecnológica. Os algoritmos aplicados à IA seriam, no caso, "um recurso crescente de agência de autoaprendizagem, interativa, autônoma, que permite que artefatos computacionais executem tarefas que, de outra forma, exigiriam que a inteligência humana fosse executada com sucesso" (TADDEO; FLORIDI, 2018, p. 751).

Aplicados em uma variedade de atividades que perpassam a vida humana, aos algoritmos são atribuídas uma série de demandas que beneficiam a sociedade da informação, evidenciam Rosetti e Angeluci. Os algoritmos conseguem reduzir custos e riscos e trazer novas soluções aos problemas complexos; podem, no entanto, conduzir a consequências nocivas e inesperadas, sobretudo quando em seu processo de execução prescinde da nossa supervisão.

Por esse motivo, para combater o viés algorítmico, é crucial uma abordagem multidisciplinar que inclui a transparência dos dados e algoritmos, revisão contínua dos resultados dos algoritmos para detectar padrões de viés, e a inclusão de uma diversidade de perspectivas no desenvolvimento e implementação de tecnologia.

---

[5] ARAÚJO, G. M. O. Ética algorítmica: questões e desafios éticos do avanço tecnológico da sociedade da informação. *Galáxia*, São Paulo, n. 47, p. 1-18, 2023.

[6] TADDEO, M. R.; FLORIDI, L. How AI can be a force for good. *Science*, v. 361, n. 6404, p. 751-752, 2018.

Em concordância com os autores acima, Carvalho (2020) ressalta que a ciência de dados é a principal forma de alimentação e desenvolvimento das ferramentas atuais de inteligência artificial. Portanto, não é surpreendente que muitos tipos de vieses causadores da discriminação algorítmica envolvam de uma maneira ou de outra o uso de dados pelo sistema.

Carvalho explica que o governo federal dos Estados Unidos, vem há alguns anos elaborando estudos científicos para subsidiar sua política nacional de inteligência artificial e optou por agrupar os diferentes tipos de vieses em duas categorias principais: a primeira relacionada aos dados de entrada (inputs) em um algoritmo e a segunda abrangendo o funcionamento interno do algoritmo em si. É necessário observar, também, que existem diversas denominações para cada tipo de viés, mas como aparentemente (ainda) não há acordo ou padronização a respeito da sua nomenclatura, aqui será indicada somente sua conceituação, deixando de lado toda a diversidade de nomes que poderiam receber.

No que se refere à primeira categoria, Carvalho exemplifica que o viés se apresenta na medida em que o algoritmo precisa decidir quais dados serão utilizados pelo sistema, em detrimento de outros. Essa é uma ação normal em todos os softwares que utilizam técnicas de inteligência artificial. Por exemplo, sua aplicação é comum em programas de navegação por satélite presentes em aparelhos móveis, nos quais os dados de velocidade de cada usuário permitem ao sistema conhecer a rota mais rápida para chegar ao destino escolhido.

Nessa situação, é algo positivo, mas sua análise permite a compreensão de modo simplificado as principais maneiras pelas quais o resultado da operação algorítmica pode resultar em práticas discriminatórias. Os vieses aqui podem ser consequências pelas deficiências na inclusão e tratamento dos dados, isto é, quando algumas informações são consideradas como mais ou menos importantes para a decisão do algoritmo. No exemplo descrito acima, o sistema de navegação poderia não incluir a opção de realizar o trajeto por meio de bicicletas ou a pé, deixando em desvantagem as pessoas que não possuem carros.

Um outro tipo de viés muito frequente ocorre quando os dados apresentam algum erro ou estão desatualizados/incompletos. Pois, no caso do trajeto retratado, ainda que o sistema oferecesse a rota utilizando o transporte público, de nada adiantaria se as informações de horários não estivessem atualizadas, outra vez prejudicando os usuários que não utilizam carros.

Existe ainda uma terceira categoria de viés algorítmico, relacionado aos dados de entrada, que compreende o viés de seleção (provavelmente o mais frequente na prática), ocorrendo quando os dados inseridos nos sistemas não condizem com a realidade. No exemplo escolhido acima, se os dados forem majoritariamente obtidos a partir de telefones de última geração, é provável que a precisão do sistema seja prejudicada nas áreas mais pobres da região, que receberiam um serviço de menor qualidade, já que essa população de menor poder aquisitivo não participaria do serviço da mesma forma.

A quarta categoria de viés interno, abrange os mecanismos de tomada de decisão que pressupõem equivocadamente que a correlação estatística necessariamente implica na causalidade. Ou seja, um programador (humano) ou o próprio sistema algorítmico pode analisar dois fatores que frequentemente ocorrem juntos (por exemplo, pertencer a um determinado grupo étnico e ter um certo nível de renda), e presumir que há necessariamente algum nexo de causalidade entre os dois. Consequentemente, criar uma relação causal quando existe somente correlação entre os dados pode levar à discriminação de usuários específicos. Carvalho cita um exemplo de estudo feito nos Estados Unidos, sobre análises demográficas promovidas por seguros de saúde e, para uma mesma classificação de risco de adoecer, os pacientes negros na verdade já apresentam condições de saúde mais agravadas do que os pacientes brancos. Essa diferença se dá porque os negros em geral possuem uma renda menor e, portanto, possuem menos acesso aos serviços de saúde (privados) naquele país. Uma das consequências é que os médicos podem utilizar essa classificação de risco enviesada para formular o tratamento mais adequado para cada paciente, no caso dos negros podem estar aquém da sua real necessidade.

Por último, há também uma quinta categoria de viés interno, que ocorre quando o sistema trabalha com conjuntos de dados que não correspondem à realidade daquilo que está sendo representado, isto é, quando as informações estão incompletas ou representam desproporcionalmente certos estratos da população. Nesse contexto, o resultado da operação algorítmica será distorcido, sedimentando de modo eficaz a discriminação devido à falha dos dados de entradas iniciais. Carvalho afirma que é necessário que os algoritmos sejam elaborados de modo que lidem com a diversidade, tendo em vista que, a disponibilidade de dados, o acesso à tecnologia e a participação nas ferramentas digitais estão distribuídos de maneira desigual na sociedade. Caso contrário, barreiras econômicas, linguísticas, estruturais ou socioeconômicas podem ser ampliadas pelos mecanismos de inteligência artificial. Sem ajuste adequado, essa característica pode se

tornar uma falha sistêmica que reforça os padrões existentes de discriminação por meio de uma descrição estatística pouco representativa que será utilizada pelo algoritmo, gerando tratamento desigual para as pessoas envolvidas.

Esses tipos de vieses podem representar grandes problemas para os usuários, considerando a disseminação cada vez maior das técnicas de inteligência artificial, pois no caso do aprendizado de máquina, tal como os códigos de programação, são complexos e por vezes incompreensíveis até mesmo para seus programadores.

Dessa forma, já que o uso da inteligência artificial deve ser torna ainda mais presente na vida cotidiana das pessoas, será cada vez mais necessário dedicar atenção para compreender a maneira de operação desses sistemas, avaliando continuamente os dados que alimento o algoritmo e seu mecanismo de atuação, bem como os resultados produzidos. Além do mais, ainda é comum a crença de que os números seriam "naturalmente" neutros e imparciais, e que representariam sempre a mesma realidade objetiva, o que traz sérias dificuldades e contribui para ofuscar situações discriminatórias que afetam negativamente a vida das pessoas.

Assim, os desafios relacionados a essa nova realidade tecnológica incluem sobretudo os resultados das operações automatizadas, que devem sempre garantir que as informações sobre lugares, pessoas, comportamentos etc. sejam usados de forma adequada, ética e na promoção de princípios democráticos, como a não-discriminação e igualdade de oportunidades.

CAPÍTULO II
INTELIGÊNCIA ARTIFICIAL NA ATUALIDADE

Certamente a ética em IA não é apenas um assunto técnico ou filosófico - é também um desafio global que demanda atenção a perspectivas e assuntos interculturais. À medida que os sistemas de IA são desenvolvidos e implantados em diferentes partes do mundo, encontra-se diversos contextos sociais, culturais e políticos. É essencial assegurar que os sistemas de IA sejam desenvolvidos de modo que respeitem os valores locais e costumes, e que não reforcem inadvertidamente estruturas de poder existentes ou desigualdades – tanto no nível local quanto global. No Brasil tem um crescente interesse em ética na IA, com muitos pesquisadores e desenvolvedores trabalhando numa ampla gama de assuntos: das implicações éticas da tecnologia de reconhecimento facial ao impacto da IA no futuro do trabalho. Há também um forte foco no papel da IA na promoção de justiça social e enfrentamento da desigualdade, particularmente nas áreas de saúde, educação e políticas públicas.

Neste capítulo, pretende-se realizar uma análise da relação entre a ética e a inteligência artificial, sob um olhar filosófico de Dora Kaufman[7] e Mark Coeckelbergh, delimitando a importância de separar a ficção da realidade, afinal, a IA implementada atualmente em larga escala deve ser encarada como parceira dos profissionais humanos nos processos de decisão, e não soberana, ou seja, capaz de contribuir para aumentar a inteligência humana especializada, e não substituí-la. Kaufman destaca a necessidade de uma "ética algorítmica", que compreenda as implicações morais e sociais da IA levando em consideração os impactos no trabalho, na privacidade e nas desigualdades sociais. Para ela, os algoritmos não são neutros, e é fundamental que as decisões automatizadas sejam transparentes e responsáveis.

[7] KAUFMAN, Dora. É uma pesquisadora dos impactos sociais da Inteligência Artificial, atuando no Instituto de Estudos Avançados (IEA) e no Centro de Pesquisa em Redes Digitais Atopos na Escola de Comunicação e Artes (ECA), ambos da Universidade de São Paulo (USP). Também é pós-doutoranda em Filosofia no Programa de Pós-graduação Tecnologias da Inteligência e Design Digital (TIDD) na PUCSP. Disponível em: https://online.pucrs.br/professores/dora-kaufman. Acesso em: 29 jul. 2025.

Por sua vez, Coeckelbergh[8] adota uma abordagem filosófica mais ampla, discutindo a "ética da tecnologia", que não se limita aos aspectos técnicos, mas examina as relações humanas com a IA. Ele questiona como a IA altera nossas interações sociais e nosso senso de moralidade, sugerindo que a ética deve acompanhar o desenvolvimento tecnológico, adaptando-se às novas realidades trazidas pela inteligência artificial.

O tema da ética permeia a sociedade humana desde Aristóteles e foi mudando de sentido ao longo da história, resguardando, contudo, a crença de que apenas o ser humano é dotado da capacidade de pensar criticamente sobre valores morais e dirigir suas ações em termos de tais valores. Com o avanço recente das tecnologias de inteligência artificial, as questões éticas estão em pauta, associados à robótica ou mediando as interações sociais e os processos decisórios, os algoritmos de IA agregam diversos benefícios, mas simultaneamente, carecem de transparência, são difíceis de ser explicados e comprometem a privacidade. Diariamente, aparecem caos ilustrativos no Brasil e no mundo afora.

[8] MARK COECKELBERGH. É um filósofo belga da tecnologia. Ele é professor de Filosofia da Mídia e Tecnologia no Departamento de Filosofia da Universidade de Viena. É especialista em ética da inteligência artificial e é mais conhecido por seu trabalho em filosofia da tecnologia e ética da robótica. Também publicou nas áreas de filosofia moral, filosofia ambiental e, mais recentemente, filosofia política. Disponível em: https://en.wikipedia.org/wiki/Mark_Coeckelbergh. Acesso em: 29 jul. 2025.

2.1 O QUE É A INTELIGÊNCIA ARTIFICIAL ATUALMENTE?

Assim como toda tecnologia, a IA é social e humana, seus efeitos derivam do que os seres humanos fazem com ela, como a percebem, como a experimentam e usam, como a inserem nos ambientes técnicos-sociais. É inegável que a IA faz parte da vida cotidiana dos seres humanos. As pessoas conseguem acessar a sistemas inteligentes para programar itinerário como o Waze, pesquisar no Google e receber da Netflix e do Spotify recomendações de filmes e músicas. A Amazon, por exemplo, captura as preferências dos usuários no fluxo de dados que coleta a partir das interações com a plataforma. A Siri, da Apple, e a Alexa, da Amazon, são assistentes pessoais digitais inteligentes que ajudam na localização de informações úteis com acesso por meio de voz.

Não existe uma definição universal de inteligência artificial. A IA pode ser definida como inteligência exibida ou simulada por código (algoritmos) ou máquinas. Mas, filosoficamente falando, é um conceito vago. Essa definição levanta a questão de como definir a inteligência. Uma comparação óbvia com o conceito vago é a inteligência de tipo humano. Por exemplo, Philip Jansen e colaboradores definem a IA como "ciência e a engenharia de máquinas com capacidades que são consideradas inteligentes pelo padrão de inteligência humana" (COECKELBERGH, p. 65).

Neste contexto, a IA diz respeito à criação de máquinas inteligentes que pensam ou (re)agem como seres humanos. Porém, muitos pensadores da IA consideram que a inteligência não precisa ser semelhante à humana e optam por uma definição mais neutra, formulada em termos independentes da inteligência humana e dos objetivos associados à IA geral ou forte.(COECKELBERGH, 2024, p. 66).Enumeram todos os tipos de funções e tarefas cognitivas, como aprendizado, percepção, planejamento, processamento de linguagem natural, raciocínio, tomada de decisão e resolução de problemas – o último também frequentemente equiparado à inteligência em si.

Para o termo "inteligência" também é difícil de designar uma definição exata. De modo geral, o termo se refere a capacidade de um agente atingir objetivos determinados em uma ampla gama de domínios e, comumente, é associado à espécie humana. Definições específicas incluem atributos intangíveis, como a capacidade de fazer analogias e compreender significados, além de consciência, intencionalidade, livre arbítrio, ética, moral. De acordo com estes parâmetros, no atual estágio de desenvolvimento, a inteligência artificial não é inteligente.

A IA também não é “artificial”, seu desenvolvimento e uso extrapolam a esfera abstrata dos algoritmos, dependem de infraestruturas físicas, que, por sua vez, dependem de recursos naturais, principalmente o lítio. De cor prateada, maleável, chamado de “ouro branco”, o mineral é estratégico para o setor de tecnologia: os equipamentos eletrônicos portáteis (smartphones, notebooks, câmeras digitais, assistentes virtuais), interface entre os usuários e os sistemas de inteligência artificial, funcionam com baterias de íons de lítio (LIB). Sucessivamente, elas estão sendo implantadas nos data centers, devido à sua maior densidade energética e vida útil significativamente mais longa, além de "pegada ecológica" (KAUFMAN, 2022, p. 227).

Antes de discutir de modo mais concreto e detalhado os problemas éticos da IA, é preciso de um melhor entendimento da tecnologia e de suas aplicações. Deixando assim de lado a ficção científica transumanista e a especulação filosófica sobre a IA geral. É possível notar como as definições de IA e outros termos são objetos de disputas.

A história da IA está intimamente ligada à ciência da computação e disciplinas relacionadas, como matemática e filosofia, e, portanto, remonta pelo menos ao começo dos tempos modernos (Gottfried Wilhelm Leibniz e René Descartes, como exemplo), ou mesmo aos tempos antigos, com suas histórias sobre artesão criando seres artificiais e engenhosos em figuras animadas na Grécia antiga ou figuras mecânicas em forma humana na antiga China). Mas, como uma disciplina por si só, a IA é geralmente vista como tendo a origem na década de 1950, após a invenção do computador digital programável na década de 1940 e do nascimento da disciplina da cibernética, definida por Noberto Wiener em 1948 como o estudo científico do “controle e comunicação no animal e na máquina”.

A inteligência artificial hoje não é inteligente, não é artificial, nem objetiva e neutra, que toma decisões sem a direção humana como ressalta Kaufman. Quando pensamos na IA como um ser onipotente desprovido de qualquer contexto material, isso lhe confere uma certa qualidade mágica e teológica. Logo, os sistemas de IA “estão embutidos nos mundos social, político, cultural e econômico, moldados por humanos, instituições e imperativos que determinam o que eles fazem e como o fazem.” (CRAWFORD, 2021)[9].

[9] CRAWFORD, Kate. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven: Yale University Press, 2021.

2.2 ÉTICA E VIÉS ALGORÍTMICO: OS ALGORÍTMOS PODEM SER ÉTICOS?

Com o crescimento do uso de inteligência artificial (IA) em diversas áreas da sociedade, questões éticas têm ganhado destaque, especialmente no que diz respeito ao viés algorítmico. Esse fenômeno ocorre quando algoritmos produzem resultados tendenciosos devido às premissas utilizadas no seu desenvolvimento ou aos dados de treinamento. Essa situação levanta várias questões éticas importantes que necessitam ser abordadas para garantir que as tecnologias de IA sejam justas e não discriminatórias. Viés algorítmico acontece quando um sistema de IA reproduz preconceitos ou discriminações existentes na sociedade. Isso pode ocorrer por causa dos dados enviesados utilizados para treinar o algoritmo ou por premissas inadequadas no seu design. A falta de controle sobre esses fatores pode perpetuar desigualdades sociais em vez de reduzi-las.

Do ponto de vista ético, é primordial que os desenvolvedores e empresas responsáveis por algoritmos sejam transparentes sobre os critérios e os dados utilizados para treinar suas máquinas. A falta de transparência pode levar a uma confiança cega em sistemas que perpetuam desigualdades e dificultam a correção de erros quando eles são descobertos. Por isso, é fundamental que os responsáveis pelos sistemas de IA assumam responsabilidade por suas criações, corrigindo vieses quando identificados. Além disso, os algoritmos devem ser projetados para serem justos e não perpetuar preconceitos sociais existentes. Isso envolve a inclusão de uma variedade de perspectivas e backgrounds no processo de desenvolvimento, além de uma análise crítica dos conjuntos de dados utilizados para treinamento, que muitas vezes podem conter preconceitos históricos. Por exemplo, um estudo mostrou que sistemas de recrutamento automatizados descartavam candidaturas de mulheres para áreas de tecnologia, devido ao histórico de predominância masculina nesses cargos. Esse tipo de viés deve ser prevenido com avaliações constantes e ajustes nos algoritmos.

A ética na tecnologia também envolve minimizar danos. Algoritmos com viés podem causar impactos significativos, como no caso de sistemas de reconhecimento facial que falham em identificar corretamente pessoas de certos grupos étnicos ou algoritmos de contratação que desfavorecem candidatos com base em gênero ou idade. A responsabilidade de corrigir esses vieses, quando identificados, é fundamental. Isso pode incluir o reajuste dos modelos algorítmicos, a atualização dos conjuntos de dados ou a implementação de novos protocolos de teste para garantir a justiça. Uma parte vital da

ética no desenvolvimento de IA envolve educar aqueles que desenvolvem e implementam algoritmos sobre a importância de considerar questões de viés. Isso também se estende à sociedade em geral, que deve estar ciente de como as decisões algorítmicas são feitas e quais são seus potenciais impactos. A conscientização é crucial para garantir que os sistemas de IA sejam desenvolvidos de maneira mais justa e transparente.

Para que a inteligência artificial seja uma ferramenta de progresso para todos, é essencial que desenvolvedores, empresas e governos trabalhem juntos para criar padrões éticos que minimizem o viés algorítmico, assegurando que essas tecnologias sejam justas, transparentes e seguras para toda a sociedade. Um diálogo aberto e responsável sobre como essas tecnologias são implementadas e regulamentadas é indispensável para garantir que a IA beneficie a todos, sem perpetuar desigualdades ou discriminações.

O problema do viés está diretamente relacionado às decisões algorítmicas tendenciosas. A questão que surge é se as máquinas são capazes de valorar a realidade. De acordo com Rossetti e Angeluci, em certo sentido, elas são, pois foram programadas por seres humanos, que são, por sua vez, avaliadores intrínsecos da realidade. Dessa forma, o algoritmo em si não tem preconceito, mas acaba refletindo os valores e escolhas de seus criadores. Silva também ressalta, em seus estudos, que "problemas desses agentes artificiais, que tomam decisões de visibilidade, acesso, classificação e processamento de materiais digitais, são frequentes, muitas vezes ligados a vieses de raça, gênero, classe, localidade, neuroatipicidade e outros" (SILVA, 2020, p. 431)[10]. Assim, é o ser humano quem atribui valor à realidade.

Um exemplo de como os algoritmos refletem o viés e o preconceito humano pode ser observado em um caso citado por Rossetti e Angeluci, que relata uma pesquisa de 2015 na qual um usuário descobriu que o Google possuía um algoritmo "racista" (SALAS, 2018)[11]. A IA do Google Fotos etiquetava todos os amigos negros do usuário como "gorilas", pois não conseguia distinguir a diferença entre a pele de seres humanos e a de macacos. Esse incidente demonstra que não existe neutralidade, uma vez que os valores sociais são pré-existentes e permeiam os projetistas. Além disso, há restrições técnicas e o contexto de uso do algoritmo que condicionam e limitam sua operação. Logo, o algoritmo está sujeito a um viés social e cultural ao ser criado, como explicam Rossetti

[10] SILVA, T. Visão Computacional e Racismo Algorítmico: Branquitude e Opacidade no Aprendizado de Máquina. *Revista ABPN*, v. 12, p. 428-448, 2020.
[11] SALAS, J. Google “conserta” seu algoritmo racista apagando Gorilas. *El País*, 14 jan. 2018. Disponível em: https://brasil.elpais.com/brasil/2018/01/14/tecnologia/1515955554_803955.html. Acesso em: 30 jul. 2025.

e Angeluci. Esse é o viés ético de seu criador, transferido ao algoritmo por meio das escolhas feitas durante sua programação. Portanto, um algoritmo não é neutro, mas reflete os valores éticos de seus criadores humanos.

De acordo com a tradição filosófica que argumenta que a ética é uma característica exclusivamente humana, as coisas em si não possuem qualidades morais. Bill Gates declarou que "a tecnologia é amoral" (IHU ONLINE, 2017)[12]. Sendo assim, não seria viável discutir sobre uma ética das coisas ou das máquinas, pois, conforme a ética clássica, esse domínio é sempre humano. Segundo Lima Vaz (1999), a ética é uma análise da realidade histórico-social relacionada aos costumes tradicionais de uma sociedade e sua influência no comportamento dos indivíduos em relação aos valores e deveres de conduta.

No entanto, os desafios da IA reabrem a questão sobre a exclusividade da ética humana. Estender-se-ia aos artefatos semiautônomos feitos pelo homem? Se ampliarmos o campo da ética e aplicarmos o conceito de algoritmo, podemos obter uma "ética algorítmica", que se define como uma reflexão sobre a nova realidade histórica e social tecnológica, mediada por algoritmos. Nessa nova realidade, novos hábitos e costumes do mundo tecnológico, bem como o comportamento dos indivíduos e o funcionamento das máquinas, exigem uma revisão de seus valores e obrigações. Nesse sentido, é possível falar em ética algorítmica como uma forma de ética aplicada.

Contudo, diante do fenômeno dos algoritmos, principalmente os de aprendizado, que imitam a inteligência humana e são capazes de tomar decisões, surge uma questão ontológica sobre a natureza autônoma dos algoritmos, um feito que os aproximaria cada vez mais do humano. Rossetti e Angeluci enfatizam que essa nova realidade levanta inúmeras questões éticas relativas à transparência dos dados e do processo, à responsabilidade pelos danos causados, aos efeitos discriminatórios, à privacidade informacional, à autonomia das máquinas e dos usuários, e à fabricação de resultados tendenciosos. A ética, portanto, desempenha um papel central na garantia de que os algoritmos possam exercer seu potencial, mitigando seus riscos. Como aponta Silva (2020), "os algoritmos muitas vezes reproduzem o que já está impregnado na sociedade", perpetuando vieses sociais preexistentes.

---

[12] Dilemas morais da quarta revolução industrial. *Ihu Online*, 2017. Disponível em: http://www.ihu.unisinos.br/78-noticias/571796-os-dilemas-morais-da-quarta-revolucao-industrial. Acesso em: 30 jul. 2025.

2.3 SUPERINTELIGÊNCIA E TRANSUMANISMO: POSSIBILIDADES DE UMA IINTELIGÊNCIA ARTIFICIAL GERAL

A visão transumanista do futuro tecnológico deduz que a inteligência artificial geral (ou IA forte) é possível. Qual a possibilidade? Isso significa que é possível criar máquinas com capacidades cognitivas semelhantes às humanas? Se a resposta for *não,* então toda a visão de superinteligência é irrelevante para a ética de IA. Pois se a inteligência geral humana não é possível em máquinas, não precisa preocupar com a superinteligência.

De maneira mais ampla, a avaliação das pessoas sobre a IA parece depender do que pensam a respeito do que a IA seja e que pode se tornar futuramente, e de como pensam sobre a diferença entre humanos e máquinas. Pelo menos desde meados do século XX, filósofos e cientistas têm debatido o que os computadores são capazes de fazer e se torna, bem como quais são as diferenças entre humanos e máquinas, que será apontado mais à frente.

Os computadores podem ter inteligência, consciência e criatividade? Podem dar sentido às coisas e entender significados? Existe uma história de críticas e ceticismo em torno da possibilidade de uma IA semelhante à inteligência humana. No capítulo intitulado 'Tudo sobre os humanos' de Coeckelbergh, aborda sobre Hubert Dreysfus, um filósofo com formação em fenomenologia, que publicou um livro chamado *What Computeres Can't Do* em 1972, [O que os computadores não podem fazer].

Desde a década de 1960, Dreyfus foi muito crítico sobre a base filosófica da IA e questionou suas promessas: argumentou que o programa de pesquisa em IA estava fadado ao fracasso. Antes da sua mudança para Berkeley, trabalhava no MIT, um lugar importante para o desenvolvimento da IA, que na época se baseava principalmente na manipulação simbólica. Dreyfus ponderou que o cérebro não é um computador e que a mente não opera por meio de manipulação simbólica. O ser humano desfruta de um repertório inconsciente de saberes de senso comum baseado na experiência e no que Heidegger denominaria do "ser -no-mundo". Tal conhecimento é tácito e não pode ser formalizado. Dreyfus argumentava que, a competência humana está baseada em *know-how* (saber como) e não em *know-what* (saber o quê).

A IA não pode capturar esse significado e conhecimento de fundo; se é isso que a IA objetiva, então ela é basicamente alquimia e mitologia. Apenas os seres humanos

podem discernir o que é relevante porque, como seres corporificados e existenciais, estão envolvidos no mundo e são capazes de responder às demandas de cada situação.

Na época, Dreyfus encontrou muita oposição, porém mais tarde muitos pesquisadores de IA deixaram de prometer ou prever a IA geral. A pesquisa se afastou da dependência da manipulação de símbolos, dirigindo-se para novos modelos, incluindo aprendizado de máquina com base em estatística. E, embora nessa época ainda houvesse uma enorme lacuna entre a fenomenologia e a IA, hoje muitos pesquisadores de IA adotam abordagens corporificadas e situadas de ciência cognitiva, que afirmam estar mais próximas da fenomenologia.

Diante disso, as objeções de Dreyfus ainda são relevantes e demonstram como as visões do ser humano muitas vezes se chocam com as visões de mundo científicas, especialmente, mas não apenas na chamada filosofia ocidental. Os filósofos europeus continentais costumam enfatizar que os seres humanos e sua mente são fundamentalmente diferentes das máquinas e se concentram na experiência (auto) consciente e na existência humana, que não podem e não devem ser reduzidas a descrições formais e explicações científicas. Porém, outros filósofos, muitas vezes da tradição analítica da filosofia, endossam uma visão do ser humano que serve de apoio aos pesquisadores de IA que pensam que o cérebro e a mente humana *realmente são e funcionam* como seus modelos computacionais.

Mas nem todos os filósofos da tradição analítica pensam que a IA geral ou forte seja possível. Outra crítica filosófica muito conhecida da IA vem de John Searle, que argumenta contra a ideia de que programas de computador poderiam ter estados cognitivos genuínos ou entender significados. O experimento mental que ele propõe, chamado de argumento do quarto chinês, é da seguinte forma: Searle está trancado num quarto e recebe escritos chineses, mas não sabe chinês. No entanto, ele pode responder a perguntas feitas a ele por falantes de chinês que estão fora da sala, pois usam um livro de regras que lhe permite produzir as respostas certas *(output)* com base nos documentos (input) que recebe. Ele pode fazer isso com sucesso sem entender chinês. Da mesma maneira, argumenta Searle, os programas de computador podem produzir uma saída a partir de uma entrada por meio de regras que lhes são dadas, sem entender o que fazem.

Em termos filosóficos mais técnicos: os programas de computador não têm intencionalidade, e a compreensão genuína não pode ser gerada pela computação formal. Como Boden coloca, a ideia é que o significado vem dos humanos. No fundo dos debates sobre a IA estão, portanto, profundas divergências sobre a natureza do ser humano, sobre

inteligência humana, mente, compreensão, consciência, criatividade, significado, conhecimento humano, ciência e assim por diante. Se afinal se trata de uma "batalha", ela concerne tanto ao humano quanto à IA.

DOMINAÇÃO DAS MÁQUINAS

Nos filmes como *Matrix* (Lily e Lana Wachowski, 1999), o *hacker* Neo descobre que o mundo é uma mera simulação de computador, e que as máquinas inteligentes declaram guerra à humanidade, assim como no filme *The Terminator*, mais conhecido como *O Exterminador do Futuro*, é um filme americano (James Cameron, 1984) onde um ciborgue (androide cujo esqueleto é recoberto por tecido vivo) com inteligência artificial, é transportado no tempo de 2029 até ao dia 12 de maio de 1984, com o objetivo de alterar o curso da história e, consequentemente, o futuro. Em uma palestra em São Paulo, o filósofo francês Edgar Morin, com a lucidez de seus 97 anos, ressaltou que o papel das artes é de sensibilizar e fazer compreender. Porém, no caso da inteligência artificial, a ficção confunde mais do que esclarece. De acordo com Kaufman, "não há nenhum indício científico de um futuro com robôs subjugando humanos".

O avanço da inteligência artificial intensificou o debate sobre quais são os limites das máquinas, tanto na ficção quanto na ciência. Mas a inteligência não pode ser confundida com aleatoriedade, potencialmente responsável pela capacidade das máquinas de surpreender ao entregarem resultados distintos da programação (ou distintos da intenção inicial). Mas antes de focar no futuro da IA, é preciso atenção nos modelos estatísticos de probabilidade (modelos de *machine learning* e *deep learning),* que são a maior parte das aplicações atuais de IA, Stuart Russel, um dos mais renomados cientistas da computação, em seu livro *Human Compatible: Artificial Intelligence and the Problem of Control* (na edição brasileira, 2021, *Inteligência Artificial a Nosso Favor: como manter o controle sobre a tecnologia*) sugere pensar sobre o que significa "perder o controle" sobre as máquinas.

Para Russel, a crença de que os humanos estariam na eventualidade de perder o controle sobre "as máquinas inteligentes" baseia-se em um erro inicial na definição de IA. A princípio, as máquinas não são "inteligentes" no sentido dado pelos seres humanos – ser capaz de agir para atingir objetivos próprios -, pelo contrário, elas não possuem objetivos, são os seres humanos que atribuem seus objetivos nos sistemas inteligentes (são máquinas de otimização). No entanto, não impede que as máquinas inteligentes

encontrem soluções melhores do que os seres humanos, o que é uma realidade com as tecnologias de IA.

Posteriormente, Russel propõe abandonar a ideia de "máquinas inteligentes" em favor de "máquinas benéficas" à medida que se espera que suas ações alcancem os objetivos humanos, contribuindo para o aprimoramento da vida em sociedade. "O que temos hoje é uma relação binária, uma propriedade do sistema composto pela máquina e pelos seres humanos, com resultados superiores aos obtidos individualmente. As máquinas funcionam como uma espécie de metade de um sistema combinado com os seres humanos" argumenta.

Dois filósofos do século XX, Gilbert Simondon e Martin Heidegger, apresentaram novas perspectivas, rompendo com a supremacia interpretativa de Aristóteles. Em 1958, Simondon associou a relação com a técnica em duas atitudes culturais contraditórias: de um lado, os objetos técnicos são tratados como puro conjunto de materiais, destituído de qualquer significação, representando apenas um utilitário; e de outro, supõe-se que eles sejam animados de intenções hostis ao homem, representando uma permanente e perigosa ameaça de agressão, uma insurreição.

Trata-se da idolatria da máquina, que atribui a esses objetos uma existência separada autônoma. A segunda perspectiva iguala-se às "tecnorreligiões" apontadas por Yuval Harari em seu livro *Homo Deus*, particularmente o dataísmo, originado no Vale do Silício, em que os humanos gradativamente perdem relevância. Nem neutra nem determinista, no estágio atual a tecnologia é um elemento facilitador e alavancador de estratégicas definidas pelos humanos. Umas positivas, outras nem tanto.

As máquinas inteligentes estão sendo concebidas para, com base em grandes conjuntos de dados, "aprender" por meio de processos não totalmente explicáveis, isto é, os desenvolvedores dessas máquinas não sabem exatamente como elas aprendem para desempenhar as tarefas (a chamada "caixa-preta", que não deve ser confundida com autonomia).

Além disso, tanto os transumanistas quanto os pós-humanistas podem concordar que, em vez de competir com uma IA por determinada tarefa, pode-se também definir um objetivo comum, que é alcançado em colaboração com humanos e agentes artificiais, mobilizando o melhor que cada ser humano pode oferecer para se aproximar desse objetivo comum.

2.4 A IMPORTÂNCIA DE DISCUTIR PROBLEMAS ÉTICOS E SOCIAIS

A ética é objeto da ação humana, as tecnologias não possuem objetivos próprios nem autonomia. No caso da inteligência artificial, em seu estágio atual de desenvolvimento, em que o humano detém a prerrogativa de controle, não há como conceder a esses sistemas o status moral. A IA não tem uma ética própria, trata-se de elaborar um conjunto de melhores práticas que possa ser replicado em uma ampla variedade de configurações. O que não é nada trivial, dada a complexidade de seus sistemas.

O filosofo italiano Luciano Florido, em um artigo publicado em coautoria com o pesquisador Josh Cowls, reconhece que a inteligência artificial esteja disseminando na sociedade com potencial de aliviar ou ampliar desigualdades, e que os princípios éticos gerais propostos por instituições mundo agora são limitados como subsídios à criação de leis, regras, normas técnicas e melhores práticas.

Entre as questões éticas, talvez o maior desafio a enfrentar seja a ameaça à privacidade. Como todas as tecnologias conectadas, os dispositivos da IoB geram grandes volumes de dados, particularmente dados sensíveis, categorizados pela Lei Geral de Proteção de Dados Pessoais (LGPD) como aqueles que revelam atributos pessoais, incluindo informações genéticas, biométricas e de saúde. Essa coleta, armazenamento, uso e compartilhamento em larga escala de dados pessoais, em geral, não estão previstos nas regulamentações vigentes.

A ampliação do agenciamento da IA, principalmente quando substitui a agência humana, remete a uma questão ética cada vez mais urgente: a responsabilidade. A sociedade humana, desde Aristóteles, considera que cada um é responsável pelas consequências de seus atos. Em Ética a Nicômaco, o filósofo adiciona uma condição à responsabilidade moral, que é estar ciente do que se está fazendo, o que não é o caso dos algoritmos de IA: como agente, esses algoritmos são capazes de realizar ações e tomar decisões com consequências éticas, mas não são capazes de um pensamento moral, portanto, não podem ser responsabilizados pelas consequências de seus atos. Isso significa que, os algoritmos de IA são agentes, mas não agentes morais, porque, entre outros, carecem de consciência, emoções e sentimentos, intencionalidade. Segundo Aristóteles, somente os seres humanos realizam atos voluntários, cabendo aos humanos a responsabilidade pelos atos das máquinas. Ao delegar o agenciamento à IA, os seres humanos retêm a responsabilidade.

Conferir aos humanos a responsabilidade sobre as decisões automatizadas, ou mesmo sobre meras previsões, não é trivial. Os sistemas de IA mais complexos, portanto, com maior probabilidade de causar danos, agregam contribuições de diversos desenvolvedores, utilizam bases de dados originadas de múltiplas fontes para treinar seus algoritmos e, por fim, são aplicados para executar tarefas às vezes, em distintos domínios. A responsabilidades, desse modo, deveria ser distribuída entre todas as partes envolvidas (o que nem sempre é fácil averiguar).

A responsabilidade e, consequentemente, a explicabilidade são um dos temas éticos mais sensíveis: por um lado, têm a opacidade intrínseca de redes neurais profundas (como algoritmos correlacionam os parâmetros contidos nos dados), o que limita a capacidade dos humanos de explicar plenamente a decisão do sistema; por outro lado, dado o estágio de desenvolvimento da técnica, seus resultados refletem, em grande parte, decisões humanas, que como explicado anteriormente, são fragmentadas, dificultando uma explicação abrangente.

A privacidade é outra preocupação central na ética da IA. Com a capacidade da IA de analisar grandes volumes de dados pessoais, há um risco significativo de invasão de privacidade e uso indevido de informações sensíveis. É primordial que os sistemas de IA sejam projetados com salvaguardas robustas para proteger os dados dos usuários e garantir que esses dados sejam utilizados de forma ética e dentro dos limites legais. A consideração ética nesse contexto deve incluir não apenas a conformidade com as regulamentações existentes, mas também a antecipação de possíveis futuros dilemas éticos à medida que a tecnologia continua a avançar.

A inteligência artificial que está sendo utilizada em larga escala são modelos estatísticos de probabilidade. Esses modelos não são inteligentes, não têm capacidade de compreender o significado, não entendem o contexto social, o que se constitui em sua principal fragilidade (ou vulnerabilidade). O papel desses modelos é subsidiar o processo de tomada de decisão dos profissionais humanos.

CAPÍTULO III
DILEMAS ÉTICOS NA PRÁTICA: ESTUDOS E CASOS

Historicamente, desde a organização dos primeiros workshops, conferências e publicações acadêmicas sobre a IA, um número reduzido de pesquisadores se dedicava à análise e investigação temática da ética em tecnologia ou éticas das máquinas. Ambos os temas, *Ética em IA* e *IA ética*, por sua vez, não eram investigados ou eram a preocupação principal entre os pesquisadores da área.

Foi somente no século XXI, a partir do impacto público da IA, primeiramente na academia e posteriormente causado pelas empresas de tecnologia, o físico Stephen Hawking (Universidade de Cambridge), assim como os empresários de tecnologia Bill Gates (cofundador da Microsoft) e Elon Musk (CEO da Tesla), expressaram e anteciparam publicamente suas preocupações sobre o impacto (inclusive ético e moral) da IA sobre a vida humana, o futuro do trabalho e da sociedade.

Essa problemática evoca discussões clássicas da ética e da filosofia da tecnologia, especialmente no que diz respeito à noção de agência, responsabilidade e intencionalidade. A IA, nesse contexto, deixa de ser apenas uma ferramenta e passa a ocupar um lugar ambíguo entre objeto técnico e agente moral, desafiando os marcos tradicionais que regem a conduta humana e a imputação de culpa.

Neste capítulo, serão examinados três dilemas contemporâneos que ilustram as complexidades éticas envolvidas no uso da inteligência artificial anunciadas no capítulo I, que estabeleceu as bases teóricas, explorando a ética aristotélica e o conceito de viés algorítmico. O capítulo II delineou o panorama atual e as capacidades da IA, e este último, trará esses conceitos para a realidade.

Primeiramente, a questão dos carros autoguiados e o problema do bonde colocam em pauta a moralidade das decisões de vida ou morte tomadas por máquinas. Em seguida, será discutida a produção de conteúdo por IA e os desafios envolvendo direitos autorais, criatividade e cópia. Por fim, serão analisadas as implicações éticas mais amplas da IA na sociedade, especialmente no que diz respeito à justiça, à originalidade e à responsabilidade social.

3.1 IA E CARROS AUTOGUIADOS: O PROBLEMA DO BONDE E A RESPONSABILIDADE MORAL

Antes de abordar o complexo problemas do bonde e suas e as suas profundas implicações morais para a IA, é fundamental que antes tenha-se uma base de entendimento sobre o que são, de fato, os carros autoguiados e cimos eles operam. A ascensão desses veículos não é apenas uma revolução tecnológica no transporte, mas também um catalisador para problemas éticos que desafiam as concepções humanas tradicionais de responsabilidade e decisão. Compreender a essência da autonomia veicular é o primeiro passo para dimensionar os desafios inerentes à programações de máquinas para navegar não apenas no tráfego, mas também em cenários de vida ou morte.

A chegada dos veículos autônomos (Autonomous Vehicles – Avs) no mercado mundial é a mais impactante mudança não somente no tráfego em geral, mas também na saúde pública, com uma estimativa de 10 milhões de vidas salvas por década graças à redução de acidentes causados por motoristas humanos (90% do total). A transferência de tarefas e responsabilidade para a máquina, contudo, originou discussões éticas em vários aspectos, tendo a busca pelo mal menor em acidentes inevitáveis ganhou destaque.

Os carros autônomos são classificados em diferentes níveis, que indicam o quanto eles conseguem dirigir sozinhos. Os níveis 0 a 2 de carros autônomos já estão nas lojas. Nesses carros, o motorista ainda é o principal responsável pela direção, mesmo que o veículo tenha algumas ajudas, como freio automático ou controle de velocidade adaptativo. O nível 3 representa um grande avanço. Aqui, o carro já consegue dirigir sozinho por um tempo em certas situações, mas o motorista precisa estar sempre atento e pronto para assumir o controle se o carro pedir.

Já os níveis 4 e 5 são tecnologias de ponta, onde a intervenção humana não é mais necessária. Carros de nível 4 conseguem fazer uma viagem completa sem a ajuda de uma pessoa, mas isso só acontece em condições específicas (como em uma área pré-determinada ou em um tipo de estrada). Carros de nível 5 são totalmente autônomos. Eles podem ir a qualquer lugar e em qualquer situação sem precisar de um motorista. Esses veículos, por exemplo, nem teriam volante ou pedais.

O problema do bonde, também chamado de dilema do bonde, é um experimento mental ético, formulado pela filósofa Philippa Foot em 1967. O dilema consiste em uma situação em que o condutor de um bonde ou trem em alta velocidade está indo em direção a cinco pessoas amarradas nos trilhos. Ao seu lado está uma alavanca que pode alterar a

trajetória do bonde. Ao puxar a alavanca, o bonde mudará a sua rota para um trilho secundário e desviará das cinco pessoas, salvando suas vidas. No entanto, no outro trilho há outra pessoa amarrada. As opções, portanto, são duas: não fazer nada e assim, o bonde atropelará e matará as cinco pessoas ou puxar a alavanca, mudando a rota do bonde, que matará apenas uma pessoa.

Assim, muitas pesquisas feitas evidenciam que a automação de veículos pode ser aliada para uma condução mais segura e eficiente, trazendo assim, uma redução no número de acidentes de trânsito, redução do congestionamento nas estradas e uma melhoria da mobilidade urbana.

Diversos cenários podem ser formados para tentar buscar um consenso sobre decisões a serem tomadas no caso de acidentes inevitáveis, visando qual é a escolha mais ética, mas essa situação hipotética, por mais que seja bastante discutida, não está isenta de falhas. Uma falha perceptível é a simplificação excessiva da tomada de decisões morais. Dilemas éticos do mundo real são frequentemente complexos e envolvem uma infinidade de fatores que o problema do bonde não leva em conta. No mundo real, as decisões raramente são tão binárias, e os fatores que influenciam as escolhas são mais complicadas do que um simples cenário de um contra cinco.

Outra crítica é a falta de realismo emocional e psicológico no problema do bonde. Na realidade, os indivíduos que enfrentam tais situações provavelmente experimentariam uma gama de emoções, estresse e dissonância cognitiva. A natureza estéril e distanciada do cenário do bonde não consegue captar todo o espectro de emoções humana que desempenham um papel na tomada de decisões morais. O cenário também pressupõe um certo nível de onisciência por parte do tomador de decisões, pois espera-se que ele tenha pleno conhecimento das consequências de suas ações. Assim, problemas éticos que, de forma utilitarista, poderiam ser encarados de forma simples, tornam-se rapidamente mais complexos (GODDALL, 2014), o que faz com que não seja possível chegar a um consenso social.

O desafio para os veículos autônomos, não se resume apenas a obedecer às leis de trânsito de onde se encontram, mas também os valores universais/morais. Uma das preocupações vem justamente do fato de algumas IAs tomarem decisões de forma isolada e não permitirem a intervenção dos humanos, como em casos em que falsos positivos em sistemas de freios de emergência (quando um veículo autônomo detecta, por exemplo, uma pessoa passando à sua frente) não permitem que seus motoristas sobreporem a

aceleração ou movimentarem o volante, gerando um risco adicional por conta de colisões traseiras.

Porém, não é apenas a capacidade de tomar decisões difíceis que devem ser consideradas na ética dos veículos autônomos. Quando os próprios humanos não conseguem chegar a acordo sobre qual o quadro ético que melhor responderia ao Problema do Bonde, como é que pretendemos programar carros autónomos para ponderar compromissos como estes sob uma única ideologia? Que valores e princípios básicos deveríamos programar na IA? Deveríamos querer dar prioridade aos deveres positivos: o número de vidas salvas, ou aos deveres negativos: minimizar os danos ativos causados?

A complexidade de programar máquinas para tomar decisões morais, especialmente em cenários de acidentes inevitáveis, leva à necessidade de considerar diferentes enquadramentos éticos. As discussões filosóficas sobre o Problema do Bonde, por exemplo, muitas vezes se baseiam em duas das principais correntes da ética normativa: o Utilitarismo e a Deontologia. Ambas oferecem perspectivas distintas sobre como a IA deveria ser programada para agir, mas apresentam desafios significativos quando aplicadas ao direcionamento moral de máquinas inteligentes, como apontado por BUARQUE; CATÃO 2021. Mas no que consiste o utilitarismo? Existiriam dificuldades no enquadramento ético de decisões tomadas por máquinas? Como deve um robô se comportar quando se depara com situações de escolhas morais?

No utilitarismo, uma ação seria considerada, portanto, a depender da quantidade de prazer e dor que provoca, devendo-se priorizar aquela que propiciasse a maior quantidade de prazer possível ao maior número de pessoas. Bentham dedicou o princípio da utilidade como fundamento para a sua filosofia moral, que se fundamentava na maximização da felicidade e na análise das consequências cujo interesse está em jogo. Assim, o utilitarismo projeta, dessa forma, uma ética teleológica, cuja operacionalização deve sempre visar a maximização do prazer. (BUARQUE; CATÃO 2021)

Um sistema programado sob uma lógica utilitarista tenderia a tomar decisões que visa o maior bem para o maior número, ou seja, minimizar o número total de mortes ou ferimentos em um cenário de acidente inevitável. Por exemplo, em uma situação hipotética onde um carro autônomo se depara com a escolha entre atropelar um pedestre na pista e desviar e colidir com um grupo de cinco pedestres na calçada. Nesse cenário, a abordagem utilitarista sugeriria que o AV continuasse seu curso, atropelando o único pedestre, pois isso resultaria na menor perda de vidas.

Outro exemplo seria um AV que, para evitar uma colisão frontal com outro veículo que transporta uma família grande, decide desviar drasticamente para uma barreira rígida sacrificando intencionalmente seus próprios ocupantes (que são em menor número) para salvar um grupo de maiores pessoas. Em ambos os casos, a lógica é baseada em um cálculo frio de vidas salvas versus vidas perdidas.

Embora a minimização do dano global pareça lógica em teoria, ela pode implicar no sacrifício intencional de um indivíduo ou de um grupo minoritário para o benefício da maioria. Pesquisas, como os estudos conduzidos pelo MIT (Award et al., 2018), demonstram uma clara contradição: as pessoas geralmente esperam que os carros autônomos minimizem o número total de vítimas, mas se mostram relutantes em comprar um veículo que poderia ser programado para sacrificar seus próprios ocupantes em prol de outras vidas. Essa discrepância expõe o dilema entre o que é estatisticamente "melhor" e o que é socialmente aceitável ou moralmente esperado, dificultando a implementação puramente utilitarista.

Já na perspectiva da ética kantiana, se pauta na ideia de imperativo categórico, onde uma ação é moralmente correta quando a sua máxima possa ser universalizada para todas as pessoas, tratando-as como um fim em si mesmo. Trata-se, assim, de uma deontologia que foca na moralidade intrínseca do ato, desconsiderando as consequências que dele advêm. A ação moralmente correta torna-se, assim, um dever incondicional, e esse dever ser o motivo de sua prática, ressalta (BUARQUE; CATÃO 2021). Aplicada aos veículos autônomos, uma programação deontológica poderia se basear em princípios como "nunca causar dano intencional" ou "proteger sempre os ocupantes do veículo", independentemente do número de vidas envolvidas em outros cenários.

Por exemplo, um A programado sob uma ética kantiana não desviaria intencionalmente para atropelar um pedestre, mesmo que isso salvasse mais vidas em outra trajetória. A ação de "matar intencionalmente" seria vista como moralmente errada em si, independentemente das consequências. Outra aplicação seria a priorização absoluta da segurança do passageiro a bordo, ou a adesão estrita às regras de trânsito - como não ultrapassar o limite de velocidade ou não avançar um sinal vermelho -, pois o dever de obedecer à lei seria incondicional.

O grande desafio da deontologia reside em sua rigidez e inflexibilidade. O que acontece quando diferentes deveres entram em conflito? Por exemplo, o dever de proteger os passageiros do carro pode entrar em conflito com o dever de não prejudicar inocentes na via. Em cenários de acidente inevitável, a adesão estrita a uma regra pode,

paradoxalmente, resultar em um número maior de vítimas do que uma abordagem mais flexível. A falta de um mecanismo claro para lidar com esses conflitos entre deveres torna a aplicação puramente deontológica extremamente desafiadora para emergências onde decisões rápidas e contextuais são cruciais e não permitem soluções binárias de "certo ou errado".

A aplicação do utilitarismo e da deontologia na IA em veículos autônomos não é nada trivial, a verdadeira complexidade surge na hora de traduzir conceitos filosóficos abstratos em algoritmos e código binário.

Uma das principais dificuldades é quantificar ideias subjetivas e abstratas. Por exemplo, como um algoritmo utilitarista pode calcular "prazer" ou "dor" de forma objetiva? A vida humana tem um valor inestimável e não existe um consenso universal sobre como ponderar o "valor de diferentes vidas (jovens versus idosos, passageiros versus pedestres) ou sobre como medir a intensidade de ferimentos. Programar um sistema para fazer essas escolhas exige que se atribua um peso numérico a algo que, para humanos, é incomensurável, o que levanta profundas questões morais.

No caso da deontologia, o desafio está em definir deveres absolutos para um algoritmo que precisa operar em um mundo imprevisível. Embora a ética kantiana forneça regras morais claras, a realidade dos acidentes de trânsito é muito mais complexa do que cenários simples. É praticamente impossível prever todas as variáreis e as inúmeras nuances e uma emergência em tempo real. Como um algoritmo pode lidar com conflito entre diferentes deveres (por exemplo, o dever de proteger os ocupantes do veículo versus o dever de evitar danos a terceiros) quando não há uma hierarquia clara para esses deveres? A tradução de conceitos como "intenção" ou "dignidade", são cruciais na ética, para o código computacional é um obstáculo significativo.

Em suma, o artigo de Buarque; Catão (2021) evidencia que as filosofias éticas, por mais sólidas que sejam no campo teórico, enfrentam um abismo quando confrontadas com a necessidade de serem implementadas em sistemas autônomos. A dificuldade em prever todas as variáveis de um acidente, a complexidade em transformar valores humanos em cálculos ou regras rígidas, e a ausência de um consenso social sobre qual "moralidade" programar nos veículos, tornam o direcionamento ético de máquinas inteligentes um dos maiores desafios da era da IA.

Diante da imensa complexidade e da ausência de uma única solução ética universalmente aceita para os veículos autônomos, a comunidade global tem se dedicado intensamente a criar diretrizes e marcos regulatórios que possam guiar o desenvolvimento

e a operação da Inteligência Artificial em AVs. O objetivo principal é estabelecer um quadro ético que priorize a segurança e a dignidade humana, buscando um equilíbrio entre o avanço tecnológico e os valores morais fundamentais.

Um exemplo notável desses esforços são as Diretrizes de Ética para Veículos Autônomos da Alemanha, publicadas em 2017. Consideradas pioneiras, elas estabelecem princípios cruciais, como: a vida humana tem prioridade sobre danos materiais ou animais; em situações de risco inevitável, nenhum tipo de vida humana pode ser priorizado sobre outro com base em características pessoais (idade, gênero, deficiência, etc.); e, fundamentalmente, deve haver um princípio de "não sacrifício", significando que o veículo não pode ser programado para matar pessoas intencionalmente, mesmo que isso pudesse salvar um número maior de vidas.

Essas diretrizes alemãs refletem uma busca por um quadro ético híbrido, que tenta conciliar aspectos do utilitarismo (como a minimização de danos gerais quando possível) com princípios deontológicos (como a não instrumentalização de vidas e a proteção da dignidade individual). Elas demonstram um esforço contínuo para desenvolver uma ética aplicada que seja não apenas pragmática e tecnicamente viável, mas também socialmente aceitável e capaz de ser traduzida em padrões de engenharia. A busca por esse consenso é essencial para garantir que a revolução dos veículos autônomos ocorra de forma responsável e alinhada aos valores humanos.

3.2 IA E DIREITOS AUTORAIS: CRIAÇÃO, CÓPIA E OS LIMITES ÉTICOS

Desde algum tempo, é possível ter uma interação pessoa-máquina por meio de sistemas de linguagem natural (Chat-bot). Em muitas ocasiões x não existe a clara percepção de que esta comunicação esteja sendo feita com as máquinas e não com outras pessoas. Na década de 1970, Jacques Monod já alertava que era cada vez mais difícil estabelecer o limite entre o natural e o artificial. A simulação ou a substituição de atividades reais, cada vez mais semelhantes às realizadas por mecanismos e sistemas artificiais, gera esta ambiguidade de percepção.

A crescente sofisticação da inteligência artificial generativa (IAG) possibilitou a geração de imagens, vídeos e músicas extremamente realistas, indistinguíveis das reais à primeira vista, mas também elevando novos e complexos desafios para o conceito tradicional de direitos autorais. A base da legislação atual foi construída em um mundo

onde a criação artística e intelectual era intrinsecamente ligada à mente e ao esforço humano. No entanto, com a capacidade da IA de gerar obras que muitas vezes são indistinguíveis das criadas por humanos, somos forçados a uma reavaliação urgente de pilares fundamentais do direito autoral, como a própria definição de autoria, o que constitui originalidade, e as implicações de cópia e plágio na era algorítmica.

A expressão generativa, foi criada por Noam Chomsky, em 1965, quando propôs a "gramática generativa", ou também chamada de "gramática gerativa". Essa alternativa buscava explicar que a linguagem se estabelece por meio de competências e desempenhos para produzir frases e de quais salão os caminhos para o seu estabelecimento. Esta ideia foi baseada no pensamento de Wilhelm com Humboldt, que em 1836, disse que a linguagem faz um uso infinito de meios finitos.

A capacidade dessas IAs de produzir obras que "imitam, ou até mesmo superam, a criatividade humana" (FERREIRA, 2024, p. 02)[13] coloca em xeque as concepções tradicionais de autoria e originalidade. Alguns exemplos de IA generativa em ação incluem plataformas como DALL.E, que geram imagens a partir de descrições textuais; modelos como ChatGPT e GEMINI, que escrevem artigos, histórias e resumos automáticos; e algoritmos capazes de compor peças musicais originais. Antes, chatbots como o ELIZA (1960) operavam com respostas limitadas a scripts fixos, sem compreensão real do contexto, mas a IAG representa um salto qualitativo ao gerar conteúdo genuinamente novo e complexo.

Basicamente, os modelos de IA generativa funcionam usado redes neurais para identificar padrões de grandes conjuntos de dados e, em seguida, gerar dados ou conteúdos novos e originais. Em termos simples, redes neurais usam nós interconectados inspirados em neurônios do cérebro humano (essa é apenas uma representação simbólica, pois redes neurais são de fato operações matemáticas com matrizes, etc).

Essas redes são a estrutura dos modelos de aprendizado de máquina (Machine Learning) e aprendizado profundo (Deep Learning), que usam uma estrutura complexa de algoritmos para processar grandes quantidades de dados, como texto, código ou

[13] FERREIRA, Jorge Carlos Felz. Imagens fotográficas geradas por Inteligência Artificial - limites éticos e impacto na formação profissional. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO (INTERCOM), 40., 2024, Curitiba (PR). *Anais [...]*. São Paulo: Intercom, 2017. p. 1-15. Disponível em: https://sistemas.intercom.org.br/pdf/link_aceite/nacional/17/1001202409314966fbebb5ca714.pdf. Acesso em: 28 jul. 2025.

imagens. O treinamento dessas redes neurais envolve o ajuste dos pesos ou parâmetros das conexões entre os neurônios para minimizar a diferença entre as saídas previstas e desejadas, o que permite que a rede aprenda com os erros e faça previsões mais precisas com base nos dados.

Frequentemente, esses grandes conjuntos de dados incluem vastas quantidades de obras protegidas por direitos autorais, como textos literários, artigos científicos, imagens artísticas e composições musicais disponíveis publicamente na internet. A IA não "copia e cola" essas obras; na verdade, ela as analisa para internalizar padrões, estilos, gramáticas e vocabulários. O modelo capta a estrutura e a semântica subjacente ao conteúdo, permitindo-lhe gerar novas criações que se assemelham aos estilos absorvidos, não sendo uma reprodução direta.

Essa prática levanta sérias discussões sobre a propriedade intelectual e a originalidade, pois o uso desses dados protegidos no treinamento da IA tem sido alvo de debates intensos e processos judiciais. Como destaca Lantyer (2024), a capacidade da IA generativa de produzir conteúdo sofisticado "deriva diretamente de seu acesso irrestrito e análise de um corpus de dados que, invariavelmente, abrange uma miríade de criações intelectuais protegidas, levantando complexas questões sobre autoria e originalidade no ambiente digital."

Além disso, existe a questão da legalidade do uso de tais dados sem o consentimento explícito dos criadores, pois, embora a IA não reproduza o original, a base de seu aprendizado de certa forma é construída sob material alheio e isso coloca em xeque as noções de "uso justo" ou "fair use", porque o propósito do uso de dados pela IA não é a exibição ou distribuição do conteúdo original, mas sim a criação de um modelo que gera novas obras. O problema central é até que ponto o treinamento de IA com conteúdo protegido se enquadra no "fair use" e suas implicações legais, éticas e criativas, Lantyer (2024).

O termo "fair use" ou "uso justo" é uma doutrina legal que possibilita o uso limitado de material protegido por direitos autorais sem a necessidade de obter permissão dos detentores desses direitos. Essa doutrina é particularmente predominante no direito autoral dos Estados Unidos, mas conceitos semelhantes existem em outras jurisdições, embora com diferentes nomenclaturas e aplicações. O "fair use" é codificado na Seção

107 do Título 17 do Código de Direitos Autorais dos Estados Unidos. Essa seção oferece uma lista não exaustiva de propósitos para os quais a reprodução de uma obra protegida por direitos autorais pode ser considerada justa, como crítica, comentário, reportagem de notícias, ensino, bolsa de estudos ou pesquisa (United States, 1976). Contudo, a lei não especifica quantitativamente o que constitui "uso justo", deixando a interpretação para ser determinada caso a caso (United States, 1976).

A comunidade global e as instituições têm se dedicado a criar diretrizes e marcos regulatórios para estabelecer um quadro ético que priorize a segurança, a dignidade humana e a responsabilidade, buscando um equilíbrio entre o avanço tecnológico e os valores morais. Dentre as principais iniciativas que viam guiar o desenvolvimento e a operação da IA de forma ética, destacam-se a Recomendação da UNESCO sobre a Ética da Inteligência Artificial (2021): Um documento de grande relevância, aprovado por 193 países e que defende o uso ético da IA com base nos direitos humanos, inclusão, segurança e responsabilidade. Essa recomendação também serve como um guia para os governos e os desenvolvedores; Os princípios da OCDE sobre a Inteligência Artificial (2019): Essas recomendações internacionais tem como foco a IA confiável e promovem o bem-estar, a transparência, a segurança, a privacidade e a responsabilidade humana como pilares essenciais para o desenvolvimento de sistemas inteligentes; Diretrizes Éticas para uma IA Confiável - União Europeia (2019): Desenvolvidas por um grupo de especialistas, essas diretrizes apontam sete requisitos fundamentais para que a IA seja considerada ética, incluindo a necessidade de supervisão humana e a não discriminação, servindo como base para futuras legislações na Europa, como o AI Act e por último, a Estratégia Brasileira de Inteligência Artificial (2021): No Brasil, este documento oficial do governo define princípios e linhas de ação para promover o desenvolvimento e o uso de IA de maneira ética, responsável e inclusiva no país, alinhando-se aos debates internacionais.

É fundamental compreender que a concepção contemporânea sobre os direitos autorais não encontra um paralelo direto nas civilizações da Grécia e Roma Antigas. Embora essas sociedades sejam reconhecidas como o berço da cultura ocidental, observa-se que essas civilizações não protegiam direitos autorais como os de reprodução, circulação e execução das obras. Compreendia-se que o criador intelectual não deveria ser submeter à condição de comerciante. Contudo, mesmo nesse cenário pouco protetivo, havia uma forte reprovação, por parte da sociedade, de práticas de plágio e pirataria

(Paranaguá; Branco, 2009). A Grécia Antiga foi marcada pela tradição oral, lembra Vieira (2018). Sócrates, por exemplo, não deixou qualquer obra escrita. Já na Roma Antiga, apesar de não existir uma cultura livreira, o processo era bastante rudimentar, o que não facilitava para a noção de que obras intelectuais poderiam ser consideradas como objeto de direitos. (SARDINHA, 2023, p.05)

Nos casos em que a criação envolve o uso de IA, há autores que defendem já ser possível vislumbrar situações em que a IA cumpra os requisitos de autoria, mas essa relação ainda desperta diversos questionamentos que não podem ser facilmente respondidos pela legislação em vigor. Algumas questões vieram à tona promovendo debates, inclusive pela via judicial. Ainda não se sabe se a IA poderia ser detentora de direitos autorais; se seria necessário existir uma personalidade jurídica eletrônica, como assegurou o Parlamento Europeu; por quanto tempo será possível atribuir direitos autorais a uma IA, se ela puder designar direitos; "ou quais pessoas devem ser as titulares dos direitos autorais em obras de IA: a própria IA? A empresa desenvolvedora do software? A detentora do banco de dados? Como fica a questão do domínio público?" (SARDINHA, 2023, p.11)

Existem situações reais desses questionamentos, como exemplo o caso envolvendo Stephen Thales, que, em 2022, elaborou um pedido de reconhecimento de direitos autorais em favor da IA Creativity Machine, que desenvolveu a obra digital: A Recent Entrance to Paradise, a partir de comandos dados pelo usuário. Mas o pedido foi negado pelo Escritório de Direitos Autorais dos Estados Unidos (USCO) sob alegação da não existência da autoria humana na obra (Marzani et al., 2023). Nota-se que essa situação demonstra que a participação do usuário no direcionamento da obra foi ignorada pelo juiz que avaliou o caso, apontando a IA como autora, algo que não encontra respaldo legal no cenário atual, já que as leis presentes, normalmente, citam expressamente que o autor é uma pessoa natural.

Um outro exemplo é o da obra Teatro de Ópera Espacial feita pela IA Midjourney, que venceu um prêmio da feira de arte do Colorado. No entanto, a legislação vigente não reconhece os direitos autorais da IA, de forma que tais direitos poderiam ser usufruídos pelo ser humano que se envolveu no processo de criação, ou pela empresta que administra a IA – sendo a empresa já beneficiada diretamente pela exploração econômica do software, o que ocasionaria um enriquecimento exagerado dessas organizações. Uma

alternativa que vem sendo impulsionada é a que defende que a titularidade seja de domínio público (Marzani et al,2023).

Falando em concursos e premiações, o prestigiado Prêmio Jabuti, considerado o mais importante prêmio do cenário literário brasileiro, retirou da lista de semifinalistas o livro Frankenstein, um clássico de May Shelley, desenhado com ilustrações inseridas com o uso de IA. A obra, assinada pelo designer Vicente Pessôa, publicada pela editora Clube de Literatura Clássica em 2022, estava entre as de selecionadas na categoria de ilustração. A organização do evento informou que casos como o de uso de IA nas obras não estavam contemplados no regulamento e seriam deliberados pela curadoria, que decidiu pela exclusão da obra.

A justificativa para a eliminação da obra Frankenstein da lista de finalistas ao prêmio expôs o desconhecimento, por parte dos jurados, de estarem julgando uma obra híbrida, fato que afetaria o julgamento da criação analisada. A situação elevou o debate e fez surgir a ideia de se criar uma categoria específica de obras incluindo o uso de IA, porém, o ilustrador da obra eliminada do Jabuti que essa não é a melhor saída, pois não existem categorias distintas para técnicas diferentes como: ilustração com serigrafia, ilustração com o uso do Photoshop, com xilogravura, entre outras (Moura, 2023). De todo modo, a resolução necessita passar por ponderações éticas e que levem em consideração os direitos autorais envolvidos no processo de criação, algo que seguramente a organização do evento vai levar em conta.

Esses exemplos mostram como não há um retorno preciso para os questionamentos feitos acima, sobre a autoria de obras que envolvem o uso da IA, e ressalta o quanto esse assunto vem inquirindo as relações sociais atualmente, e, ainda, valida que haverá efeitos negativos e positivos a partir da adoção de cada linha de resposta, que vão desde “privilegiar certos grupos econômicos até o desestímulo dos avanços tecnológicos” (SARDINHA, 2023, p.12), revelando que a questão ultrapassa o campo técnico e exige uma reflexão filosófica e política mais ampla.

4. CONSIDERAÇÕES FINAIS

A presente pesquisa buscou analisar os complexos desafios adquiridos pela ascensão da Inteligência Artificial e a sua relação com a filosofia e a ética. Nesse contexto de rápidas transformações e incertezas jurídicas e morais, o papel da ética no avanço tecnológico da IA assumiu uma centralidade inegável. Não se trata apenas de definir o que a IA pode fazer, mas sim o que ela deve fazer, e como garantir que seu desenvolvimento e aplicação estejam alinhados aos valores humanos e aos direitos fundamentais.

A análise de cenários como o problema do bonde e as abordagens utilitarista e deontológica demonstrou que a tomada de decisão em sistemas autônomos e a atribuição de responsabilidade por suas ações são questões que desafiam os paradigmas morais estabelecidos. Conforme a perspectiva aristotélica, a IA não pode ser considerada um agente moral em si, pois carece de consciência, emoções e intencionalidade. Assim, a responsabilidade ética recai sobre os seres humanos que criam, programam e utilizam sistemas de IA, o que, em sistemas complexos, gera uma "lacuna de responsabilidade" devido à fragmentação das contribuições e à opacidade dos algoritmos.

Nos capítulos anteriores, foram apresentados diversos argumentos como a não neutralidade da IA, sobre como os algoritmos de IA podem reproduzir e até amplificar preconceitos presentes nos conjuntos de dados com os quais são treinados, sobre o mito de perder o controle sobre "as máquinas inteligentes" culminando em cenários de submissão ou aniquilação, entre outros que são fundamentais para demonstrar que a ética não é um adendo, mas um componente indissociável no desenvolvimento e uso da Inteligência Artificial.

A neutralidade da IA desafia a percepção comum de que a tecnologia, por ser um produto da ciência, é inerentemente objetiva e imparcial. Pelo contrário, sistemas de IA são criados por seres humanos, que possuem vieses conscientes ou inconscientes, e são treinados com dados que refletem a sociedade existente – com todas as suas complexidades, preconceitos e desigualdades históricas. Essa não neutralidade se manifesta de diversas formas como:

Os algoritmos de IA podem reproduzir e até amplificar preconceitos presentes nos conjuntos de dados com os são treinados. Isso significa que, se um algoritmo de

reconhecimento facial é treinado predominantemente com rostos de determinadas etnias, ele pode ter um desempenho inferior ou apresentar vieses discriminatórios ao analisar outras etnias. Assim como nos sistemas de IA usados em processos seletivos que podem perpetuar desigualdades de gênero ou raciais se os dados de treinamento refletirem práticas de contratação historicamente enviesadas. Essa perpetuação de preconceitos não é um mero erro técnico, mas uma consequência direta da replicação de padrões sociais existentes.

Os sistemas de IA, especialmente aqueles baseados em aprendizado de máquina, precisam de uma vasta quantidade de dados para funcionar. A coleta, armazenamento e análise desses dados, muitas vezes sem consentimento explícito e informado, interferem na autonomia individual e a na capacidade de controle sobre informações pessoais. A vigilância massiva, possibilitada pela IA, pode levar à criação de perfis detalhados de indivíduos, que podem ser usados para manipulação, discriminação ou controle social. O que prejudica o direito fundamental à privacidade.

Outro exemplo é sobre a análise da tomada de decisões autônomas por sistemas de IA, que por mais que possam processar informações e volumes inatingíveis para humanos, a ausência de transparência - chamado de "problema de caixa preta"- representa um grande problema ético, pois quando um algoritmo determina a concessão de um empréstimo, a elegibilidade para um benefício social, ou até mesmo a liberdade de um indivíduo no sistema de justiça criminal, sem que seus critérios de decisão sejam compreensível, a capacidade de contestação e de responsabilização é severamente comprometida.

Portanto, o presente trabalho buscou analisar as implicações éticas da Inteligência Artificial a partir de uma abordagem filosófica crítica, permitindo evidenciar que a apesar da crescente autonomia da IA, a responsabilidade última por suas ações recai sobre os humanos que a projetam, implementam e regulamentam, uma vez que as máquinas, desprovidas de consciência e intencionalidade própria, não podem ser consideradas agentes morais. Consequentemente, a falta de cuidado na concepção, no treinamento e na aplicação da IA não apenas reproduz, mas perigosamente alimenta as desigualdades sociais, transformando a promessa de progresso tecnológico em um espelho distorcido de nossos próprios preconceitos e injustiças, ameaçando a própria coesão de uma sociedade

que se ilude com a neutralidade de um poder que, em essência, é e sempre será intrinsecamente humano.

Uma das dificuldades que permearam a pesquisa foi o fato de o tema ser extremamente atual e ainda estar em desenvolvimento. Muita especulação ainda existe, e a materialidade completa de suas implicações ainda está por acontecer. Nesse cenário de incerteza e constante inovação, a reflexão ética emerge não como um freio ao progresso, mas como uma bússola indispensável. O papel da ética, e consequentemente dos seres humanos, é o de moldar ativamente o desenvolvimento da IA garantindo que as futuras direções tecnológicas estejam alinhadas com valores de justiça, equidade, autonomia e dignidade.

Longe de pretender exaurir o tema, este trabalho busca lançar holofotes sobre pontos específicos da discussão, destacando questões que frequentemente passam despercebidas ou são tratadas de forma acrítica.

Concluiu-se que a IA, longe de ser neutra, carrega escolhas morais e políticas que precisam ser trazidas à luz pela filosofia aplicada, pois cada linha de código, cada modelo de decisão e cada aplicação prática reflete valores, prioridades e visões de mundo de seus criadores e das instituições que a utilizam. Mais do que respostas técnicas, o debate sobre a Inteligência Artificial exige perguntas éticas. E é neste campo que a filosofia segue insubstituível, pois apenas ela nos permite compreender não apenas a técnica, mas o sentido e as consequências do seu uso. Se a IA é reflexo de nossas escolhas, pensar sobre ela é, em última em última instância, pensar sobre nós mesmos.

REFERÊNCIAS

AB2L. **Filosofia e algoritmos: o dilema moral dos carros autônomos**. AB2L, [S. l.], 14 out. 2021. Disponível em: https://ab2l.org.br/filosofia-e-algoritmos-o-dilema-moral-dos-carros-autonomos-2/?utm_source=chatgpt.com. Acesso em: 28 jul. 2025.

ARAÚJO, G. M. O. Ética Algorítmica: questões e desafios éticos do avanço tecnológico da sociedade da informação. **Galáxia**, São Paulo, n. 47, p. 1-18, 2023. Disponível em: https://www.scielo.br/j/gal/a/R9F45HyqFZMpQp9BGTfZnyr/?format=pdf&lang=pt. Acesso em: 23 jul. 2025.

BOSTROM, N.; YUDKOWSKY, E.; BATISTA, T. P. A. A ética da inteligência artificial. **Fundamento**, [S. l.], v. 1, n. 3, 30 maio 2012. Disponível em: https://periodicos.ufop.br/fundamento/article/view/2270. Acesso em: 28 jul. 2025.

BUARQUE, G.; CATÃO, A. de L. O utilitarismo e a ética kantiana: dificuldades no direcionamento moral de máquinas inteligentes. **Revista Direitos Culturais**, [S. l.], v. 16, n. 40, p. 219-228, 2021. Disponível em: https://doi.org/10.20912/rdc.v16i40.49. Acesso em: 28 jul. 2025.

CARREIRA JUNIOR, G. Ética aplicada à inteligência artificial: interfaces com a teoria de Martha Nussbaum. **REVISTA INSTANTE**, [S. l.], v. 5, n. 2, p. 160–176, 2023. Disponível em: https://revista.uepb.edu.br/revistainstante/article/view/2707.

COECKELBERGH, Mark. **Ética na Inteligência Artificial**. São Paulo: Editora Atlas, 2020.

FERREIRA, Jorge Carlos Felz. Imagens fotográficas geradas por Inteligência Artificial - limites éticos e impacto na formação profissional. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO (INTERCOM), 40., 2024, Curitiba (PR). **Anais [...].** São Paulo: Intercom, 2017. p. 1-15. Disponível em: https://sistemas.intercom.org.br/pdf/link_aceite/nacional/17/10012024093149 66fbebb5ca714.pdf. Acesso em: 28 jul. 2025.

FILHO, José Carlos Rodrigues Pinto. A imprevisibilidade aceitável na direção autônoma: porque a ausência de respostas éticas não deve impedir a adoção de carros autônomos. In: CONGRESSO DE ADMINISTRAÇÃO, SOCIEDADE E INOVAÇÃO, 10., 2018, Petrópolis (RJ). **Anais do X CASI - X Congresso de Administração, Sociedade e Inovação**. Petrópolis (RJ): FMP-FASE, 2018. Disponível em: https://www.even3.com.br/anais/xcasi/57684-a-imprevisibilidade-aceitavel-na-direcao-autonoma--porque-a-ausencia-de-respostas-eticas-nao-deve-impedir-a-adocao. Acesso em: 28 jul. 2025.

KAUFMAN, Dora. **Desmistificando a Inteligência Artificial**. São Paulo: Autêntica Business, 2020.

LANTYER, Victor Habib. Inteligência Artificial Generativa e Direito Autoral: Investigando Os Limites Do Uso Justo Na Era Da Tecnologia. **Revista de Direito, Inovação, Propriedade Intelectual e Concorrência**, Florianópolis, Brasil, v. 10, n. 1, 2024. DOI: 10.26668/IndexLawJournals/2526-0014/2024.v10i1.10352. Disponível em: https://www.indexlaw.org/index.php/revistadipic/article/view/10352. Acesso em: 28 jul. 2025.

NUSSBAUM, Martha C. A fragilidade da bondade: Fortuna e ética na tragédia e na filosofia grega. Trad.: Ana Aguiar Cotrim. São Paulo: Martins Fontes, 2009.

SARDINHA, Adriane Nascimento Celestino. Direitos autorais e inteligência artificial: uma análise do cenário atual e perspectivas para o futuro. **Res Severa Verum Gaudium**, Porto Alegre, v. 8, n. 1, p. 152-180, dez. 2023. Disponível em: https://seer.ufrgs.br/index.php/resseveraveraverumgaudium/article/download/137800/90994/600864. Acesso em: 28 jul. 2025.