



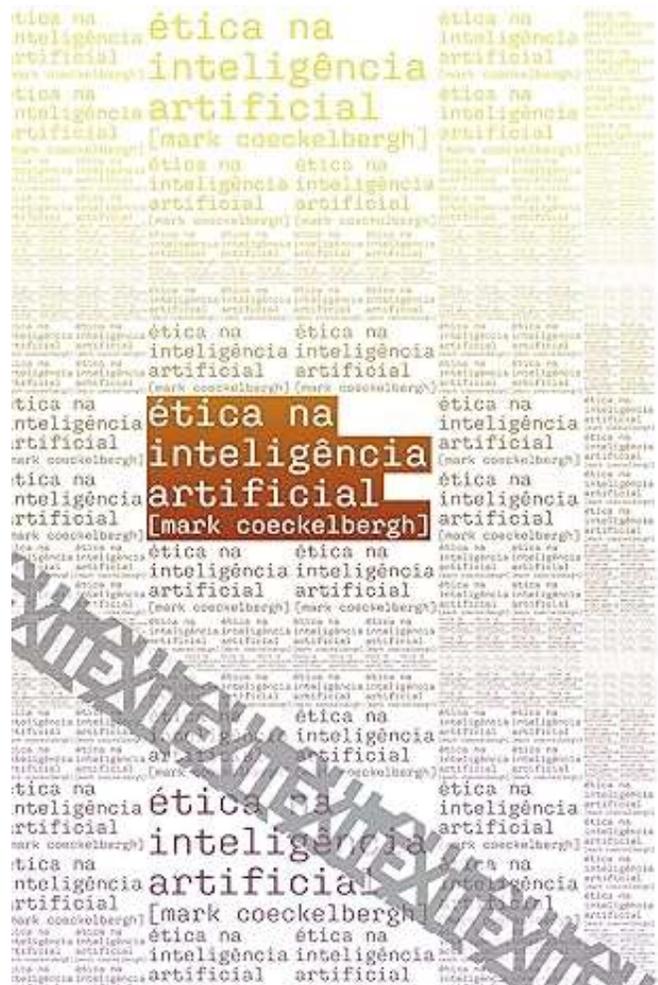
DOI: <https://doi.org/10.22484/2318-5694.2024v12id5584>

Ética na inteligência artificial

Ethics in Artificial Intelligence

Ética en la Inteligencia Artificial

Celio Aparecido Garcia – Faculdade de tecnologia de Santana de Parnaíba | Santana de Parnaíba | SP | Brasil. E- mail: celio.garcia@fatec.sp.gov.br



COECKELBERGH, Mark. **Ética na inteligência artificial**. São Paulo: UBU; Rio de Janeiro: PUC, 2023.

Recebido em: 04 out. 2024 | Aprovado em: 18 nov. 2024 | Revisado em: 10 dez. 2024
Editora de Seção: Luciana Coutinho Pagliarini de Souza | Editora de Layout: Silmara Pereira da Silva Martins



Mark Coeckelbergh publicou, em 2020, o texto IA Ethics, o qual foi traduzido para português brasileiro por Clarice de Souza *et al.* e publicado no Brasil, em 2023, com o título *Ética na Inteligência Artificial*, pelas editoras Ubu e PUC-Rio.

Com o objetivo de refletir sobre as questões éticas, os impactos econômicos, sociais e ambientais, advindos das inteligências artificiais generativas, as repercussões negativas provocadas pela falta de princípios em sistemas de IA. Coeckelbergh atinge tanto os leitores menos técnicos quanto os profissionais que desenvolvem sistemas de IA. Coeckelbergh apresenta esse contexto em 12 capítulos. Nos quatro capítulos iniciais, ele expõe os problemas filosóficos em relação à hipótese de uma inteligência artificial geral, aos seus impactos morais, sociais, econômicos e aos riscos de uma superinteligência em um futuro distante como demonstrado nas ficções científicas. A partir do quinto capítulo, até o último, são apontadas questões práticas e desafios éticos individuais e coletivos em relação às IAs generativas.

O prefácio, da edição em português, que retrata o contexto brasileiro do desenvolvimento de sistemas para Inteligência Artificial e as implicações éticas, de acordo com os editores, foi desenvolvido pelo ChaGPT, com base em um Prompt proposto, e revisado pelo autor.

Na introdução, primeiro capítulo, se acentua a soberania da IA ao vencer dois campeões mundiais, Garry Kasparov, no xadrez, e Lee Sedol, no jogo Go. Em seguida, por meio de perguntas diretas e Big Data são discutidos “o impacto real e pervasivo da IA”, e, com base no desenvolvimento de carros autônomos, abre a discussão sobre as questões éticas e sociais em relação aos impactos individuais e às transformações na sociedade e na economia.

Em seguida, amparado nos termos superinteligência e transumanismo, segundo capítulo, Coeckelbergh ressalta o entusiasmo para muitos, o pesadelo de alguns e as duas situações para outros, em relação à possibilidade de uma máquina assumir o controle. Nesse contexto, enfatiza-se que para construir novas narrativas para o futuro da IA é importante entender o porquê de certas narrativas serem dominantes, por quem são criadas e, também, menciona as responsabilidades dos cientistas pelas suas criações.

Em relação ao transhumanismo, é evidenciado o contexto histórico sobre a escatologia para explicar os medos de um final apocalíptico da sociedade ocidental, em relação ao pensamento oriental, que retrata as máquinas como colaboradoras, isto é, têm uma cultura amigável entre os humanos, os robôs e a IA. No final, deste capítulo, são sugeridas quatro possibilidades para a discussão sobre a hipervalorização apocalíptica da superinteligência.

No capítulo seguinte, terceiro, o foco é a possibilidade de criar máquinas com capacidades cognitivas semelhantes aos humanos, conforme a visão transhumanista. Todavia, com base em alguns filósofos, Coeckelbergh ressalta que a competência humana é baseada no saber como e não em saber o que, pois, os seres humanos, pelas vivências



e experiências, são capazes de responderem às demandas de cada situação. Já os sistemas computacionais podem propor soluções com base em algoritmos, sem entender o que fazem.

Na conclusão da discussão sobre humanismo, pós-humanismo e pós-fenomenologia sobressai a orientação de que a IA não precisa ser a imagem dos humanos, mas sim colaborativa. E, de acordo com Coeckelbergh, ao invés de ver a tecnologia como ameaça, ele lembra que os seres humanos são tecnológicos. Contudo, acentua que essa percepção pode parecer otimista e distante da prática científica.

Para responder à pergunta "Simplesmente máquina?", conteúdo do quarto capítulo, o filósofo expressa, por meio do conceito de moral plena sobre os animais, ou seja, aquilo que não é humano também é importante em relação à moralidade. Isso para realçar que aqueles que entendem a IA apenas como máquina cometem o mesmo erro por não valorizar o que não é humano. Como argumentado por Kant, citado por Coeckelbergh, é errado atirar em um cão, não por violar o dever para com o animal, mas por ferir as qualidades amáveis e humanas de si mesmo. E acrescenta que a suscetibilidade moral envolve o tratamento humano em relação às máquinas, ou seja, aqui é exposto o conflito entre os que destacam a possibilidade da IA ter uma agência moral plena e aqueles que defendem que uma agência moral plena é integralmente humana, pois, segundos estes, a moralidade não pode ser reduzida apenas a seguir regras.

Em relação às questões práticas, Coeckelbergh inicia essa parte com a apresentação dos conceitos; aplicações da tecnologia de inteligência artificial; as diferentes abordagens e subcampos e as aplicações de impactos.

Com base em seu contexto histórico, quinto capítulo, a IA é definida tanto como uma ciência quanto uma tecnologia que pode auxiliar no entendimento do ser humano e de outros seres que com inteligência natural. Em seguida, Coeckelbergh ressalta a relação da IA com várias ciências e disciplinas, por exemplo, IAs que interagem com a neurociência, a robótica e Big Data. Outro aspecto abordado nessa conjuntura são as diversas formas de incorporação das IAs, que não estão limitadas aos robôs humanóides. Por isso, como apontado por ele, se faz necessário entender como os algoritmos de IA funcionam e o que eles podem fazer. Pois, os diferentes paradigmas de pesquisas; os subcampos; o aprendizado de máquina e o aprendizado profundo criaram tipos de IA que superaram a IA simbólica.

Para finalizar, nesse ponto, Coeckelbergh expõe os benefícios da IA na agricultura, na educação, na saúde, para citar alguns exemplos, e os impactos negativos. No entanto, salienta a importância de avaliar, em relação à tecnologia de IA, os aspectos éticos e as consequências sociais.

Na sequência, o sexto capítulo tem como base o aprendizado de máquina, supervisionado, o não supervisionado e o aprendizado por reforço, ainda que não conceituado como aprendizado, pela falta de uma real cognição, podem, pelo processo



estatístico, serem utilizados em várias tarefas, porque os algoritmos conseguem identificar padrões para explicar dados e fazer previsões, que, em algumas análises, superam as regras dadas pelo programador. E, devido ao custo reduzido, o acesso e o armazenamento dos dados auxiliam no desenvolvimento e treinamento de IA e da ciência de dados.

Embora a ciência e dados extraiam padrões significativos que favorecem a tomada de decisões com inúmeras aplicações, conforme Coeckelbergh, a participação humana é necessária, pois as escolhas envolvem um componente ético e a privacidade de dados.

A respeito da privacidade de dados, exposta no sétimo capítulo, Coeckelbergh ressalta a falta de transparência e o controle exercido por algumas empresas de tecnologia por meio de recursos da IA. No entanto, ele afirma que essa problemática pode ser resolvida com o exemplo da ética na pesquisa. E acrescenta que, sem a devida transparência em relação aos dados dos usuários, pode ocorrer a manipulação e a exploração de usuários vulneráveis, uma espécie de mão de obra não remunerada.

Nesse cenário o é ressaltado formas de dominação e exploração de consumidores, tal como a vigilância e totalitarismo em sociedades sem políticas totalitárias. Acrescenta, também, que a IA pode ser utilizada para manipular a política de forma direta, por exemplo, as fake news, o discurso de ódio e informações falsas. Outro aspecto de destaque, nessa parte, é a ilusão de companheirismo da IA que tem desestabilizado as relações pessoais.

Com relação ao respeito e à privacidade é concluído com a explanação sobre segurança e proteção em um ambiente compartilhado com máquinas e seres humanos, como pressuposto para discutir outro problema ético: a responsabilidade.

Os conceitos filosóficos, com base no texto *Ética a Nicômaco*, de Aristóteles, o autor reforça, no oitavo capítulo, que o ser humano é responsável pelas suas decisões, todavia, ele faz o seguinte questionamento: o que ocorre com a IA são decisões? Nessa circunstância uma IA não pode ser agente moral, porque não tem consciência, livre-arbítrio, emoções e atitudes. Dessa forma, faz-se necessário a transparência e explicabilidade. Os seres humanos têm como base as explicações sociais pautadas nos valores, no entanto, independentemente se a IA pode ou não fornecer explicações, os seres humanos, segundo Coeckelbergh, devem responder a seguinte pergunta: Por quê? Sobretudo em situações que podem surgir o viés.

Para desenvolver o tema do enviesamento de dados, capítulo nono, é retomado o caso do COMPAS, que auxilia na tomada de decisões sobre a liberdade condicional e o PredPol, utilizado para prever a probabilidade de crimes. No entanto, esses sistemas podem ser afetados pela falta de diversidade entre os desenvolvedores de IA e o fato da aprendizagem de máquina se basear em informações que incluem a cultura humana cotidiana e os preconceitos. Por isso, Coeckelbergh chama a atenção para os riscos de confiar demais na tecnologia. Ele acrescenta que a questão do viés não é apenas técnica,



mas política e filosófica, pois envolve justiça e equidade sobre os humanos e a sociedade. Como o futuro do trabalho, que no contexto capitalista dá sentido à vida.

Fundamentado nessas questões, este conteúdo é finalizado com duas perguntas: o que significa tomada de decisão democrática em relação à IA? E, qual é o conhecimento necessário, sobre IA, por parte dos cidadãos e das políticas?

No que diz respeito às propostas de políticas, no décimo capítulo, acentua o que é essencial e que os formuladores de políticas precisam responder sobre o que deve ser feito, por quê, quando, quanto, por quem, a natureza, a extensão e a urgência do problema. Após um detalhamento desses tópicos, são apresentados, de forma sequencial, algumas iniciativas que visam os princípios éticos e as justificativas, com respeito às diferenças culturais, como a "IA confiável". Do mesmo modo, salienta que as políticas sobre IA não são de responsabilidade exclusiva dos comitês e órgãos governamentais; por isso os acadêmicos, algumas empresas e organizações não governamentais têm tomado algumas iniciativas para as soluções tecnológicas, métodos e operacionalização. Coeckelbergh, ainda, menciona o hiato entre citar os princípios éticos e descobrir como executá-los na prática. Pois, isso exige uma ética proativa com a incorporação de valores.

No que se refere à ética proativa, penúltimo capítulo, é ressaltado a necessidade de pensar os conceitos éticos no início do projeto para evitar problemas após a concepção de sistemas de IA. Isso, com base em uma governança inclusiva mediante o envolvimento dos interessados, debate público e intervenção social. Só assim, acrescenta, será construída uma inovação responsável. Já que, no entendimento de Coeckelbergh, os documentos de políticas de IA são desenvolvidos de cima para baixo e são abstratos. Nesse ínterim a inovação responsável enfrenta as mesmas condições de um acordo para evitar o aquecimento global e outros problemas que afligem humanos e não humanos.

Com a afirmação de que a ética não é sobre banir as coisas e, em seu aspecto positivo, o filósofo mostra que as questões éticas em IA não são apenas tecnológicas porque se referem, também, à vida e à prosperidade humanas e todo o necessário para a sobrevivência humana, não humana e ao futuro do planeta. E, para esse desenvolvimento, se faz necessário e interdisciplinaridade e a transdisciplinaridade entre as pessoas das humanidades, das ciências sociais, das ciências naturais, que precisam entender e utilizar os recursos tecnológicos, e as pessoas da engenharia, cientista e engenheiros, que precisam de mais sensibilidade em relação aos aspectos éticos e sociais no desenvolvimento e uso da tecnologia para não incentivarem a alienação.

No último capítulo, Coeckelbergh discorre, com base em um questionamento sobre a centralidade da ética na IA no ser humano, sobre os desafios e as prioridades, isto é, quais são as repercussões da IA em outros seres vivos e a tensão entre valores e interesses humanos e não humanos. Depois, enfatiza que a tecnologia deve beneficiar os seres humanos e servi-los, e não ao contrário. Ainda, nesse contexto, ele aponta vários problemas relacionados à sustentabilidade e as implicações entre a IA, mudanças climáticas e o antropoceno, uma vez que não são apenas os humanos que precisam ser



gerenciados, pois, o planeta também precisa ser reprojetoado. E, questiona se a IA distrai a sociedade desses assuntos importantes, sobretudo, o fato de destinarem energias e investimentos para o desenvolvimento que busca, com os recursos da IA, a sobrevivência em outro planeta. Isso pode provocar o abandono da condição existencial, corpórea, terrestre e dependente. Assim como as discrepâncias socioeconômicas.

Nas palavras de Coeckelbergh a matemática e sua utilização nas tecnologias são ferramentas importantes, mas não conseguem entender os problemas humanos e lidar com eles. A IA é eficiente em edificar padrões, no entanto, de acordo com as reflexões de Coeckelbergh, não se atribui sabedoria às máquinas.

Mark Coeckelbergh no texto **Ética na Inteligência Artificial**, com base em suas competências filosóficas, tecnológicas e exemplos práticos, disserta sobre temas complexos e polêmicos, mas emprega uma linguagem que possibilita a compreensão tanto do público não especializado em tecnologia quanto os desenvolvedores de tecnologias em IA. Este texto não se delimita apenas a propor regras em relação à ética na Inteligência Artificial, mas promover uma reflexão sobre as prioridades humanas, a inteligência humana e a inteligência artificial.

Em relação à educação, a interação humana com a Inteligência Artificial, é evidente as contribuições filosóficas e práticas proposta por Coeckelbergh, porque, nessa área, se faz necessário ampliar as discussões e as práticas para fomentar as regulamentações e formas de desenvolvimento intelectual em um cenário que privilegia a técnica e o desenvolvimento, porém sem as discussões necessárias sobre os possíveis aspectos negativos da IA na educação. E, para facilitar esse processo, as grandes organizações de tecnologia, que disponibilizam ferramentas de Inteligência Artificial, precisam de mais transparência nos processos e aproximação entre as inteligências individual, social e artificial. Isso posto, essa obra é indicada àqueles que desejam entender os impactos sociais, econômicos e éticos desse admirável mundo novo.

Mark Coeckelbergh fez a graduação em ciências sociais, em 1997, e, no mesmo ano, obteve o título de mestre em ciências políticas na Universidade de Leuven. Em 1999, concluiu o mestrado em filosofia social pela Universidade de East Anglia, o doutorado, em filosofia, concluído em 2003, na Universidade de Birmingham, também no Reino Unido. Desde 2015 é professor titular de Filosofia da Mídia e Tecnologia no Departamento de Filosofia da Universidade de Viena, na Áustria. Também contribui com vários projetos em Filosofia, tecnologia, inteligência artificial e ética.