KAUFMAN, Dora. Resenha de *Ética na inteligência artificial*, de Mark Coeckelbergh. TECCOGS – Revista Digital de Tecnologias Cognitivas, n. 28, 2023, p. 151-155.

Recebido em: 10 fev. 2024
Aprovado em: 15 mar. 2024

dx.doi.org/10.23925/1984-3585.2023i28p151-155

Resenha de *Ética na inteligência artificial*, de Mark Coeckelbergh[1]

Dora Kaufman[2]

A ética em pauta

Mark Coeckelbergh é um autor contumaz, com valiosas contribuições ao debate contemporâneo. No livro *Green Leviathan or the Poetics of Political Liberty: Navigating Freedom in the Age of Climate Change and Artificial Intelligence,* de 2021, Coeckelbergh recorre à ficção para refletir sobre as mudanças climáticas em um mundo mediado pela inteligência artificial (IA). Libertado da prisão após 20 anos, militante de um "futuro verde" constata encantado que seu sacrifício não foi em vão: a humanidade e a natureza foram salvas! A catástrofe climática, contudo, foi evitada com a IA controlando e manipulando o comportamento humano, ou seja, o custo foi a liberdade. Abordando diversos temas relacionados à crise climática, o autor deixa claro, contudo, que a solução não é a supressão da democracia.

No ano seguinte, 2022, Coeckelbergh publicou dois livros. Em *The Political Philosophy of AI* (2022, MIT Press), Coeckelbergh pondera que a IA coloca em xeque os significados de conceitos políticos-filosóficos tradicionais como liberdade, igualdade, democracia e poder, diante de "uma forma de capitalismo bruto, que expropria a

1 COECKELBERGH, Mark. *Ética na inteligência artificial*. Tradução: Clariss de Souza *et al*. São Paulo / Rio de Janeiro: UB / Editora PUC-Rio, 2023.

2 Dora Kaufman é professora do Programa Tecnologias da Inteligência e Design Digital da PUC SP. Doutora pela ECA-USP com estágio sanduíche na Université Paris Sorbonne IV, possui dois pós-doutorados: na COPPE-UFRJ e no TIDD-PUC SP. Autora de vários livros, entre eles *A inteligência artificial irá suplantar a inteligência humana?* e *Desmistificando a inteligência artificial*. Colunista da Época Negócios e colaboradora do Globo e do Valor Econômico. ORCID: https://orcid.org/0000-0001-7060-4887

experiência humana e impõe um novo tipo de controle: concentração de conhecimento significa concentração de poder", com as grandes corporações moldando o futuro da sociedade.

O argumento central do livro *Self-Improvement: Technologies of the soul in the age of artificial intelligence* é que o autoaperfeiçoamento não é mais opcional, dado que os algoritmos de IA são parte essencial da cultura de autoaperfeiçoamento, além de um imperativo na medida em que as tecnologias digitais não nos oferecem apenas informações, mas nos convidam a nos comparar uns com os outros. Essa constante comparação "nos arrasta para regimes de autodisciplina, autovigilância e conhecimento quantitativo duros e insuportáveis". Com a IA, nosso self é quantificado, tornando-se um "data-self". Coeckelbergh advoga a favor de modelos que incentivem a percepção de que o crescimento individual sadio é relacional, transformando os "outros" não em competidores, mas em parceiros.

Poucos temas escapam à investigação e reflexão de Coeckelbergh. No artigo *Can machine create art?*, de setembro de 2017, por exemplo, ele oferece uma estrutura conceitual para o debate filosófico sobre o status da arte da máquina e da criatividade da máquina, defendendo que devem ser consideradas formas não humanas de criatividade. Seria o algoritmo de IA um agente artístico? Coeckelbergh nos faz refletir sobre a natureza da arte e da criatividade humanas.

Entretanto, a contribuição mais efetiva de Coeckelbergh talvez esteja no livro *Ética na inteligência artificial*, publicado em primeira edição na "Essential Knowledge Series" da MIT Press em 2020. O livro aborda temas desde o histórico ao cultural e técnico da IA, contendo um bom número de referências, mesmo que parte dessas referências não sejam devidamente comprovadas pelos autores citados, comprometendo os argumentos construídos a partir delas. Coeckelbergh, como professor de filosofia da mídia e da tecnologia na Universidade de Viena, tem legitimidade para versar sobre uma visão geral da ética da IA; além de professor, ele é membro de vários conselhos consultivos de ética para robótica e IA, incluindo conselhos consultivos políticos como o "High-level Expert Group on Artificial Intelligence" constituído pela Comissão Europeia no início do processo de regulamentação da IA em 2018.

O livro inicia com uma discussão sobre a possibilidade da IA geral, ou seja, sistemas de IA com cognição no nível humano, na qual o autor não assume uma posição explícita, mas apresenta diversas visões filosóficas com o pressuposto de que no fundo do debate sobre IA estão

divergências profundas sobre a natureza do ser humano, inteligência humana, mente, compreensão, consciência, criatividade, significado, conhecimento humano, e ciência, portanto com ligações a muitas disciplinas, incluindo matemática, engenharia, linguística, ciência cognitiva, ciência da computação, psicologia e até mesmo filosofia. Para ele, tanto o filósofo quanto o cientista de IA estão interessados em compreender a mente e fenômenos como inteligência, consciência, percepção, ação e criatividade. São abordadas as mudanças tecnológicas e seu impacto na vida das pessoas, além de transformações na sociedade e na economia, com base no entendimento de que a tecnologia é sempre social e humana, pois seu uso ocorre em um contexto social. Neste sentido a inteligência artificial não é apenas sobre tecnologia, mas igualmente sobre o que os humanos fazem com ela, como a usam, como a percebem e experimentam, e como a inserem em ambientes técnico-sociais mais amplos. Ou seja, o autor sugere pensar a ética, que diz respeito às decisões humanas, em uma perspectiva histórica e sociocultural.

Em seguida, Coeckelbergh aborda a questão do estatuto moral da IA, os temas do "agenciamento" e da "responsabilização" procurando descrever os argumentos que entrelaçam o debate sem tomar uma posição específica. Os sistemas de IA recebem o atributo de "agenciamento" no sentido de que executam ações no mundo, e essas ações têm consequências morais. Contudo, isso não significa necessariamente atribuir à IA o status de "agente moral". O autor agrega diversas visões filosóficas sobre a atribuição de "agente moral" aos sistemas ou algoritmos de IA, desde questionamentos sobre a capacidade necessária para a agência moral, salientando que a IA é produzida e usada por humanos, portanto, a tomada de decisões morais em práticas tecnológicas é da competência humana.

Do outro lado do espectro, estão os que pensam que as máquinas podem ser agentes morais semelhantes aos seres humanos, inclusive afirmam que é possível e desejável dar às máquinas um tipo de moralidade humana, que podem até ser melhores do que os seres humanos no raciocínio moral pois são mais racionais e não influenciáveis por emoções (posição que pressupõe, do ponto de vista do autor, uma ideia equivocada a respeito da natureza da moralidade ao reduzi-la a seguir regras com o risco de gerar uma "IA psicopata", perfeitamente racional e insensível às preocupações humanas porque carece de emoções). Por essas razões, pondera o autor, poderíamos rejeitar a própria ideia de agência moral plena, ou poderíamos tomar uma posição intermediária, a saber, temos que dar à IA algum tipo de moralidade, algo como "moralidade funcional"

proposta por Wendell Wallach e Colin Allen ou uma "moralidade estúpida", não baseada nas propriedades humanas como propõem Luciano Floridi e J.W. Sanders.

Outra alternativa seria tornar a agência moral dependente de um nível mínimo de interatividade, autonomia e adaptabilidade, ou aplicar critérios não antropocêntricos para o agenciamento moral. De qualquer forma, a ênfase é em uma abordagem do status moral relacional e socialmente enraizada, que não é nem abstrata nem formalizante, nem é baseada em atitudes superiores e hegemônicas. A ética da IA, consequentemente, nos obriga a reconsiderar as nossas atitudes morais peculiarmente humanas e a questionar a própria natureza humana e o nosso futuro.

Para Coeckelbergh a atribuição de agente moral está diretamente associada à responsabilidade: se você tem efeito no mundo e nos outros, você é responsável pelas consequências. Recorrendo à "condição de controle" de Aristóteles, o autor argumenta que a IA pode realizar ações e tomar decisões que têm consequências éticas, mas não está ciente do que faz e não é capaz de pensamento moral e, portanto, não pode ser moralmente responsável pelo que faz. "As máquinas podem ser agentes, mas não agentes morais, uma vez que carecem de consciência, livre arbítrio, emoções, a capacidade de formar intenções e assim por diante", logo o recomendado é preservar para os humanos a responsabilidade sobre as consequências.

Responsabilidade significa responsabilização e explicabilidade, se algo der errado precisamos de uma resposta e de uma explicação, contudo essas exigências não são factíveis com a natureza das redes neurais ("black box", não explicabilidade, opacidade e outros termos que definem a não transparência de como as redes neurais chegam ao resultado). Diante da limitação técnica dos sistemas de IA baseados em redes neurais, Coeckelbergh pondera que "não se trata principalmente de explicar 'como funciona', mas de como eu, como ser humano de quem se espera que seja responsável e aja com responsabilidade, posso explicar minha decisão", demanda cujo grau de sensibilidade é função do setor e da tarefa a ser empreendida pelo sistema de IA.

Em seguida, Coeckelbergh discute os impactos reais da IA atual (IA estreita) como privacidade e proteção de dados, manipulação, desinformação, totalitarismo, segurança e proteção, preconceito, futuro do trabalho e mudanças climáticas. Adicionalmente, versa sobre alguns aspectos do processo regulatório europeu da IA e, por fim, pondera sobre a abordagem antropocêntrica da IA.

Para o autor, a ética não deve ser vista como um tópico marginal que tem pouco a ver com sua prática tecnológica, mas como parte essencial dela, contemplando o acúmulo de riscos tecnológicos e o consequente crescimento das vulnerabilidades humanas, sociais, econômicas e ambientais. A ética precisa ser levada em consideração no estágio inicial do desenvolvimento da tecnologia, denominada de "ética by design", superando o caráter vago e abstrato dos princípios éticos, enfrentando o desafio de construir uma ponte entre esses princípios éticos e legais abstratos generalistas e as práticas de desenvolvimento e uso da tecnologia em contextos específicos. Uma barreira a ser enfrentada é a falta de interdisciplinaridade e transdisciplinaridade, consequência da lacuna significativa na formação e no entendimento entre, por um lado, especialistas das ciências humanas e sociais e, por outro lado, especialistas das ciências naturais e da engenharia, tanto dentro quanto fora da academia: especialistas com formação em ciências humanas precisam se conscientizar da importância de pensar sobre as novas tecnologias, como a IA, e adquirir conhecimento básico dessas tecnologias e do que elas fazem. Por outro lado, cientistas e engenheiros precisam ser mais sensíveis aos aspectos éticos e sociais do desenvolvimento e uso da IA.

Com a premissa de que a tecnologia não é apenas um instrumento nem tampouco neutra, ou um recurso externo, mas molda nossas ações e nossas narrativas, Coeckelbergh defende que, se queremos um futuro para a IA diferente, precisamos de histórias diferentes e tecnologias diferentes. Se não gostamos de uma história específica sobre a IA, o que precisamos não é apenas rejeitar a história (muito menos ignorá-la), mas reescrever a história ou escrever uma nova história.