

ética na  
inteligência  
artificial  
[mark coeckelbergh]

tradução  
Clarisse de Souza  
Edgar Lyra  
Matheus Ferreira  
Waldyr Delgado

Dados Internacionais de Catalogação na Publicação (CIP)  
Bibliotecário Vagner Rodolfo da Silva – CRB 8/9410

C672e Coeckelbergh, Mark [1975-]

Ética na inteligência artificial / Mark Coeckelbergh;  
título original: *AI Ethics*; traduzido por Clarisse de  
Souza et al.

São Paulo / Rio de Janeiro: Ubu Editora / Editora PUC-Rio,  
2023 / 192 pp. / Coleção Exit  
ISBN UBU 978 85 7126 124 2  
ISBN PUC-RIO 978 85 8006 310 3

1. Tecnologia. 2. Inteligência artificial. 3. Big Techs.  
4. Internet. 5. Ética. I. Souza, Clarisse de. II. Lyra, Edgar.  
III. Ferreira, Matheus. IV. Delgado, Waldyr v. Título  
VI. Série

2023-3036 CDD 006.3 CDU 004.81

Índice para catálogo sistemático:

1. Tecnologia: Inteligência artificial 006.3
2. Tecnologia: Inteligência artificial 004.81

fonte Edita e Aglet Mono  
papel Alta alvura 90 g/m<sup>2</sup>  
impressão Margraf

Hofstadter. Trata-se de um legado que qualquer discussão sobre ética em IA precisa considerar, mas que, como todo debate filosófico, é denso, complexo e de difícil acesso. É esse o principal mérito desta obra, de levar em conta os debates que a precederam e fazer uma discussão sobre as principais questões éticas envolvendo a IA de forma ampla, mas acessível; abrangente, mas não superficial. Mais ainda, este livro introduz as ideias básicas das tecnologias envolvidas, para situar o leitor menos técnico no contexto que estamos, sem nunca perder de vista possíveis impactos negativos do seu uso. Ao mesmo tempo, para os desenvolvedores de sistemas de IA, para os jovens hackers de Python e TensorFlow, o livro traz de forma sucinta questões éticas fundamentais que todos que trabalham com IA diretamente, mãos na massa, devem conhecer.

Não podemos decidir o futuro de sistemas inteligentes sem um debate consciente e bem fundamentado sobre seu uso ético e sobre a miríade de possibilidades que essa tecnologia nos abre. E esse debate passa pelo entendimento, por todas as partes, dos conhecimentos e ideias contidos neste livro, brilhantemente apresentados por Coeckelbergh para um público técnico e não técnico.

Parem as máquinas! As da euforia e da catástrofe em IA, e leiam este livro.

Claudio Pinhanez

Pode, entretanto, surgir a preocupação de que as visões pós-humanistas e pós-fenomenológicas não sejam críticas o suficiente, porque são muito otimistas e muito distantes da prática científica e de engenharia, portanto insuficientemente sensíveis aos perigos reais e às consequências éticas e sociais da IA. Atravessar fronteiras nunca antes cruzadas não é necessariamente algo isento de problemas e, na prática, essas ideias pós-humanistas e pós-fenomenológicas podem ser de pouca ajuda contra a dominação e a exploração que podemos enfrentar vindas de tecnologias como a IA. Pode-se também defender uma visão mais tradicional do humano ou clamar por um novo tipo de humanismo, em vez de pós-humanismo. Por isso, o debate continua.

## [4] Simplesmente máquinas?

Questionando o status moral da IA: agência moral e suscetibilidade<sup>1</sup> moral

Um dos assuntos que surgiram no capítulo anterior foi se não humanos também são importantes. Hoje muitas pessoas acham que os animais importam, moralmente falando. Mas isso nem sempre foi assim. Aparentemente, estávamos errados sobre os animais no passado. Se muitos hoje acham que as IAs são apenas máquinas, estão cometendo erro similar? As IAs superinteligentes, por exemplo, mereceriam status moral? Deveriam ser-lhes concedidos direitos? Ou é uma ideia perigosa até mesmo pôr em questão se máquinas podem ter status moral?

Um modo de discutir o que uma IA é e o que ela pode se tornar é perguntar sobre o seu status moral. Nesse ponto tocamos questões filosóficas com relação à IA, não por meio da metafísica, da epistemologia, da história das ideias, mas pela filosofia moral. O termo *status moral* (também chamado de *posição moral*) pode se referir a dois tipos de questões. A primeira se ocupa do que a IA é capaz de fazer moralmente falando – em outras palavras, se ela pode ter o que os filósofos chamam de *agência moral*, e se, então, poderia ser um *agente moral pleno*. O que isso significa? Parece que as ações das IAs hoje já têm consequências morais. A maioria das pessoas concordará que a IA tem uma forma “fraca” de agência moral nesse sentido, o que é similar, por assim dizer, à maio-

<sup>1</sup> O termo *moral patiency* foi traduzido por “suscetibilidade moral” para não ser confundido com a virtude da *paciência* em relação a questões morais. O sentido dado pelo autor é o da possibilidade ou ato de ser afetado moralmente. [N.T.]

ria dos carros de hoje: estes também podem ter consequências morais. Mas, considerando que a IA está se tornando mais inteligente e autônoma, pode vir a ter uma forma mais forte de agência moral? Ela deveria ganhar, ou desenvolver, alguma capacidade de raciocínio, julgamento ou decisão de caráter moral? Por exemplo, carros autônomos que usam IA podem ou deveriam ser considerados agentes morais? Essas questões são sobre ética na IA, no sentido de: *Que tipo de capacidades morais uma IA tem ou deve ter?* Mas questões sobre o “status moral” também podem se referir a como nós devemos tratar uma IA. A IA é “só uma máquina”, ou ela merece alguma forma de consideração moral? Deveríamos tratá-la de modo diferente, por exemplo, de uma torradeira ou de uma máquina de lavar? Teríamos de conferir direitos a uma entidade artificial altamente inteligente, se tal entidade for algum dia desenvolvida, mesmo não sendo humana? Isso é o que filósofos chamam de questões relacionadas à *suscetibilidade moral*. Essa questão não é sobre a ética na IA ou por ela, mas sobre *nossa ética para com a IA*. Aqui, a IA é objeto de preocupação ética, em vez de um potencial agente ético em si.

## Agentes morais

Vamos começar com a questão da agência moral. Se uma IA fosse mais inteligente do que hoje é possível, podemos supor que ela seria capaz de desenvolver raciocínio moral e de aprender como seres humanos tomam decisões sobre problemas éticos. Mas seria isso suficiente para uma agência moral plena, isto é, para um agente moral como o ser humano? A questão não é inteiramente uma ficção científica.

Se hoje já delegamos algumas de nossas decisões a algoritmos, por exemplo, em carros ou tribunais, então parece que seria importante tais decisões serem moralmente robustas. Mas não está claro se máquinas podem ter as mesmas capacidades morais de humanos. É dada agência a elas no sentido de que fazem coisas no mundo e essas ações possuem consequências morais. Por exemplo, um carro autônomo está sujeito a causar um acidente, ou uma IA pode recomendar que determinada pessoa seja presa. Esses comportamentos e escolhas não são moralmente neutros: há evidentemente consequências morais para as pessoas envolvidas. Mas, para lidar com esse problema, deveríamos conceder agência moral às IAs? Elas podem ter agência moral plena?

Há várias posições filosóficas sobre essas questões. Alguns dizem que máquinas jamais poderão ser agentes morais. Máquinas, argumentam, não possuem capacidades necessárias para agência moral, tais como estados mentais, emoções ou livre arbítrio. Portanto, é perigoso supor que elas possam tomar decisões morais robustas, delegando totalmente tais decisões morais a elas. Por exemplo, Deborah Johnson<sup>2</sup> argumentou que sistemas computacionais não possuem agência moral por si só: são produzidos e utilizados por seres humanos, e somente esses humanos possuem liberdade e são capazes de agir e decidir moralmente. Do mesmo modo, pode-se dizer que as IAs são construídas por seres humanos e que por isso a tomada de uma decisão moral em práticas tecnológicas deve ser executada por seres humanos. Do outro

<sup>2</sup> Deborah G. Johnson, "Computer Systems: Moral Entities but not Moral Agents". *Ethics and Information Technology*, v. 8, n. 4, pp. 195-204, 2006.

lado do espectro estão os que pensam que as máquinas podem ser agentes morais plenos, assim como os seres humanos. Pesquisadores como Michael e Susan Anderson, por exemplo, alegam que, em princípio, é possível e desejável dar a máquinas um tipo humano de moralidade.<sup>3</sup> É possível dotar as IAs de princípios, e as máquinas chegariam a ser até melhores do que os seres humanos em raciocínio moral, pois elas são mais racionais e não se deixam levar por emoções. Contra essa posição, alguns têm argumentado que regras morais frequentemente entram em conflito (considere, por exemplo, as histórias de Asimov sobre robôs, nas quais as leis da robótica<sup>4</sup> sempre põem robôs e seres humanos em apuros) e que o projeto geral de construir "máquinas morais" dando a elas regras é baseado em suposições equivocadas sobre a natureza da moralidade. A moralidade não pode ser reduzida a seguir regras e não é integralmente uma questão de emoções humanas – embora estas últimas possam muito bem ser indispensáveis para julgamentos morais. Se uma IA geral é realmente possível, então não havemos de querer um tipo de "IA psicopata" que seja perfeitamente racional,

<sup>3</sup> Michael Anderson e Susan Anderson, "General Introduction", in M. Anderson e S. Anderson (orgs.), *Machine Ethics*. Cambridge: Cambridge University Press, 2011, pp. 1-4.

<sup>4</sup> As Três Leis da Robótica, imaginadas por Isaac Asimov, são as seguintes: 1) Um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano sofra algum mal. 2) Um robô deve obedecer a ordens dadas por seres humanos, exceto nos casos em que tais ordens entrem em conflito com a Primeira Lei. 3) Um robô deve proteger sua própria existência desde que tal proteção não entre em conflito com a Primeira ou Segunda Leis. Asimov explorou a complexidade dessas leis em sua literatura, mostrando como são incompletas e paradoxais, levando robôs a conflitos éticos.



porém insensível às preocupações humanas porque é desprovida de emoções.<sup>5</sup>

Por essas razões, poderíamos rejeitar inteiramente a própria ideia de agência moral plena, ou então poderíamos tomar uma posição intermediária: oferecer à IA algum tipo de moralidade, mas não uma moralidade plena. Wendell Wallach e Collin Allen usam o termo “moralidade funcional”.<sup>6</sup> Sistemas de IA precisam ter alguma capacidade de avaliar as consequências éticas de suas ações. A justificativa para essa decisão é evidente no caso dos carros autônomos: o carro provavelmente entrará em uma situação em que uma escolha moral tenha que ser feita, sem que haja tempo para a tomada de decisão ou intervenção humana. Algumas vezes essas escolhas tomam a forma de dilemas. Filósofos falam sobre os *dilemas do bonde*, nomeados por conta de um experimento mental em que um bonde desce disparado pelos trilhos e você deve escolher entre não fazer nada, o que matará cinco pessoas amarradas aos trilhos, ou acionar uma alavanca e desviar o curso do bonde para outro trilho, onde há uma só pessoa amarrada, mas ela é alguém que você conhece. Qual é a coisa moralmente certa a fazer? De forma análoga, os proponentes dessa abordagem argumentam que um carro autônomo pode ter de fazer uma escolha moral entre, por exemplo, matar pedestres atravessando a rua ou decidir ir em direção a uma parede, matando assim o motorista. O que o carro deve escolher? Parece que teremos de tomar essas decisões morais (de

5 M. Coeckelbergh, “Moral Appearances: Emotions, Robots, and Human Morality”. *Ethics and Information Technology*, v. 12, n. 3, pp. 235-41, 2010.

6 Wendell Wallach e Colin Allen, *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press, 2009, p. 39.

antemão) e garantir que os desenvolvedores as implementem nos carros. Ou talvez precisemos construir carros que aprendam com as escolhas humanas. Contudo, é possível questionar se oferecer regras às IAs é um bom modo de representar a moralidade humana; se a moralidade pode realmente ser representada e reproduzida; e se dilemas do bonde podem capturar algo que é central para a vida e experiência morais. Ou, a partir de uma perspectiva completamente diferente, cabe perguntar se seres humanos são de fato bons em realizar escolhas morais. E por que imitar a moralidade humana? Transumanistas, por exemplo, podem argumentar que as IAs terão uma moralidade superior porque serão mais inteligentes do que nós.

Esse questionamento com foco no humano nos leva a outra posição, que não requer agência moral plena e tenta sair da posição ética antropocêntrica. Luciano Floridi e J. W. Sanders<sup>7</sup> defenderam uma moralidade não mental, não baseada em propriedades humanas. Poderíamos fazer com que a agência moral dependesse de um nível suficiente de interatividade, autonomia e adaptatividade, e de ser capaz de uma ação moralmente qualificada. De acordo com esses critérios, um cão de busca e resgate é um agente moral, assim como um robô de busca com IA que filtra e-mails indesejados. Do mesmo modo, é possível aplicar critérios não antropocêntricos para a agência moral de robôs, como proposto por John Sullins:<sup>8</sup> se uma IA é autônoma com relação aos programadores e pode-

7 Luciano Floridi e J. W. Sanders, “On the Morality of Artificial Agents”. *Minds and Machines*, v. 14, n. 3, pp. 349-79, 2004.

8 John Sullins, “When Is a Robot a Moral Agent?”. *International Review of Information Ethics*, v. 6, pp. 23-30, 2006.

-essa pergunta expõe o fundo  
a humanidade da posição sobre a  
IA.

mos explicar seu comportamento atribuindo intenções morais a ela (como a intenção de fazer o bem ou de machucar), e se ela se comporta de um modo que mostra um entendimento de sua responsabilidade em relação a outros agentes morais, então essa IA é um agente moral. Assim sendo, tais visões não requerem uma agência moral plena, se isso significa agência moral humana, mas, pelo contrário, definem agência moral de um modo que é, em princípio, independente da agência moral humana plena e das capacidades humanas necessárias para tal. Contudo, seria tal agência moral artificial suficiente se julgada por padrões humanos? A preocupação prática é que, por exemplo, carros autônomos possam não ser morais o suficiente. Já a preocupação baseada em princípios é que nos distanciemos muito da moralidade humana. Muitas pessoas pensam que a agência moral está e deveria estar conectada à essência do humano e da personalidade. Elas não estão dispostas a endossar noções pós-humanistas ou transumanistas.

### Suscetibilidade moral

Outra controvérsia se refere à suscetibilidade moral em IA. Imagine que temos uma IA superinteligente. É moralmente aceitável desligá-la, “matá-la”? E, mais próximo à IA de hoje: é aceitável chutar um cachorro-robô construído com IA?<sup>9</sup> Se as IAs estão fadadas a ser parte da vida cotidiana, como vários pesquisadores predizem, então tais casos inevitavelmente

<sup>9</sup> Um caso real disso foi o cão-robô Spot, que foi lançado por seus desenvolvedores para testar isso, algo que conseguiu respostas surpreendentemente empáticas. Cf. “Watch robot dog ‘Spot’ run, Walk... and get kicked”. *On Demand News*, 11 fev. 2015.

surgirão e levantarão a questão de como nós, seres humanos, devemos lidar com esses entes artificiais. Novamente, não precisamos olhar para o futuro distante ou para a ficção científica. Pesquisas mostraram que, já nos dias atuais, pessoas criam empatia por robôs e hesitam em matá-los ou torturá-los,<sup>10</sup> mesmo que esses robôs não possuam IA. Seres humanos parecem exigir muito pouco de agentes artificiais para projetar sobre eles traços de personalidade ou humanidade e criar empatia por eles. Se esses agentes agora se tornam artificialmente inteligentes, o que potencialmente os torna mais parecidos com humanos (ou com animais), isso parece tornar a questão sobre a suscetibilidade moral ainda mais urgente. Por exemplo, como devemos reagir às pessoas que criam empatia por uma IA? Elas estão erradas?

Dizer que as IAs são apenas máquinas e que as pessoas quem têm empatia por elas estão simplesmente equivocadas em seus julgamentos, emoções e experiência moral é talvez a posição mais intuitiva. Em um primeiro momento, parece que não devemos nada às máquinas. Elas são coisas, não pessoas. Muitos pesquisadores de IA seguem essa linha de pensamento. Por exemplo, Joanna Bryson argumentou que robôs são ferramentas e posses e que não temos obrigações para com

<sup>10</sup> Yutaka Suzuki et al., “Measuring Empathy for Human and Robot Hand Pain Using Electroencephalography”. *Scientific Reports*, v. 5, n. 15.924, 2015; Kate Darling, Palash Nandy e Cynthia Breazeal, “Empathic Concern and the Effect of Stories in Human-Robot Interaction”, in *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 24. New York: IEEE, 2015, pp. 770-75.

eles.<sup>11</sup> Os que mantêm essa posição podem concordar que, se IAs fossem conscientes, tivessem estados mentais e assim por diante, deveríamos dar a elas status moral. Mas dirão que essa condição não é satisfeita atualmente. Como vimos nos capítulos anteriores, alguns argumentarão que isso não poderá jamais ser satisfeito; outros dirão que em princípio isso pode ser realizado, mas que não tão cedo. A conclusão sobre a questão do status moral é que hoje e em um futuro próximo as IAs serão tratadas como coisas até prova em contrário.

Um dos problemas com essa posição é, contudo, que isso não explica nem justifica as intuições e experiências morais que nos dizem haver *algo* de errado em “maltratar” uma IA, mesmo se a IA não tenha propriedades similares às humanas ou às dos animais, como consciência e senciência. Para encontrar tais justificativas, pode-se recorrer a Kant, que argumentou ser errado atirar em um cachorro, não porque atirar em um cachorro viole algum dever para com o cachorro, mas porque quem assim o faz “fere as qualidades amáveis e humanas de si mesmo, as quais deve exercitar em virtude de seus deveres para com a humanidade”.<sup>12</sup> Hoje tendemos a pensar de maneira diferente sobre os cães (embora não todos, nem em todos os lugares). Mas parece que o argumento poderia ser aplicado às IAs: poderíamos dizer que não devemos nada a uma IA, mas, mesmo assim, não deveríamos chutar ou “torturar” a IA porque isso nos torna insensíveis com os pró-

11 Joanna Bryson, “Robots Should Be Slaves”, in Yorick Wilks (org.), *Close Engagements with Artificial Companions: Key Social, Psychological, Ethical and Design Issues*. Amsterdam: John Benjamins, 2010, pp. 63-74.

12 Immanuel Kant, *Lectures on Ethics* [1930], Peter Heath e J. B. Schneewind (orgs.), trad. Peter Heath. Cambridge: Cambridge University Press, 1997.

prios seres humanos. Poderia ser usado o argumento da ética da virtude, que também é um argumento indireto, visto que concerne a seres humanos, e não a IAs: “maltratar” uma IA é errado não por conta do dano causado à IA, mas porque nosso caráter moral é prejudicado ao fazer isso. Isso não nos torna pessoas melhores. Contra essa abordagem poderíamos argumentar que no futuro algumas IAs podem vir a possuir valor intrínseco e merecer nossa preocupação moral, contanto que tenham propriedades como senciência. Uma abordagem pelo dever indireto ou virtude não parece levar a sério esse “outro” lado da relação moral. Ela se preocupa somente com os seres humanos. E as IAs? Mas, como perguntou David Gunkel,<sup>13</sup> IAs e robôs podem mesmo ser um outro? Novamente, o senso comum parece dizer: não, as IAs não têm as propriedades necessárias para isso.

Uma abordagem completamente diferente propõe que o modo pelo qual questionamos o status moral é problemático. O raciocínio moral mais comum sobre status moral é fundamentado nas propriedades morais relevantes que as entidades têm – por exemplo, consciência ou senciência. Mas como sabemos se uma IA possui certas propriedades morais relevantes ou não? Estamos seguros quanto a isso no caso de seres humanos? O cético diz que não. No entanto, mesmo sem essa certeza epistemológica ainda atribuímos status moral a seres humanos com base na aparência. Isso provavelmente aconteceria se no futuro as IAs tivessem aparência e comportamentos humanos. Parece que, independentemente do que

13 David Gunkel, “The Other Question: Can and Should Robots Have Rights?”. *Ethics and Information Technology*, n. 20, pp. 87-99, 2018.



o status moral atribuído a partir da relação

seja considerado moralmente *correto* pelos filósofos, os seres humanos de qualquer modo atribuirão status moral a tais máquinas e, por exemplo, darão a elas direitos. Além do mais, se olharmos mais atentamente para como os seres humanos *realmente* atribuem status moral, vemos que, por exemplo, relações sociais existentes e linguagem têm um papel nisso. Vejamos. Se tratamos nosso gato bem, não é por conta do nosso envolvimento moral com o gato, mas porque já temos uma espécie de relação social com ele. Ele já é um animal de estimação e companheiro antes de realizarmos o trabalho filosófico de lhe atribuir um status moral – se é que sentiríamos um dia a necessidade de fazer tal exercício. E, se damos ao nosso cachorro um nome pessoal, então – em contraste com os animais sem nome que comemos – já conferimos a ele um status moral independente de suas propriedades. Utilizando uma abordagem relacional, crítica e não dogmática como essa,<sup>14</sup> poderíamos argumentar que, de maneira similar, o status das IAs será atribuído por seres humanos e dependerá de como elas estiverem inseridas em nossa vida social, na linguagem e na cultura humana.

Além disso, como tais condições variam historicamente – lembremo-nos de como tratávamos os animais e pensávamos sobre eles –, talvez alguma precaução moral seja necessária antes de “fixarmos” o status moral da IA em geral ou de uma IA em particular. E por que, aliás, falar de uma IA em geral ou de modo abstrato? Parece que há algo de errado com o procedimento moral de atribuir um status:

<sup>14</sup> M. Coeckelbergh, *Growing Moral Relations: Critique of Moral Status Ascription*. New York: Palgrave Macmillan, 2012.

para julgá-lo, tiramos o ente de seu contexto relacional e, antes de termos o resultado de nosso procedimento moral, já o tratamos, de maneira bastante hierárquica, paternalista e hegemônica, como um ente sobre o qual nós, juízes humanos superiores, tomaremos decisões. Parece que, antes de realizarmos nosso raciocínio sobre o seu status moral de fato, já o posicionamos e talvez até já tenhamos cometido alguma violência, tratando-o como objeto de nossa tomada de decisão e nos colocando como deuses poderosos, centrais, oniscientes e dotados do direito de conferir status moral a outros entes sobre a Terra. Nós também já tornamos invioláveis todas as condições e contextos situacionais e sociais. Como no dilema do bonde, reduzimos a ética a uma caricatura. Com tal raciocínio, filósofos morais parecem fazer o que filósofos dreyfusianos acusaram os pesquisadores da IA simbólica de fazer: formalizar e abstrair a riqueza da experiência e conhecimento morais com o ônus de deixar de lado o que nos faz humanos e – adicionalmente – com o risco de dar por resolvida a questão mesma do status moral dos não humanos. Independentemente de qual “é”, de fato, o status moral das IAs – como se isso pudesse ser definido de maneira totalmente apartada da subjetividade humana –, é válido examinar de forma crítica nossa própria atitude moral e o projeto em si de um raciocínio moral abstrato.

Rumo a questões éticas mais práticas

Como mostram as discussões deste capítulo e dos anteriores, pensar sobre a IA não nos ensina apenas coisas sobre a IA. Também nos ensina algo sobre nós mesmos: sobre como

pensamos e como nos relacionamos e deveríamos nos relacionar com não humanos. Se olharmos para os fundamentos filosóficos da ética na IA, observaremos desacordos profundos sobre a natureza e o futuro da humanidade, da ciência e da modernidade. Questionar a IA abre um abismo de questões críticas sobre o conhecimento humano, a sociedade humana e a natureza da moralidade humana.

Essas discussões filosóficas são menos inverossímeis e menos “acadêmicas” do que se possa imaginar. Continuarão vindo à tona quando, mais adiante neste livro, considerarmos questões éticas, legais e políticas mais concretas levantadas pela IA. Se tentarmos atacar tópicos como responsabilidade e carros autônomos, a transparência da aprendizagem de máquina, IA com vieses ou a ética relacionada a robôs sexuais, rapidamente nos defrontaremos com elas novamente. Se a ética na IA quiser ser mais do que uma lista de pendências, também deve ter algo a dizer sobre tais questões.

Dito isso, agora é hora de lidarmos com assuntos mais práticos. Eles não estão relacionados nem com problemas filosóficos levantados pela hipótese de uma inteligência artificial geral, nem com os riscos de uma superinteligência em um futuro distante, ou outros monstros espetaculares da ficção científica. Concernem às menos visíveis e quem sabe menos atraentes, mas ainda muito importantes, realidades das IAs que já estão em operação. A IA em funcionamento nos tempos atuais não tem o papel de um monstro de Frankenstein ou dos espetaculares robôs inteligentes que ameaçam a civilização, mas é mais do que um experimento de pensamento filosófico. A IA tem a ver com as tecnologias menos visíveis, subjacentes, porém pervasivas, poderosas, e cada vez

mais inteligentes que já moldam nossa vida hoje. A ética na IA está, portanto, relacionada aos desafios éticos colocados pela IA atual e do futuro próximo, bem como a seus impactos em nossas sociedades e democracias vulneráveis. A ética na IA diz respeito à vida das pessoas e à política. Trata da nossa necessidade, como pessoas e como sociedade, de lidar com as questões éticas *agora*.