

# Os desafios éticos da inteligência artificial e dos objetos autônomos: um preâmbulo

Los desafíos éticos de la inteligencia artificial y los objetos autónomos: un preámbulo

The ethical challenges of artificial intelligence and autonomous objects: a preamble

[Luís Fernando Lopes](#)  [Alvino Moser](#)  [André Luiz Moscaleski Cavazzani](#) 

## Destaques

As inteligências artificiais apresentam benefícios, mas também riscos significativos para a humanidade.

O debate ético sobre as inteligências artificiais se torna cada vez mais urgente.

O utilitarismo consequencialista caracteriza as reflexões éticas atuais sobre inteligências artificiais.

## Resumo

Analisa-se os desafios éticos suscitados pelo vertiginoso avanço dos sistemas de inteligência artificial (IA) e dos objetos autônomos incorporados em seu funcionamento. Problematizam-se as perspectivas éticas subjacentes às recentes comissões criadas em instâncias internacionais (Organização das Nações Unidas e Parlamento Europeu) para avaliar os riscos das IAs. Pontua-se que as reflexões éticas sobre IAs estão permeadas por um utilitarismo consequencialista, visto que as comissões geralmente são formadas por agentes, em sua maioria, diretamente interessados nos avanços e na lucratividade das IAs.

[Resumen](#) | [Abstract](#)

## Palavras-chave

Ética. Inteligência Artificial. Humanismo.

Recebido: 09.08.2023

Aceito: 08.11.2023

Publicado: 14.11.2023

DOI: <https://doi.org/10.26512/lc29202350406>

## | Introdução

A inteligência artificial (IA), com seu avanço nas últimas décadas, desde Marvin Lee Minsky (1927-2016), proporcionou efeitos impactantes aos produtos nos quais está incorporada. Como exemplo, podem ser mencionados os simples automatismos, antes mecânicos, agora incomparavelmente melhores e mais eficientes.

Vale mencionar que, em 1968, pretendia-se, com a ciência da IA construir máquinas para fazer coisas que requerem inteligência quando feitas pelo homem. É o caso dos robôs e demais objetos autônomos, inclusive as máquinas de guerra, sem esquecer suas aplicações posteriores nos videogames. Nesse sentido, há produtos, como os robôs, que servem ao homem como eficientes auxiliares, nos serviços corriqueiros e em caso de doenças, sobretudo pelo que vêm possibilitando à Internet das Coisas (IdC) ou *Internet of Things* (IoT).

Tem-se controle sobre a casa, a geladeira, os tapetes colocados ao lado das camas dos idosos para detectar se estão bem ao acordar, e até para diagnosticar doenças e socorrer as pessoas ou solicitar auxílio. Há, ainda, robôs criados para serem “cuidadores” de pessoas idosas ou pessoas com necessidades especiais. Trata-se do “maravilhoso mundo da IoT”. Emergem a Transformação Digital, a indústria 4.0 e o orbe dos aplicativos. Benefícios são esperados dos carros autônomos, pois, segundo as estatísticas do Observatório Nacional de Segurança Viária (ONSV), 90% dos acidentes automobilísticos ocorrem por falhas humanas (Portal ONSV, 2015).

A indústria do lazer, por sua vez, fatura bilhões de dólares com os videogames. Há, inclusive, usos mais heterodoxos e discutíveis, por exemplo, os jogos sexuais, substituindo as relações humanas pela interação com as máquinas “inteligentes”, emuladoras e estimuladoras da sexualidade humana (Souza, 2019). Contudo, nesse contexto, há possibilidades e usos ainda mais preocupantes. Os sistemas inteligentes de aviões bombardeiros supersônicos, por exemplo, permitem aos pilotos visarem seus alvos sem que sua sensibilidade seja afetada. Como se fossem objetos em um videogame, matam, destroem edifícios e cidades, friamente, ao som da música de sua preferência. Citamos, ainda, as máquinas de guerra autônomas. O que era ou ainda é ficção científica está, em vertiginosa velocidade, tornando-se realidade.

Surgem, então, algumas perguntas: o que se pretende com a IA? O que estamos tentando fazer? Será que a criação de autômatos mais inteligentes e fortes do que os humanos são desejáveis? Será justo escravizar esses autômatos (aliás, robô significa escravo)? Manipulada e aprimorada por nanotecnologia em breve, a espécie *sapiens* poderá continuar a ser assim chamada? (Harari, 2017). Essas questões nos introduzem no mundo da ética e da moral.

Desde há aproximadamente cinco anos, nota-se uma profusão dos estudos e das iniciativas em relação à crescente utilização das IAs. Conforme Perrault et al.

(2019, p. 5), entre 1998 e 2018, o número de artigos com revisão por pares sobre Inteligência Artificial (IA) cresceu mais de 300%, representando 3% das publicações em revistas revisadas por pares e 9% dos artigos de conferência publicados no mundo. Tal recrudescimento reflete a crescente popularidade da IA e do Aprendizado de Máquina entre os estudantes. Em 2018, mais de 21% dos doutorandos em Ciência da Computação se especializaram em IA e Aprendizado de Máquina (Perrault et al., 2019, p. 6).

Diante de algo que poderíamos classificar como um primado de preocupações técnicas, é urgente resgatar a perspectiva humanística e as implicações éticas desses modelos autônomos ou, então, objetos automatizados que simulam autossuficiência, mas, na verdade, são criados e controlados por meio de programas cujo desenvolvimento interno é processado em caixa-preta.

Com efeito, é importante sublinhar que os modelos de IA e, além disso, a popularização e difusão de seu emprego, respiram os ares do nosso tempo. Ou seja, seu rápido e agressivo desenvolvimento se dá num ritmo de sincronicidade com a temporalidade dos autores e da própria escritura em que se produz estas linhas. O fenômeno e sua interpretação são afetados pelo calor do presente, o que interfere nas nossas conclusões e pode produzir um viés. De todo modo, não buscamos, neste caso, uma pretensa imparcialidade. Antes, assumimos o lugar de perspectiva crítica ao fenômeno das IAs. Faz-se urgente refletir sobre algo que não deve ser compreendido isoladamente como um domínio da sofisticação técnica, que pode e, de certa forma, já está a revolucionar todo o sistema jurídico/organizacional de nosso tempo.

Para tanto, a perspectiva teórico-metodológica adotada neste ensaio é a hermenêutica, considerada aqui como uma epistemologia da interpretação (Moser & Lopes, 2016, p. 113). Nesse campo, são referência os teóricos Friedrich Schleiermacher, Wilhelm Dilthey, Martin Heidegger e Hans-George Gadamer que, segundo Palmer (1989), transformaram a hermenêutica em seu tempo. A esses quatro teóricos é preciso acrescentar o nome de Paul Ricoeur (1913-2005), que fez da hermenêutica uma teoria do texto (ou teoria da interpretação), cujo fundamento se coloca em uma ontologia da escritura. Nesse sentido, Ricoeur (1989, p. 97), conforme Heidegger, compreende a hermenêutica como uma “metodologia das ciências históricas do espírito”, na qual fica evidenciada a problemática da interpretação, que exige uma reflexão sobre ela mesma. A hermenêutica se coloca como uma teoria das operações da compreensão, em que a interpretação dos textos funciona como um jogo que determina os valores presentes em um repertório textual. A interpretação do texto requer que nos aproximemos dele, mas também de nós mesmos, pois, à medida que nos compreendemos, melhor nos explicamos. Dessa maneira, segundo Ricoeur (1989, p. 92), a tarefa da hermenêutica incide em estabelecer teoricamente a validade universal da interpretação.

Assim, ensaiamos aqui um preâmbulo crítico, que tem como fio condutor a problematização das perspectivas éticas subjacentes às recentes comissões criadas em instâncias internacionais – Organização das Nações Unidas e

Parlamento Europeu – para avaliar os riscos das IAs. Atravessados por um utilitarismo consequencialista, os debates nessas instâncias parecem contaminados pelos interesses daqueles que participam da lucratividade que pode ser gerada pelas IAs.

Assume-se aqui que este ensaio ainda oferece mais perguntas que respostas ou certezas. Contudo, demarca-se a posição de que os conhecimentos técnicos não devem – sob risco de desastre – sobrepujar-se à pessoa humana. E defende-se, claro, que o vigor das discussões e dos questionamentos acerca das implicações éticas desses sistemas operacionais deve se dar *pari passu* com seu desenvolvimento técnico. Ainda que preambulares, espera-se que estas discussões provoquem novos estudos, novas visadas, novos olhares no sentido de marcar uma posição com princípios humanistas diante, ao que parece, do inescapável desenvolvimento das IAs.

## **| A questão ética**

Os objetos autônomos cada vez mais se aproximam da inteligência humana, a tal ponto que podem tomar decisões nem sempre consideradas, com consequências, por sua vez, imprevisíveis. Eles agem de maneira autônoma, pois tomam decisões segundo as circunstâncias e as necessidades. Enfim, diz-se que são inteligentes, mas há decisões que podem ter consequências danosas, que não constavam nos programas previstos pelos desenvolvedores.

Foi o que ocorreu, por exemplo, com o acidente envolvendo um carro autônomo em Tempe, no Arizona (EUA). Tratava-se de um carro fabricado pela Tesla, de propriedade do Uber. O acidente ocorreu em março de 2018 e uma mulher, Joshua Brown, de 40 anos, morreu. Segundo dados divulgados pela polícia, o condutor estava assistindo a um filme do Harry Potter (Maussion, 2018).

Esse acidente e outros possíveis impulsionam a reflexão acerca da responsabilidade dos robôs. Entretanto, as opiniões não são unânimes. Trata-se de um problema jurídico e ético, e não apenas tecnológico. Diante dessas possibilidades, após ressaltar os enormes e sólidos benefícios que resultam da aplicação da IA, em 2015, Stephen Hawkins, Elon Musk, centenas de cientistas e tecnólogos assinaram uma carta alertando, além dos cuidados e das precauções, sobre possíveis erros e armadilhas da IA, que podem ocorrer conforme a metodologia empregada para programá-la (Griffin, 2015).

As considerações assinalam preocupações a curto prazo, como os acidentes possíveis com carros autônomos e os decorrentes das decisões dos drones de combate e, ainda, problemas com a confidencialidade. Quanto às preocupações a longo prazo, seguem as do diretor de pesquisa da Microsoft, Eric Horvitz (2014, p. 5, tradução nossa):

Especulações sobre o surgimento dessas inteligências de máquina incontroláveis têm chamado a atenção para diferentes cenários, incluindo trajetórias que tomam um curso lento e insidioso de refinamento e evolução mais rápida dos sistemas em direção a uma poderosa "singularidade" de

inteligência. Esses desfechos distópicos são possíveis? Em caso afirmativo, como podem surgir estas situações? Quais são os caminhos para esses temidos resultados? O que podemos fazer proativamente para efetivamente abordar ou diminuir a probabilidade de tais resultados e, assim, reduzir essas preocupações? Que tipo de pesquisa nos ajudaria a entender melhor e abordar as preocupações sobre o surgimento de uma superinteligência perigosa ou a ocorrência de uma "explosão de inteligência"? As preocupações sobre a perda de controle dos sistemas de IA devem ser abordadas por meio de estudo, diálogo e comunicação. As ansiedades precisam ser resolvidas, mesmo que sejam injustificadas. Estudos poderiam reduzir ansiedades mostrando a inviabilidade de resultados preocupantes ou fornecendo orientação sobre esforços proativos e políticas para reduzir a probabilidade dos resultados temidos.

A possibilidade de superinteligência é um tema popular em ficção científica, mas também é um tópico sério de pesquisa acadêmica. Alguns especialistas acreditam que a superinteligência é apenas uma questão de tempo, enquanto outros acreditam que se trata de algo impossível. De qualquer modo, há uma série de riscos potenciais associados à superinteligência, o que torna fundamental a tarefa de desenvolver salvaguardas para garantir que a superinteligência seja usada de forma segura e responsável. Segundo Russell et al. (2015, p. 107-108, tradução nossa):

Diferentes maneiras pelas quais um sistema de IA pode falhar no desempenho desejado correspondem a diferentes áreas de pesquisa de robustez: Verificação: Como provar que um sistema satisfaz certas habilidades formais desejadas. (Eu construí o sistema certo?) Validade: Como garantir que um sistema que atenda aos seus requisitos formais não tenha comportamentos e consequências indesejadas. (Eu criei o sistema certo?) Segurança: Como evitar a manipulação intencional por partes não autorizadas. Controle: Como permitir um controle humano significativo sobre um sistema de IA depois que ele começa a operar (OK, eu construí o sistema errado; posso corrigi-lo?).

Como é possível notar, os autores argumentam que quatro áreas de pesquisa são essenciais para garantir que os sistemas de IA sejam robustos e confiáveis: 1) verificação, 2) validade, 3) segurança e 4) controle. Dessa maneira, sem uma atenção adequada à robustez, os sistemas de IA podem falhar em atingir seus objetivos e até mesmo causar danos. A verificação é o processo de provar que um sistema de IA satisfaz certos requisitos formais. Isso pode ser feito usando uma variedade de técnicas, incluindo modelagem matemática, simulação e teste de unidade. A validade, por sua vez, é o processo de garantir que um sistema de IA não tenha comportamentos e consequências indesejadas. Isso pode ser feito identificando e mitigando potenciais riscos, como viés, preconceito e segurança cibernética. Já a segurança é o processo de evitar a manipulação intencional por partes não autorizadas. Isso pode ser feito usando uma variedade de técnicas, incluindo autenticação, autorização e criptografia. Por fim, o controle é o processo de permitir um controle humano significativo sobre um sistema de IA depois que ele começa a operar. Isso pode ser feito fornecendo aos usuários a capacidade de definir parâmetros, realizar configurações e revisar resultados (Russell et al., 2015).

Carros autônomos, algoritmos de recomendação, drones cujas ações suscitam a questão da ética e da moralidade. Sobretudo os efeitos, que dependem de redes

neurais artificiais (*deep learning*), oferecem flancos à crítica por serem efetuados em “caixa-preta”, logo, opacos; isso porque não se pode conhecer como funcionam os raciocínios dos algoritmos.

A ética e a moral são consideradas a bússola da boa conduta por meio de normas que delas emanam. “A ética é a teoria ou ciência do comportamento moral dos homens em sociedade” (Vázquez, 1989, p. 12), e a moral tem como critério o bem. Salientamos que os princípios e as normas ético-morais são autoimpostos pelo sujeito, que se torna autoimputável, sendo responsável por seus atos e suas consequências. Ressalta-se que liberdade e responsabilidade exigem um sujeito consciente; ao contrário das leis e normas jurídicas ou outras que são impostas por autoridade ou instâncias externas. Nesse caso, se houver infração das leis ou normas jurídicas, o infrator está sujeito à sanção ou pena. Contudo: “*Nulla poena sine lege*” (Não há crime, nem pena, sem prévia lei).

Fundamentalmente, a questão ética dos robôs e objetos autônomos requer que se esclareça se eles são ou não são conscientes, embora possam responder que estão conscientes e sabem o que fazem. Pois, para que se autoimponham uma norma, é preciso estar consciente e ser sincero. Ora, como se pode saber se o robô é sincero ou não? Negando essa possibilidade, a questão da ética dos robôs se torna insolúvel. Mas, como são criados e programados por humanos, a responsabilidade recai sobre os seus criadores, desenvolvedores e programadores ou seus usuários (Soulez, 2018).

Uma máquina ultrainteligente, por sua vez, seria capaz de projetar máquinas ainda melhores do que ela mesma, causando uma “explosão de inteligência”, que deixaria a inteligência humana para trás. A ideia de uma máquina ultrainteligente é um tópico popular de ficção científica, mas também é um campo de pesquisa sério em inteligência artificial. Alguns especialistas acreditam que o desenvolvimento de tal máquina é apenas uma questão de tempo, enquanto outros acreditam que é impossível (Russell, 2023).

De todo modo, já há razões suficientes para, no limite, ainda que de forma hipotética, considerar os riscos potenciais de um empreendimento dessa natureza. Por exemplo, uma máquina ultrainteligente poderia ser usada para criar armas autônomas, que seriam capazes de matar sem intervenção humana; operar seleções genéticas; sem falar, claro, em pressões e ebulições sociais causadas por demissões e alterações radicais no mercado de trabalho.

Nesse sentido, o que preocupa empresas e governos é o perigo de que os robôs, sobretudo os de guerra, programados para agirem de modo autônomo, revoltam-se contra a humanidade e passem a causar catástrofes, ferindo as três leis da Robótica que Isaac Asimov havia proposto em um de seus enredos de ficção científica: um robô não pode ferir um humano ou permitir que um humano sofra algum mal; os robôs devem obedecer às ordens dos humanos, exceto nos casos em que essas ordens entrem em conflito com a primeira lei; um robô deve proteger sua própria existência, desde que não entre em conflito com as leis anteriores



(Asimov, 2004). Entretanto, para causar desastres, os robôs não precisam de consciência: basta um erro ou uma combinação não prevista pelos programadores.

Segundo Hammerschmidt (2002), o termo precaução é um princípio ético que defende a adoção de medidas de proteção contra riscos potenciais, mesmo que não haja certeza incontestável de que esses riscos sejam reais. Esse princípio foi desenvolvido em resposta ao aumento da consciência sobre os riscos ambientais e tecnológicos, e tem sido aplicado em uma ampla gama de contextos, incluindo a proteção à biodiversidade, o controle de poluição e a avaliação de riscos de novos produtos.

Assim, ressalta-se a necessidade de investimentos seguros, que ainda são poucos, para que os sistemas autônomos fiquem sob o domínio dos que os manobram. Também, para que programas de formação humanística sejam ofertados, a fim de evitar ou atenuar, no limite, vícios tecnicistas aos quais estão sujeitos os operadores desses sistemas. Nesse sentido, a responsabilidade dos robôs autônomos poderia recair sobre a inteligência humana, que está por detrás da artificial, a saber: programadores, chefes-executivos de multinacionais de tecnologia, usuários, e uma miríade de atores envolvidos nas estruturas de operações das IAs. Porém, em contrapartida, no Parlamento Europeu, há uma proposta segundo a qual os objetos autônomos seriam personalidades eletrônicas e, como tais, necessitam de um seguro para cobrir as consequências desastrosas ou não esperadas, ocasionadas pelos próprios objetos.

A Comissão de Assuntos Jurídicos do Parlamento Europeu propõe rotular os robôs como “pessoas eletrônicas” - uma nova personalidade jurídica dos robôs, que deve permitir dotá-los de “direitos e deveres específicos”. O status legal do robô, segundo o relatório, não pode ser equivalente ou igual ao modelo da pessoa humana com os direitos que possui, nem seria uma pessoa jurídica. Porque quem responde pelo robô no tribunal, pelo carro autônomo, por exemplo, são pessoas físicas, pois são os advogados e os réus que os representam. Ninguém vai interrogar uma máquina inteligente pelo fato de não ser considerada pessoa: “*persona est naturae rationalis individua substantia*” (pessoa é substância individual de natureza racional), define Boethius (s.d.).

Não obstante os impasses e discussões, o OPECST (*Office Parlementaire d'Évaluation des Choix Scientifiques et Technologiques*), Órgão Parlamentar de Avaliação de Escolhas Científicas e Tecnológicas, adotou por unanimidade, em 14 de março de 2017, um relatório intitulado: “Para uma inteligência artificial controlada, útil e desmistificada”, a fim de destacar as oportunidades e os riscos da inteligência artificial, para tranquilizar o público e desmistificar as representações tendenciosas da inteligência artificial.

Contudo, Laetitia Poulighen (2019), diretora da *NBIC Ethics* (Nanotecnologia, Bioética, tecnologia Informática e tecnologia Cognitiva) e uma das mentoras da Carta Aberta à Comissão Europeia de Robótica e IA, denunciou a resolução do Regulamento do Parlamento Europeu sobre a Lei de Robótica, adotada em fevereiro de 2017, que propõe a criação de uma personalidade jurídica específica

para robôs autônomos. O Parlamento Europeu criou um grupo que redigiu o *AI Ethical Guideline* (Guia Ético para Inteligência Artificial). Esse grupo contava com 52 especialistas, mas:

[...] sua composição parece desequilibrada, pois a grande maioria era da indústria e das federações industriais e uma quase total ausência de filósofos, éticos, líderes religiosos, sociólogos, antropólogos ou mesmo pessoal de saúde [...]. A super-representação industrial do grupo de especialistas gera temores de que o pensamento será guiado por uma análise de custo-benefício de novas tecnologias em termos econômicos, sociais e ambientais, sem incluir todo um corpo de conhecimento humano sobre a ação humana em face de espera de decisões algorítmicas. (Pouliquen, 2019, s.p.)

Como se pode notar, Pouliquen argumenta que a super-representação industrial do grupo gera preocupações de que o pensamento será guiado por uma análise tecnicista/utilitarista de custo-benefício de novas tecnologias em termos econômicos, sociais e ambientais, sem incluir todo um corpo de conhecimento humano sobre a ação humana em face de decisões algorítmicas. Essa falta de diversidade no grupo de trabalho é um sinal de que a IA vem sendo desenvolvida sem uma consideração adequada para suas implicações sociais e éticas. Fato preocupante.

## Crítica à ética utilitarista ou consequencialista

Na perspectiva da maioria dos membros do grupo, anteriormente referido, o homem é visto como consumidor e simples usuário dos produtos que incorporam a IA. Essa perspectiva, como dissemos, está viciada pela predominância de profissionais ligados à indústria, que seguem uma ética utilitarista ou consequencialista.

Cabe esclarecer o que compreendemos por utilitarismo. O utilitarismo, enunciado pelo inglês Jeremy Bentham (1748-1832), prevê o princípio da maior felicidade para o maior número de pessoas. Ou, em outros termos, uma ação seria moral se tivesse com consequência o maior benefício para o maior número de pessoas. John Stuart Mill (1806-1873), em sua obra *O utilitarismo*, assumindo essa doutrina moral, salienta o caráter altruísta dela. Nessa direção, John Stuart Mill (2000, p. 49) afirma que:

A utilidade ou o princípio da maior felicidade, como fundamento da moral, sustenta que as ações são certas na medida em que elas tendem a promover a felicidade e erradas quando tendem a produzir o contrário da felicidade. Por felicidade entende-se prazer e ausência de dor, por infelicidade, dor e privação do prazer.

Nesse caminho, o utilitarismo também assume um caráter consequencialista: “qualquer teoria consequencialista deve aceitar a afirmação de que rotulei o ‘consequencialismo’, a saber, que certas propriedades normativas dependem apenas de consequências. Se essa afirmação for abandonada, a teoria deixa de ser consequencialista” (Sinnott-Armstrong, 2003, s. p.).



Vários filósofos, tais como McNaughton e Rawling (1991), Howard-Snyder (1994), Pettit e Slote (1997), afirmam que uma teoria moral não deve ser classificada como consequencialista, a menos que seja neutra por seus agentes. Essa definição mais restrita é motivada pelo fato de que muitos críticos autodenominados do consequencialismo argumentam contra a neutralidade do agente.

Entre os críticos do utilitarismo ou consequencialismo está John Rawls (2000), que problematiza esse sistema ético a partir das seguintes indagações: como é possível que a soma de todos os indivíduos da sociedade tenha os mesmos interesses e considere a felicidade como algo comum a todos? Como chegar a uma satisfação maximal da sociedade, se cada indivíduo, cada sujeito é um único ser? Qual garantia se tem que todos os membros da sociedade serão alcançados por esse cálculo da somatória de todos os sujeitos isolados? Trata-se de uma utopia. Não um “eu universal”, pois os que existem são eu(s) empíricos com sua cultura e suas determinações. Logo, o utilitarismo consequencialista é, portanto, controverso. Como uma única instância de subjetividade pode determinar o maior bem ou a maior satisfação de todos os membros de uma sociedade? Dito de forma mais rápida, “o utilitarismo não leva a sério a diferença entre as pessoas” (Rawls, 2000, p. 30).

Anderson e Anderson (2014), por sua vez, têm como perspectiva a ética utilitarista ou o consequencialismo, que vai contra a ética deontológica de Kant e de Rawls, a ética intuicionista, a ética aristotélico-tomista, entre outras. Anderson e Anderson (2014) propõem “embutir nos sistemas autônomos”, com a base em “*GenEth uses inductive logic programming*” (ILP), isto é, uma programação lógica indutiva que raciocina a partir de exemplos, o que é, no mínimo, problemático para os especialistas em ética, pois: “o utilitarismo é um empirismo que parte dos fatos, [...] considera as tendências e as inclinações dos homens e se esforça, em seguida de os satisfazer” (Russ, 1994, p. 87).

## **| A proposta de John Rawls**

John Rawls (2000) estabelece que os legisladores deveriam votar envolvidos num véu de ignorância, isto é, os que ditam as leis, os que legiferam, deveriam pôr-se em situação de não considerar se a lei vai favorecer os seus interesses ou prejudicá-los, considerando o bem de todos e não de maiorias, mas essa imparcialidade de juízo é impossível.

Rawls (2000) propõe, ainda, dois princípios. De acordo com o primeiro, o da liberdade, todas as pessoas têm as mesmas demandas para liberdades básicas. O segundo princípio, por sua vez, o da igualdade, estipula que as desigualdades sociais e econômicas devem ser ordenadas de tal modo que sejam ao mesmo tempo consideradas como vantajosas para todos, dentro dos limites do razoável (princípio da diferença), e vinculadas a posições e cargos acessíveis a todos (princípio da igualdade de oportunidades).

No primeiro princípio, contemplam-se os direitos à vida, à subsistência, ao trabalho, à educação, ao lazer, os direitos civis, sociais e políticos. É o princípio da

igualdade. A distribuição de direitos, deveres e demais bens sociais. Eles podem ser aplicados (em diferentes estágios) para o julgamento da constituição política, das leis ordinárias e das decisões dos tribunais. A distribuição desses direitos e da redistribuição de renda seriam determinados pela condição hipotética assinalada, a saber, “o véu de ignorância” (Rawls, 2000, p. 146-153).

No que se refere à justiça como equidade, Rawls (2000) pretende falar de uma noção razoável de justiça, que permita mediar a convivência política através do contrato (fazendo acordos mútuos entre as pessoas em iguais condições). Na justiça como equidade, o conceito do certo vem antes do conceito de bom. Contudo, a teoria de Rawls não deixa de ser uma tentativa utópica e que tem como fundamento o liberalismo.

Nesse sentido, é preciso considerar que o liberalismo não impera em todas as nações e, mesmo que se aceitassem os princípios de Rawls, que visam “aperfeiçoar” o consequencialismo, persiste o desafio ético: em um contexto no qual as IAs são capazes de aprender e executar tarefas cada vez mais complexas, o que acontecerá com os empregos e os trabalhadores atuais, uma vez que os desempregados e subempregados superam os milhões no mundo todo? Recentemente, Catherina Thorbecke, em artigo jornalístico, fez um levantamento demonstrando que as IAs estão afetando negativamente a empregabilidade dentro do próprio setor de tecnologia. Ou seja, trata-se de um exemplo concreto, irônico e um tanto assustador: a criatura voltando-se contra o(s) criador(es) (Thorbecke, 2023).

## **| A justiça equitativa de Aristóteles**

Seguindo nas discussões éticas, podemos refletir ainda sobre o senso de justiça no contexto das IAs. O princípio da justiça como equidade, defendido por Rawls, não tem a mesma perspectiva considerada por Aristóteles (1967) quando tratou dessa temática. Na *Ética a Nicômaco*, explicitou que o equitativo é justo, mas que nem tudo o que o é justo é equitativo. Justiça significa dar a cada um o que merece, dar a cada um o que tem direito, mas há casos em que as pessoas a quem se aplica a justiça não estão nas mesmas condições. Por exemplo, um trabalhador solteiro e um trabalhador que precisa sustentar a mulher e dois filhos. A justiça é cega para essas diferenças, e a equidade não nega a justiça, mas procura corrigi-la.

Lê-se ainda, na *Ética a Nicômaco*, que a medida da equidade é semelhante à régua de Lesbos, famoso instrumento usado pelos arquitetos para conseguir melhores resultados, pois, sendo flexível, adaptava-se melhor às imperfeições e à dureza da rocha. Já a medida da lei é semelhante à régua linear inflexível.

Nesse sentido, no contexto contemporâneo, diante dos dilemas suscitados pela utilização das IAs, pergunta-se Laetitia Pouliquen (2019, s.p.): “Mas então, sem autonomia, o usuário não teria direitos? Da mesma forma, a transparência dos algoritmos, a não discriminação, a proteção de dados serão princípios suficientes para garantir o respeito de nossas liberdades? Provavelmente não”.

Ora, considerar o homem como consumidor e usuário, embora procurando não o prejudicar, é privá-lo de sua autonomia e liberdade. A autonomia, segundo a filosofia, é a capacidade de escolha e de autoimposição de normas, a saber, de legislar por si próprio e estar ciente dos princípios que adota para sua conduta de vida. Apanágio das subjetividades viventes, o Presidente da Comissão de Assuntos Jurídicos do Parlamento Europeu julga ser inadequado aplicar o termo “autonomia” a meros artefatos, mesmo que sejam sistemas avançados, sofisticados, mesmo “inteligentes” (Pouliquen, 2019, s.p.).

O acesso a uma perspectiva mais crítica e humanista – não só às conquistas, mas também aos perigos das IAs – parece estar longe de ser uma preocupação para aqueles que escrevem guias ou orientações éticas, quando os escrevem, para o uso dos objetos criados via IA. Dentro desta complexa lógica utilitarista, regida por pesados sistemas de interesse financeiro, controle, pressão e coerção das grandes organizações de tecnologia, os resultados acerca de discussões éticas são em quase sua total maioria onfaloscópicos: olham para seus interesses ou para os interesses de quem os emprega.

Nesse sentido, não se nega aqui a necessidade de um *AI Ethical Guideline* (Guia Ético para Inteligência Artificial); ousamos, porém, a expor nossa sugestão, sem desmerecer as outras. Note-se que não entramos na discussão sobre saber se um robô autônomo é consciente ou não, depois de aplicado o Teste de Turing, pois isso nos levaria a outras considerações e outros ensaios. Propõe-se, assim, que o *Guideline* seja estabelecido seguindo as ideias de Karl Otto Apel; da ética da comunicação de Habermas; e de Rorty.

## **| Karl Otto Apel, Jürgen Habermas e Richard Rorty**

Apel (1987) propõe uma ética para a idade da ciência. Para ele, a avalanche das inovações tecnocientíficas e tecnológicas, sobretudo com as aplicações da IA, apela para uma ética global.

Quem quer que reflita sobre as relações que entre ciência e ética na sociedade industrial moderna no momento da planetarização, encontra-se, a meu ver, numa situação paradoxal. De um lado, com efeito, a necessidade de uma ética universal, isto é, suscetível de engajar a sociedade humana em sua totalidade, que nunca mais foi tão premente como em nossos dias, ao mesmo tempo que assistimos através das consequências da ciência, o estabelecimento, em escala planetária, de uma sociedade unificada. Mas, de outro lado, a missão da filosofia de fundar uma ética universal nunca foi tão árdua, haja vista desesperada como na época científica. (Apel, 1987, p. 43)

Por que é uma tarefa difícil, árdua? Porque perde-se a confiança nos fundamentos metafísicos e racionais. A metafísica cai em declínio. É nesse sentido que Heidegger (1983) declara o fim da filosofia como consumação da metafísica no predomínio das ciências direcionadas à técnica, acrescentando que esse desfecho estabelece a hipótese de um novo começo como a tarefa do pensamento.

Habermas e Rorty, por sua vez, embora com diferentes perspectivas, descartam as fundamentações metafísicas da ética e da moral e outras. Propõem a ética da

discussão ou da argumentação. Gilbert Hottois (1997), citando Richard Rorty, afirma que nada há, nem de metafísico, nem de racional ou qualquer outro tipo de fundamento, que esteja acima do consenso entre os que discutem.

O conhecimento não está acima da conversação (do diálogo, observação nossa), e nunca é legítimo terminar um debate, quer se trate da autoridade de um fato dito “objetivo” ou uma revelação dita “transcendente”. As discussões podem ser unicamente fechadas legitimamente apenas se os interlocutores estiveram de acordo sobre as razões (que também são enunciados) de fechá-las, ao menos provisoriamente. (Hottois, 1997, p. 148)

Habermas (2007) pressupõe uma pragmática universal, pois os que se propõem discutir sobre moral não podem se subtrair ao contexto intersubjetivo, que é mediado pela linguagem. Então, todos aqueles que se envolvem numa prática de argumentação precisam pressupor pragmaticamente que, em princípio, todos os possíveis afetados poderiam participar, na condição de livres e iguais, de uma busca cooperativa da verdade, na qual a única coerção admitida é a do melhor argumento. Nas palavras de Habermas, trata-se da “força sem força do melhor argumento”. Acrescenta em *Consciência Moral em Agir Comunicativo* que:

Toda norma válida deve satisfazer a condição de que as consequências e efeitos colaterais, que (previsivelmente) resultarem para a satisfação dos interesses de cada um dos indivíduos do fato de ser ela universalmente seguida, possam ser aceitos por todos os concernidos. (Habermas, 1989, p. 86)

A ética da comunicação visa o acordo e o consenso, sem que possa se cogitar em tal discussão a coação, seja moral, psicológica e, sobretudo, física. Parte-se do pressuposto já afirmado da liberdade e da autonomia dos sujeitos envolvidos.

Por conseguinte, nas questões que envolvem a IA e os recursos possibilitados por ela não se deve apenas ouvir os que a produzem ou a disponibilizam e auferem lucros, mas também os usuários, sendo estes de tendências diversas e variáveis. A ética da discussão impõe-se democraticamente. E esses pressupostos têm incidências e implicações para os políticos e governantes que, como representantes, não devem se ater aos especialistas ou aos seus correligionários, mas precisam atender a toda comunidade.

Somente dessa maneira será alcançado o bem-estar de viver juntos. Para se ter um mundo melhor, é necessário “abrir mão das próprias certezas e procurar a objetividade”. Escrevem Humberto Maturana e Francisco Varela (1995, p. 262):

[...] se sabemos que nosso mundo é sempre o mundo que construímos com outros, toda vez que nos encontrarmos em contradição ou em oposição a outro ser humano com quem desejamos conviver, nossa atitude não poderá ser de reafirmar o que vemos do nosso próprio ponto de vista, e sim de considerar que nosso ponto de vista é resultado de um acoplamento estrutural dentro de um domínio experiencial tão válido quanto o do nosso oponente, ainda que o dele nos pareça menos desejável. Caberá, portanto buscar uma perspectiva mais abrangente de um domínio experiencial em que o outro também tenha lugar e no qual possamos, com ele construir um mundo [...]. A este ato de ampliar nosso domínio cognitivo reflexivo, que sempre implica uma experiência nova, só podemos chegar pelo raciocínio motivado pelo encontro

com o outro, pela possibilidade de olhar o outro como um igual, num ato que habitualmente chamamos de amor – ou, se não quisermos uma palavra tão forte, a aceitação do outro ao nosso lado na convivência. Esse é o fundamento biológico do fenômeno social: sem o amor, e sem a socialização não há humanidade.

Nosso mundo é construído em conjunto com os outros. Assim, precisamos buscar uma perspectiva mais abrangente de um domínio experiencial, em que o outro também tenha lugar onde possamos, com ele, construir um mundo melhor. Contudo, só podemos chegar a esse ato de ampliar nosso domínio cognitivo reflexivo por meio do raciocínio motivado pelo encontro com o outro como um igual. Contudo, sem jamais desconsiderar suas singularidades.

## **| Considerações finais**

Numa perspectiva ética, analisamos alguns dos avanços que a IA apresenta. Nesse sentido, destacamos principalmente as implicações éticas relacionadas aos avanços tecnológicos dos objetos autônomos que incorporam cada vez mais a utilização da IA. Os objetos autônomos são os que estão programados de maneira prevista ou previsível, ainda que alguns sejam não previsíveis, pois são programados em caixa-preta.

Diante disso, comissões e organismos são criados para avaliar benefícios e riscos, suscitando nas organizações, como a Organização das Nações Unidas (ONU); nos especialistas, como Stephen Hawkins; e nos bilionários, como Elon Musk, receios e preocupações com as possíveis consequências que podem advir do uso da IA. Daí a criação do grupo europeu que redigiu o *AI Guide Line Ethics*, que evidencia a necessidade de refletir sobre esses avanços na perspectiva da ética ou da moral.

De modo geral, é possível considerar que a maioria dos que se ocupam oficialmente dessa preocupação com o uso da IA são especialistas que se orientam pelo consequencialismo. Mas, em conclusão – seja o utilitarismo ou consequencialismo; a correção pela equitatividade de John Rawls; a teoria da justiça de Aristóteles; a ética deontológica de Kant; a ética intuicionista; o princípio de responsabilidade de Hans Jonas; a ética da comunicação, do diálogo e do consenso; e éticas outras –, a situação marcada pelos usos e avanços da IA evidencia que nos defrontamos com o redemoinho e o turbilhão das valsas éticas a nos envolver, com todos os chamativos ou sugestões, imagináveis ou não, para fazer frente aos desafios que os avanços tecnológicos desencadeiam.

Não importa qual perspectiva ou via ética adotarmos, sabemos que a IA nos fornece benefícios ou riscos e pode inclusive auxiliar na ponderação das decisões. Entretanto, em última análise, é ao ser humano consciente que compete a palavra final. Por isso, os objetos autônomos precisam ser sempre possíveis de serem assumíveis e controláveis pelo ser humano, a quem cabe a responsabilidade de escolha. E essa decisão está suspensa à liberdade, pois somos condenados à liberdade, dizem Jean-Paul Sartre (1970) e Giovanni Pico (2015). Ao ser humano, que é um ser de possibilidade, cabe o sim ou o não.

## **| Referências**

- Anderson, M., & Anderson, S. (2014). GenEth: A General Ethical Dilemma Analyzer. *Proceedings of the AAAI Conference on Artificial Intelligence*, 28(1).  
<https://doi.org/10.1609/aaai.v28i1.8737>
- Apel, K. O. (1987). *L'éthique à l'âge dala science*. Presses Universitaires.
- Aristóteles. (1967). *Obras Completas*. Aguilar.
- Asimov, I. (2004). *Eu, Robô*. Aleph.
- Boethius, A. M. S. (s.d.) *Liber De Persona et Duabus Naturis Contra Eutychem et Nestorium*.  
<https://scaife.perseus.org/reader/urn:cts:latinLit:stoa0058.stoa023.perseus-lat1:pr-4/>
- Griffin, A. (2015, janeiro 12). Stephen Hawking, Elon Musk e outros pedem pesquisas para evitar os perigos da inteligência artificial. *Independent, Tecnologia*. <https://www.independent.co.uk/tech/stephen-hawking-elon-musk-and-others-call-for-research-to-avoid-dangers-of-artificial-intelligence-9972660.html>
- Habermas, J. (1989). *Consciência moral e agir comunicativo*.
- Habermas, J. (2007). *A ética da discussão e a questão da verdade*. Martins Fontes.
- Hammerschmidt, D. (2002). O Risco na Sociedade Contemporânea e o Princípio da Precaução no Direito Ambiental. *Estudos Jurídicos e Políticos*, (45), 97-122. <https://periodicos.ufsc.br/index.php/sequencia/article/view/15317/13912>
- Harari, Y. N. (2017). *Homo Deus: uma breve história do amanhã*. Companhia das Letras.
- Heidegger, M. (1983). *O fim da filosofia e a tarefa do pensamento*. Conferências e escritos filosóficos (Tradução, introduções e notas: Ernildo Stein). Abril Cultural.
- Horvitz, E. (2014). *One Hundred Year Study on Artificial Intelligence: Reflections and Framing*.  
[https://www.microsoft.com/en-us/research/wp-content/uploads/2016/11/AI100\\_framing\\_memo.pdf](https://www.microsoft.com/en-us/research/wp-content/uploads/2016/11/AI100_framing_memo.pdf)
- Hottois, G. (1997). *De la Renaissance à la Postmodernité: Une histoire de la philosophie moderne et contemporaine*. DE Book.
- Howard-Snyder, F. (1994). *The Heart of Consequentialism: Philosophical Studies*.  
<https://www.jstor.org/stable/4320528>
- Maturana, R. H., & Varela G. F. (1995). *A Árvore do conhecimento: as bases biológicas do entendimento humano*. Editorial Psy II.
- Maussion, F. (2018, maio 9). Accident mortel en Arizona: le véhicule autonome a détecté le piéton, sans l'éviter. *LesEchos*.  
<https://www.lesechos.fr/2018/05/accident-mortel-en-arizona-le-vehicule-autonome-a-detecte-le-pieton-sans-leviter-990132>
- Mcnaughton, D., & Rawling, P. (1991). "Agent-Relativity and the Doing-Happening Distinction". *Philosophical Studies*. 63, 167-185.  
<https://www.jstor.org/stable/4320228>
- Mill, J. S. (2000). *O utilitarismo*. Iluminuras.
- Moser, A., & Lopes, L. F. (2016). *Para compreender a teoria do conhecimento*. InterSaberes.
- Palmer, R. (1989). *Hermenêutica*. Lisboa: Edições 70.
- Perrault, R., Shoham, Y., Brynjolfsson, E., Clark, J., Etchemendy, J., Grosz, B., Lyons, T., Manyika, J., Niebles, J. C., Mishra, S. (2019). *The AI Index 2019 Annual Report*. AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA.  
[https://hai.stanford.edu/sites/default/files/ai\\_index\\_2019\\_report.pdf](https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf)




- Pettit, B. M., & Slote, M. (1997). *The Consequentialist Perspective in Three Methods of Ethics*. Blackwell.
- Pico, G. (2015). *Discurso pela dignidade do homem*. Editora Fi.  
[https://www.fucap.edu.br/dashboard/livros\\_online/d3a96398e5b6a4ce2d49c05309f9ade7.pdf](https://www.fucap.edu.br/dashboard/livros_online/d3a96398e5b6a4ce2d49c05309f9ade7.pdf)
- Portal ONSV. (2015, julho 15). 90% dos acidentes são causados por falhas humanas, alerta Observatório. *Observatório Nacional de Segurança Viária*.  
<https://www.onsv.org.br/comunicacao/materias/90-dos-acidentes-sao-causados-por-falhas-humanas-alerta-observatorio>
- Pouliquen, L. (2019, fevereiro 21). A Inteligência Artificial pode ser ética? *Le Figaro*.  
<https://www.lefigaro.fr/vox/societe/2019/02/21/31003-20190221ARTFIG00141-l-intelligence-artificielle-peut-elle-etre-ethique.php>
- Rawls, J. (2000). *Uma teoria da Justiça*. Martins Fontes, Tempo Brasileiro.
- Ricoeur, P. (1989). *Do texto à ação: ensaios de hermenêutica II*. Porto Rés-Editora.
- Russ, J. (1994). *La pensée éthique contemporaine*. PUF.
- Russell, S. (2023). *O futuro a longo prazo da IA*.  
<https://people.eecs.berkeley.edu/~russell/research/future/>
- Russell, S., Dewey, D., & Tegmark, M. (2015). *Research Priorities for Robust and Beneficial Artificial Intelligence*. Association for the Advancement of Artificial Intelligence. Winter.  
[https://futureoflife.org/data/documents/research\\_priorities.pdf](https://futureoflife.org/data/documents/research_priorities.pdf)
- Sartre, J. P. (1970). *Existencialismo é um Humanismo*. Les Éditions Nagel.  
[http://www.educadores.diaadia.pr.gov.br/arquivos/File/2010/sugestao\\_leitura/filosofia/texto\\_pdf/existencialismo.pdf](http://www.educadores.diaadia.pr.gov.br/arquivos/File/2010/sugestao_leitura/filosofia/texto_pdf/existencialismo.pdf)
- Sinnott-Armstrong, W. (2003). Consequencialismo. *Stanford Encyclopedia of Philosophy*. [https://plato.stanford.edu/entries/consequentialism/?trk=article-ssr-frontend-pulse\\_x-social-details\\_comments-action\\_comment-text](https://plato.stanford.edu/entries/consequentialism/?trk=article-ssr-frontend-pulse_x-social-details_comments-action_comment-text)
- Soulez, M. (2018, fevereiro 7). IA: um robô criativo é um autor? *Ina La Revue des médias*. <https://larevuedesmedias.ina.fr/ia-un-robot-createur-est-il-un-auteur>
- Souza, F. (2019, maio 05). Robôs sexuais e produtos da Xiaomi são destaques de eletrônicos em abril. *TechTudo*.  
<https://www.techtudo.com.br/listas/2019/05/robos-sexuais-e-produtos-da-xiaomi-sao-destaques-de-eletronicos-em-abril.ghtml>
- Thorbecke, C. (2023, julho 7). Inteligência Artificial provoca demissões na própria indústria que a criou. *CNN Brasil*.  
<https://www.cnnbrasil.com.br/economia/inteligencia-artificial-provoca-demissoes-na-propria-industria-que-a-criou>
- Vázquez, A. S. (1989). *Ética*. Civilização Brasileira.

## Sobre os autores

### Luís Fernando Lopes


Centro Universitário Internacional Uninter, Curitiba, PR, Brasil

 <https://orcid.org/0000-0001-7925-9653>

Doutor em Educação pela Universidade Tuiuti do Paraná (2017). Professor do Programa de mestrado e doutorado profissional em Educação e Novas Tecnologias do Centro Universitário Internacional Uninter. E-mail: [fernandocater@gmail.com](mailto:fernandocater@gmail.com)


### Alvino Moser

Centro Universitário Internacional Uninter, Curitiba, PR, Brasil

 <http://orcid.org/0000-0001-7828-5067>

Doutor em Ética pela Université Catholique de Louvain (1973). Decano e Professor titular do Centro Universitário Internacional Uninter. E-mail: [moseral.am@uninter.com](mailto:moseral.am@uninter.com)

### **André Luiz Moscaleski Cavazzani**

Centro Universitário Internacional Uninter, Curitiba, PR, Brasil  
 <https://orcid.org/0000-0003-1512-3639>

Doutor em História Social pela Universidade de São Paulo (2013). Professor titular do Programa de mestrado e doutorado profissional em Educação e Novas Tecnologias do Centro Universitário Internacional Uninter. E-mail: [andre.ca@uninter.com](mailto:andre.ca@uninter.com)

Contribuição na elaboração do texto: os autores contribuíram igualmente na elaboração do manuscrito.

### **| Resúmen**

Se analizan los desafíos éticos que plantea el avance vertiginoso de los sistemas de inteligencia artificial (IA) y los objetos autónomos que los incorporan a su funcionamiento. Las perspectivas éticas que subyacen a las recientes comisiones creadas en organismos internacionales (Naciones Unidas y Parlamento Europeo) para evaluar los riesgos de las IA están problematizadas. Se observa que las reflexiones éticas sobre las IA están permeadas por un utilitarismo consecuencialista, ya que las comisiones generalmente están formadas por agentes, la mayoría de los cuales están directamente interesados en los avances y la rentabilidad de las IA.

**Palabras clave:** Ética. Inteligencia Artificial. Humanismo.

### **| Abstract**

The ethical challenges raised by the dizzying advancement of artificial intelligence (AI) systems and the autonomous objects that incorporate them into their operation are analyzed. The ethical perspectives underlying the recent commissions created in international bodies (United Nations and European Parliament) to assess the risks of AIs are problematized. It is noted that ethical reflections on AIs are permeated by a consequentialist utilitarianism, since the commissions are generally formed by agents, the majority of whom are directly interested in the advances and profitability of AIs.

**Keywords:** Ethic. Artificial intelligence. Humanism.

**Linhas Críticas** | Periódico científico da Faculdade de Educação da  
Universidade de Brasília, Brasil  
ISSN eletrônico: 1981-0431 | ISSN: 1516-4896  
<http://periodicos.unb.br/index.php/linhascriticas>

**Referência completa (APA):** Lopes, L. F., Moser, A., & Cavazzani, A. L. M. (2023). Os desafios éticos da inteligência artificial e dos objetos autônomos: um preâmbulo. *Linhas Críticas*, 29, e50406.  
<https://doi.org/10.26512/lc29202350406>

**Referência completa (ABNT):** LOPES, L. F.; MOSER, A.; CAVAZZANI, A. L. M. Os desafios éticos da inteligência artificial e dos objetos autônomos: um preâmbulo. *Linhas Críticas*, 29, e50406, 2023. DOI:  
<https://doi.org/10.26512/lc29202350406>

**Link alternativo:** <https://periodicos.unb.br/index.php/linhascriticas/article/view/50406>

Todas as informações e opiniões deste manuscrito são de responsabilidade exclusiva do(s) seu(s) autores, não representando, necessariamente, a opinião da revista *Linhas Críticas*, de seus editores, ou da Universidade de Brasília.

Os autores são os detentores dos direitos autorais deste manuscrito, com o direito de primeira publicação reservado à revista *Linhas Críticas*, que o distribui em acesso aberto sob os termos e condições da licença Creative Commons Attribution (CC BY 4.0):  
<https://creativecommons.org/licenses/by/4.0>

