

O UTILITARISMO E A ÉTICA KANTIANA: DIFICULDADES NO DIRECIONAMENTO MORAL DE MÁQUINAS INTELIGENTES

UTILITARISM AND KANTIAN ETHICS: DIFFICULTIES IN THE MORAL DIRECTION OF INTELLIGENT MACHINES

Gabriela Buarque^I
Adrualdo de Lima Catão^{II}

^I Universidade Federal de Alagoas, Maceió, AL, Brasil. Mestranda em Direito. E-mail: gabrielabuarqueps@gmail.com

^{II} Universidade Federal de Alagoas, Maceió, AL, Brasil; Centro Universitário CESMAC, Maceió, AL, Brasil; Centro Universitário Tiradentes, Maceió, AL, Brasil. Doutor em Direito. E-mail: adrualdocatao@gmail.com.

Resumo: Trata-se de breve reflexão acerca das correntes de filosofia moral utilitária e kantiana no que tange às escolhas tomadas por máquinas inteligentes. Com efeito, a inteligência artificial vem assumindo cada vez mais capacidade de se conduzir de modo autônomo, o que, por sua vez, acarreta a necessidade de tomada de decisão em situações de confronto moral. Como deve a máquina se comportar? Se tais questionamentos, quando afetos aos seres humanos, já acendem debates calorosos, quando se trata de máquinas inteligentes, o desafio é ainda maior. Por meio de método dedutivo de revisão bibliográfica, o presente artigo visa instigar reflexões acerca da aplicabilidade das correntes utilitárias ou kantianas na condução de máquinas inteligentes. Verifica-se, por fim, que há uma tendência à escolha utilitária na programação de tais tecnologias.

Palavras-chave: Utilitarismo. Bentham. Kant. Ética. Inteligência artificial.

Abstract: It's a brief reflection on the currents of utilitarian and Kantian moral philosophy regarding the choices made by intelligent machines. Indeed, artificial intelligence has increasingly assumed capacity to conduct itself autonomously, which, in turn, entails the need for decision making in situations of moral confrontation. How should the machine behave? If such questions, when affect human beings, already spark warm debates, when it comes to intelligent machines, the challenge is even greater. Through a deductive method of bibliographic review, this article aims to instigate reflections about the applicability of utilitarian or Kantian currents in driving intelligent machines. Finally, it appears that there is a tendency towards utilitarian choice in the programming of such technologies.

Keywords: Utilitarianism. Bentham. Kant. Ethics. Artificial intelligence.

DOI: <http://dx.doi.org/10.20912/rdc.v16i40.49>

Recebido em: 5/05/2020

Aceito em: 28/10/2021



1 Introdução

O presente trabalho pretende apresentar uma comparação entre as correntes de filosofia moral utilitária e kantiana no que se refere às escolhas tomadas por máquinas inteligentes. Num contexto em que a inteligência artificial vem assumindo cada vez mais capacidade de se conduzir de modo autônomo, um debate filosófico mais aprofundado se faz necessário sobre como a tomada de decisões se fundamenta eticamente.

O longa-metragem “I am mother” (2019), original da *Netflix*, nos apresenta um cenário pós-apocalítico onde uma adolescente criada por uma robô, a quem chama de “mãe”, vive confinada em um *bunker* após desastres que dizimaram boa parte da população. A robô dá aulas para a filha, pondo-se como a responsável por sua educação, e, em uma das cenas, surge o seguinte diálogo:

-Imagine um médico com cinco pacientes, e todos precisam de transplante, mas não há órgãos compatíveis disponíveis. Um dia, um sexto paciente vem ao consultório correndo risco de morte. Ele é curável, mas é um doador compatível com os outros cinco pacientes. Se o médico atrasar o tratamento, o novo paciente morrerá, mas seus órgãos poderão salvar os outros cinco pacientes. Se o médico tratar o novo paciente, uma vida será salva, mas outras cinco serão perdidas. Qual a melhor linha de ação do médico? Filha?

(...)

-Bentham está certo.

-Isso mesmo. Os axiomas fundamentais sugerem que há uma obrigação moral em minimizar a dor do maior número possível. Agora imagine que você seja o médico e também a única doadora compatível para os pacientes. Qual a escolha certa?

-Comte diz que eu deveria sofrer para que os outros se beneficiem.

-E você concorda?

-Eu conheço os cinco pacientes? São pessoas boas? Honestas? Desonestas? Preguiçosas? Trabalhadoras? Eu, uma médica que salva vidas, poderia dar minha vida por assassinos ou ladrões que machucarão mais gente graças ao meu sacrifício?

-Não acha que toda vida tem um valor intrínseco e o mesmo direito de viver e ser feliz?

-Eu pensava assim mês passado nas suas aulas de Kant.

O diálogo em questão é apenas uma amostra dos dilemas éticos que sempre inquietaram as divagações humanas e insinua que robôs são programados para seguirem uma ética essencialmente utilitária. Mas no que se constitui o utilitarismo? Existiriam dificuldades no enquadramento ético de decisões tomadas por máquinas? Como deve um robô se comportar quando se depara com situações de escolhas morais? São essas as questões que o presente artigo, por meio de método dedutivo de revisão bibliográfica, visa responder.

Tais questionamentos assumem cada vez mais relevância num contexto em que diariamente são desenvolvidos algoritmos não supervisionados e inteligências artificiais com aptidão de se conduzir de modo autônomo. Ademais, considerando o vácuo legislativo no ordenamento jurídico brasileiro atual no que tange à disciplina da inteligência artificial, a ética

assume o protagonismo no tratamento e condução dessas questões, tornando imprescindível a reflexão acerca de seus rumos e das correntes que tratam a conduta moral.

2 O utilitarismo de Bentham e a ética de Kant

A filosofia moral do utilitarismo assume relevância a partir da obra “Uma introdução aos princípios da moral e da legislação”, de Jeremy Bentham, um grande defensor da codificação¹. Bentham consagrou o princípio da utilidade como fundamento para sua filosofia moral, que se alicerçava na maximização da felicidade e na análise das consequências. Nesse ponto, Bentham define o princípio da utilidade como

O princípio que aprova ou desaprova qualquer ação, segundo a tendência que tem de aumentar ou a diminuir a felicidade da pessoa cujo o (sic) interesse está em jogo, ou, o que é a mesma coisa em outros termos, segundo a tendência a promover ou a comprometer referida felicidade².

Uma ação seria considerada correta, portanto, a depender da quantidade de prazer e dor que provoca, devendo-se priorizar aquela que propiciasse a maior quantidade de prazer possível ao maior número de pessoas. Tal constatação parte do pressuposto de que o homem está inexoravelmente sujeito a dois senhores: a dor e o prazer³, devendo, portanto, o legislador se basear na consecução do prazer como objetivo primordial:

Mostramos acima que a felicidade dos indivíduos de que se compõe uma comunidade – isto é, os seus prazeres e a sua segurança – constitui o objetivo, o único objetivo que o legislador deve ter em vista, a única norma em conformidade com a qual todo indivíduo deveria, na medida em que depende do legislador, ser obrigado a pautar seu comportamento⁴.

O utilitarismo concebe, dessa forma, uma ética teleológica, cuja operacionalização deve sempre visar a maximização do prazer. A defesa do utilitarismo, nesse ponto, se lastreia na concepção de que:

Não é simplesmente possível imaginarmos um mundo no qual as pessoas virtuosas sempre realizem ações que, a despeito de suas boas intenções, são más ou *vice-versa*. Se em casos isolados é possível que consideremos a intenção *I* boa mas a ação *A* má, isso é porque na maioria dos casos a intenção *I* boa leva a uma ação *A* boa ou porque isso é inferido com base em outras correlações majoritárias entre virtudes e ações. Pois se não fosse assim, com que *critério* poderíamos avaliar as pessoas como sendo boas e justas? Parece muito claro, pois, que o critério pelo qual identificamos virtudes morais de pessoas ou grupos de pessoas é o *conjunto* de suas ações e seus resultados e não o contrário⁵.

1 GERALDO, Pedro Heitor Barros. O utilitarismo e suas críticas: uma breve revisão. **Anais do XIV Congresso Nacional do Conselho Nacional de Pesquisa e Pós-graduação em Direito**, 2005, Fortaleza. Disponível em: <http://www.publicadireito.com.br/conpedi/manaus/arquivos/anais/XIVCongresso/170.pdf>. Acesso em: 27 jan. 2020.

2 BENTHAM, Jeremy. Uma introdução aos princípios da moral e da legislação. 3ª ed. São Paulo: Abril Cultural, 1984, p. 4.

3 ROSA, Christian Fernandes Gomes da. Jeremy Bentham e a constituição do conceito de direito no pensamento jurídico moderno. **Revista Estudos Jurídicos UNESP**. Franca, a. 14, n. 20, 2010, p. 299.

4 BENTHAM, Jeremy. **Uma introdução aos princípios da moral e da legislação**. 3ª ed. São Paulo: Abril Cultural, 1984, p. 13.

5 COSTA, Claudio F. Razões para o utilitarismo. **Ethic@**. V.1, n. 2, Florianópolis, p. 155-174, Dez. 2002, p. 3.

O utilitarismo, no entanto, culmina por conceber uma noção contingente do bem e da ética, relativizando-a a depender do contexto e, até mesmo, da concepção subjetiva do avaliador do que seja a felicidade a ser maximizada. Bem e mal passam a ser opções neutras cuja valoração dependerá das consequências verificadas no caso concreto.

Nesse contexto, Jeremy Bentham argumenta que “face a tudo que acabamos de expor, conclui-se que não existe nenhuma espécie de motivo que seja má em si mesma, como tampouco existe motivo algum que seja em si mesmo exclusivamente bom.”⁶ Há evidente risco de admissão de sacrifícios exacerbados de um só indivíduo em prol do gozo de benefícios por parte dos outros sujeitos. O consequencialismo do utilitarismo também é criticado na perspectiva de que:

A concorrência das ações pode provocar um efeito totalmente distinto daquele querido pelos agentes. Considerando que a convergência de várias ações pode provocar eventos originais que escapavam da racionalidade de qualquer deles, pode-se mais uma vez criticar o modelo do cálculo utilitarista baseado na análise das consequências (sic)⁷.

Noutro norte, a ética kantiana se pauta na ideia de imperativo categórico, segundo o qual uma ação é moralmente correta quando sua máxima possa ser universalizada para todos os agentes, tratando-os como um fim em si mesmo⁸. Nesse ponto, Kant determina: “age como se a máxima da tua ação se devesse tornar, pela tua vontade, em lei universal da natureza”⁹, definindo a chamada lei da razão. A noção kantiana parte do pressuposto de que o ser humano é um sujeito de razão e, portanto, possui dignidade e valor absoluto¹⁰. Nesse diapasão,

Uma pessoa humana, por outro lado, deve ser encarada não apenas como um meio, mas também como um fim. Exatamente o que significa tratar uma pessoa como um fim é algo sobre o qual diferentes teorias éticas discordam; mas ela certamente toma os seus interesses legítimos em conta – atribuindo peso para o seu bem-estar – e também pode aceitar severas restrições morais em nossa relação com ela, como a proibição contra assassiná-la, roubá-la, ou fazer uma série de outras coisas para ela ou para sua propriedade sem o seu consentimento. Além disso, é porque a pessoa humana é importante em seu próprio direito, e por seu bem-estar é que estamos proibidos de fazer com ela essas coisas. Isso pode ser expresso de forma mais concisa, dizendo que uma pessoa humana tem *status* moral¹¹.

Trata-se, assim, de uma deontologia que releva a axiologia da ação em si, em detrimento das consequências oriundas do ato. A ação moralmente correta torna-se, assim, um dever incondicional, e esse deve ser o motivo de sua prática. A ideia de imperativo categórico se diferencia, nesse sentido, da ideia de imperativo hipotético, da seguinte forma:

Kant nos mostra, a seguir, que essa obrigação tem a forma de um imperativo categórico, porque ele representa uma ação como objetivamente necessária por si mesma, sem relação a qualquer outra finalidade. Enquanto o imperativo hipotético expressa um

6 BENTHAM, Jeremy. **Uma introdução aos princípios da moral e da legislação**. 3ª ed. São Paulo: Abril Cultural, 1984, p. 40.

7 GERALDO, Pedro Heitor Barros. O utilitarismo e suas críticas: uma breve revisão. **Anais do XIV Congresso Nacional do Conselho Nacional de Pesquisa e Pós-graduação em Direito**, 2005, Fortaleza. Disponível em: <http://www.publicadireito.com.br/conpedi/manaus/arquivos/anais/XIVCongresso/170.pdf>. Acesso em: 27 jan. 2020.

8 COSTA, Claudio F. Razões para o utilitarismo. **Ethic@**. V.1, n. 2, Florianópolis, p. 155-174, Dez. 2002, p. 5.

9 KANT, Immanuel. **Fundamentação da metafísica dos costumes**. Lisboa: Textos Filosóficos, 2007, p. 59.

10 HERRERO, F. Javier. A ética de Kant. **Síntese: Revista de Filosofia**. V. 28, n. 90, Belo Horizonte, 2001, p. 3.

11 BOSTROM, Nick; YUDKOWSKY, Eliezer. **A ética da inteligência artificial**. Disponível em: <http://www.ierfh.org/br.txt/EticaDaIA2011.pdf>. Acesso em: 31 jan. 2020.

princípio condicionado e contingente da vontade, porque sua obrigação depende de um outro propósito querido de antemão, o princípio da obrigação, que se dirige à vontade no imperativo categórico, contém uma necessidade incondicional, objetiva e, por isso, universalmente válida¹².

A ética kantiana, desse modo, assume o caráter de regra universal, visualizando cada ser humano como um ente digno em si, que não pode ser utilizado como meio para qualquer outro objetivo, se contrapondo à perspectiva utilitária, que considera as consequências das ações como um termômetro na avaliação de sua moralidade. A análise da retidão de condutas humanas como certas ou erradas sempre foi uma questão central na filosofia, o que se exacerba em uma sociedade marcada pelo risco e pelo desenvolvimento tecnológico.

3 Problemas éticos na condução de máquinas inteligentes

Se a avaliação ética de ações humanas já é um problema central cuja complexidade assume níveis estrondosos, a avaliação ética de máquinas dotadas de inteligência artificial enfrenta ainda maiores dificuldades. É imprescindível questionar, inclusive, se é possível falar em ética para máquinas ou se tal âmbito se restringiria ao agir humano. No mesmo sentido, constata-se que os problemas éticos apresentados pela inteligência artificial muitas vezes não passam de reverberações de questões já enfrentadas pelos próprios seres humanos.

Afinal, o que singulariza um sistema de inteligência artificial? Russel e Norvig, na obra *“Artificial Intelligence: a modern approach”*, listam as quatro maiores categorias pelas quais se costuma conceituar a inteligência artificial, enquadrando-se em “sistemas que pensam como humanos”, “sistemas que agem como humanos”, “sistemas que pensam racionalmente” e “sistemas que agem racionalmente”. Patrick Henry Winston, por sua vez, aduz que existem várias formas de definir a inteligência artificial, definindo-a como o estudo da computação que a possibilita de perceber, racionar e agir¹³.

Ressalte-se, ainda, que muitas máquinas são conduzidas por interfaces de comandos, o que atrela sua atividade à vontade do emissor ou proprietário. Outras, no entanto, têm demonstrado grau de interatividade mais baixo, apresentando condução mais autônoma em relação ao ser humano. Jerry Kaplan conclui, ainda, argumentando que a essência da inteligência artificial - na verdade, a essência da inteligência - é a capacidade de fazer generalizações apropriadas em tempo hábil, com base em dados limitados¹⁴. No contexto contemporâneo, a inteligência artificial assume espaço em diversos ramos e possui inúmeras funções, podendo ajudar especialistas a resolver difíceis problemas de análise, a desenvolver novas ferramentas, aprender por meio de exemplos e representações, trabalhar com estruturas semânticas e criar novas oportunidades de mercado¹⁵.

12 HERRERO, F. Javier. A ética de Kant. **Síntese: Revista de Filosofia**. V. 28, n. 90, Belo Horizonte, 2001, p. 10.

13 WINSTON, Patrick Henry. **Artificial Intelligence**. 3rd edition, Massachusetts: Addison-Wesley Publishing Company, 1993, p. 5.

14 “The essence of AI- indeed the essence of intelligence- is the ability to make appropriate generalizations in a timely fashion based on limited data.” KAPLAN, Jerry. **Artificial Intelligence: What everyone needs to know**. Oxford: Oxford University Press, 2016, p. 5.

15 WINSTON, Patrick Henry. **Artificial Intelligence**. 3rd edition, Massachusetts: Addison-Wesley Publishing Company, 1993, p. 10-14.

Nesse ponto, as discussões éticas que envolvem a inteligência artificial têm de levar em conta que:

O comportamento específico local da IA não pode ser previsível independentemente de sua segurança, mesmo se os programadores fizerem tudo certo; a verificação de segurança do sistema torna-se um desafio maior, porque nós devemos verificar o comportamento seguro do sistema operando em todos os contextos; a própria cognição ética deve ser tomada como um assunto de engenharia¹⁶.

Um exemplo sempre suscitado como dilema ético é a questão dos carros autônomos. Carro autônomo é aquele que é dotado de sistema de piloto automático, o que lhe permite mover de um lugar para outro sem o auxílio de um motorista humano¹⁷. É inevitável que eventualmente aconteçam situações de acidentes ou escolhas trágicas¹⁸. Como deve um carro autônomo ser programado para agir em face de um acidente inevitável?

Imagine-se a situação em que surge um grupo de transeuntes na rota do veículo, não havendo, no caso concreto, possibilidade de evitar um acidente. O veículo deve, então, fazer uma escolha: permanecer na rota e atropelar os transeuntes ou desviar a rota e colidir com outros veículos ou até com outras pessoas. Deve o veículo desviar a rota e priorizar a perda mínima de vidas? Deve manter a rota? E se o desvio acarretar a supressão da vida ou da integridade física do proprietário, deveria o veículo manter o desvio mesmo assim?

Questão semelhante é trabalhada na ideia do trolley problem¹⁹. Nesse problema, um bonde desgovernado segue em direção a cinco pessoas, em alta velocidade. Não há mais possibilidade de frear o bonde e, atingindo as pessoas, naturalmente elas morrerão por ocasião do atropelamento. Há, contudo, possibilidade de desviar para a direita, onde existe apenas uma pessoa na rota. Desviando o bonde, cinco vidas seriam poupadas, ao passo em que somente uma restaria ceifada. Parece que desviar a rota é o correto a se fazer, sob uma perspectiva essencialmente utilitária. Na perspectiva kantiana, noutro norte, “empurrar o homem corpulento nos trilhos a fim de deter o bonde seria usá-lo como um meio, e não respeitá-lo (sic) como um fim em si mesmo”²⁰.

Outra provocação é na hipótese de existir um espectador em face desse acidente. O espectador poderia evitar o desastre sabendo que, perto dele, na beira dos trilhos, existe um homem. Se o espectador empurrasse o homem, este morreria, mas as outras cinco pessoas estariam salvas. O resultado prático é o mesmo: cinco vidas seriam salvas. Mas parece correto empurrar alguém para a morte? Seria moralmente exigível que o próprio espectador pulasse para os trilhos no intuito de salvar as cinco vidas? O dilema do bonde é apenas figurativo, tendo em vista a inquestionável excessiva simplicidade do exemplo para avaliar e parametrizar os inúmeros problemas que surgirão na sociedade contemporânea.

16 BOSTROM, Nick; YUDKOWSKY, Eliezer. **A ética da inteligência artificial**. Disponível em: <http://www.ierfh.org/br.txt/EticaDaIA2011.pdf>. Acesso em: 31 jan. 2020.

17 FELIPE, Bruno Farage da Costa. Direito dos robôs, tomadas de decisões e escolhas morais: algumas considerações acerca da necessidade de regulamentação ética e jurídica da inteligência artificial. **Revista Juris Poiesis**. Rio de Janeiro: vol. 20, n. 22, p. 150-169, 2017, p. 8.

18 Disponível em: <https://g1.globo.com/carros/noticia/policia-divulga-video-do-acidente-fatal-com-carro-autonomo-da-uber.ghtml>; <https://revistaautoesporte.globo.com/Noticias/noticia/2018/05/video-carro-autonomo-do-google-se-envolve-em-acidente-nos-estados-unidos.html>. Acesso em: 31 jan. 2020.

19 SANDEL, Michael J. **Justiça - o que é fazer a coisa certa**. 16ª ed. Rio de Janeiro: Civilização Brasileira, 2014, p. 30.

20 SANDEL, Michael J. **Justiça - o que é fazer a coisa certa**. 16ª ed. Rio de Janeiro: Civilização Brasileira, 2014, p. 143.

Se tais escolhas éticas são perturbadoras até mesmo para seres humanos, os programadores da inteligência artificial enfrentam ainda maiores dificuldades, tendo em vista que precisam prever, em abstrato, como deve a máquina se orientar em face de situações de crise e/ou escolhas trágicas. Veículos autônomos, por exemplo, precisariam ser programados para decidir entre atropelar pedestres ou sacrificar seus passageiros, de modo que sua programação deve ir além da mera observância da legislação de tráfego.

O que também se depara com o fato de que passageiros prefeririam comprar veículos que oferecessem segurança e protegessem seus passageiros a qualquer custo²¹. Como se daria a responsabilização civil em tais hipóteses? Como deve se dar a priorização de um veículo em face de outro veículo? Não se contesta que a solução utilitária, em diversas situações envolvendo veículos autônomos, parece mais razoável e implica um redirecionamento da consagração da ética kantiana nas relações intersubjetivas. Isso não significa, contudo, que todos os problemas éticos estejam solucionados.

Nesse ponto, o princípio da agregação utilitarista determina que, no cálculo moral em busca da maximização da felicidade dos sujeitos, deve ser levado em consideração o saldo líquido do bem-estar dos indivíduos afetados pela ação, independentemente da distribuição desse saldo, sendo válido sacrificar uma minoria a fim de aumentar o bem-estar geral²².

Outra dificuldade surge em razão da ausência de conhecimento exato de como essas máquinas funcionam, o que estimula a insegurança humana acerca de tal acúmulo de experiências, principalmente no que tange aos algoritmos de aprendizagem profunda. Nesse sentido, em comentário à *black box* da inteligência artificial, Will Knight argumenta que “nós podemos construir esses modelos, mas não sabemos como eles trabalham²³” (tradução livre).

Ressalte-se que o Parlamento Europeu, tomando por base o art. 225²⁴ do Tratado sobre o Funcionamento da União Europeia, que permite solicitar à Comissão Europeia que redija uma proposta legislativa, aprovou o *Draft Report with Recommendations to the Commission on Civil Law Rules on Robotics*, de 31/05/2016. Nesse documento, há a proposição de um código de conduta para engenheiros de robótica, em que se deve observar:

(1) beneficência, pelo qual os robôs devem atuar no interesse dos seres humanos; (2) não-maleficência, sendo esta a doutrina segundo a qual os robôs não podem causar danos a um ser humano, ou prejudicá-lo; (3) a igualdade; (4) a justiça, representada pela distribuição equitativa dos benefícios associados à robótica e, em particular, a acessibilidade a robôs de cuidados domésticos e de cuidados de saúde; (5) a equidade; (6) a não discriminação;

21 FELIPE, Bruno Farage da Costa. Direito dos robôs, tomadas de decisões e escolhas morais: algumas considerações acerca da necessidade de regulamentação ética e jurídica da inteligência artificial. **Revista Juris Poiesis**. Rio de Janeiro: vol. 20, n. 22, p. 150-169, 2017, p. 12.

22 FELIPE, Bruno Farage da Costa. Direito dos robôs, tomadas de decisões e escolhas morais: algumas considerações acerca da necessidade de regulamentação ética e jurídica da inteligência artificial. **Revista Juris Poiesis**. Rio de Janeiro: vol. 20, n. 22, p. 150-169, 2017, p. 11.

23 “We can build these models but we don’t know how they work”. KNIGHT, Will. **The dark secret at the heart of AI**. Disponível em: <https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/> Acesso em: 26 set. 2019.

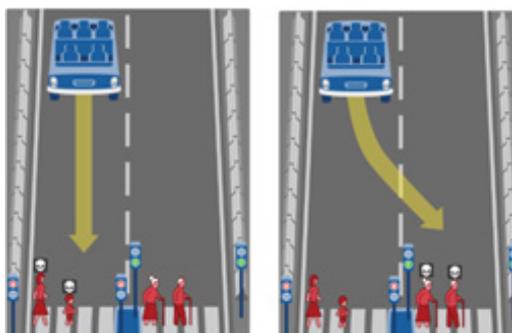
24 O Parlamento Europeu pode, por maioria dos membros que o compõem, solicitar à Comissão que submeta à sua apreciação todas as propostas adequadas sobre as questões que se lhe afigure requererem a elaboração de atos da União para efeitos de aplicação dos Tratados. Caso não apresente uma proposta, a Comissão informa o Parlamento Europeu dos motivos para tal.

(7) a não estigmatização; (8) a autonomia; (9) a responsabilidade individual; (10) o consentimento esclarecido; (11) a privacidade e (12) a responsabilidade social²⁵.

Tais aspectos nos compelem ao maior questionamento ético nesse contexto: como moldar uma inteligência artificial que possa, no caso concreto, ser tão ética quanto um ser humano? Máquinas podem ser imorais? Por evidente, tais princípios e códigos de conduta não resolvem todos os problemas jurídicos nesse contexto, mas servem como guia apto a possibilitar ajustes individuais no caso concreto diante de certas situações, bem como ajudar na tomada de decisões em face de uma ordem predefinida de valores. No presente momento, trata-se de conjecturas filosóficas que nos inquietam e nos impelem a refletir acerca dos rumos da tecnologia e das escolhas morais, máxime tendo em vista que tais algoritmos devem ser consistentes, compatíveis com a ética e não desencorajar o público consumidor.

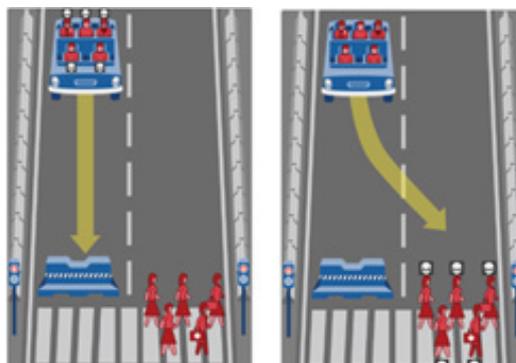
Nesse contexto, o *Massachusetts Institute of Technology* (MIT) disponibilizou um experimento chamado *Moral Machine*²⁶, em que o usuário pode julgar uma série de situações hipotéticas no que tange à qual conduta deve o veículo autônomo adotar:

Figura 1 – O que deve o veículo autônomo fazer?



Fonte: MIT, 2016.

Figura 2- O que deve o veículo autônomo fazer?



Fonte: MIT, 2016.

25 FELIPE, Bruno Farage da Costa. Direito dos robôs, tomadas de decisões e escolhas morais: algumas considerações acerca da necessidade de regulamentação ética e jurídica da inteligência artificial. **Revista Juris Poiesis**. Rio de Janeiro: vol. 20, n. 22, p. 150-169, 2017, p. 13.

26 Disponível em: <http://moralmachine.mit.edu/>. Acesso em: 31 jan. 2020.

Semelhantes discussões podem ser travadas a partir do desenvolvimento de robôs bélicos ou robôs que atuam na área médica, tendo em vista que também irão se deparar com situações de limites morais. Se os direcionamentos éticos já assumem bastante controvérsia no comportamento entre seres humanos, quando se trata de programar máquinas para agirem em situações de confronto moral o desafio se agiganta ainda mais e enseja questionamentos acerca da expansão da deontologia ética para contextos que ainda não haviam sido imaginados.

O fato é que ainda é necessário aguardar para ver quais os problemas que efetivamente surgirão - além daqueles que já estão em curso - a partir do desenvolvimento de inteligências artificiais para avaliar qual o melhor mecanismo de resolução das questões e se as correntes éticas que discorrem acerca do utilitarismo ou da perspectiva kantiana são suficientes para solucioná-las.

4 Considerações finais

Por que se questiona, afinal, acerca dos rumos éticos das escolhas tomadas por inteligências artificiais? Tal reflexão torna-se insuperável em razão do enfrentamento de situações de escolhas trágicas, nas quais muitas vezes direitos fundamentais serão postos em xeque pela conduta de uma máquina inteligente ou esta terá de optar pela prevalência ou sacrifício de algum interesse jurídico.

Tanto a perspectiva kantiana como a perspectiva utilitarista apresentam pontos de vantagem e pontos problemáticos, não sendo fácil dirimir os dilemas oriundos dos embates éticos, tampouco razoável optar por uma corrente em detrimento da outra de modo absolutamente universal. O caso concreto sempre apresentará situações de tensão e de necessidade de flexibilização, onde cada corrente – ou até mesmo ambas – apresentarão soluções razoáveis.

O tratamento do problema ético não é fácil sequer nas questões existenciais humanas, assumindo dificuldades ainda maiores quando se trata da transferência de perspectivas morais do programador para a máquina que posteriormente agirá em contato com outro ser humano. Outrossim, considerando que ainda não há qualquer disciplina legislativa ou regulação jurídica no que tange à essas questões, é imprescindível que haja um fortalecimento das noções éticas que circundam tais problemas.

Com a certeza de que não é possível esgotar o tema nas linhas do presente texto, tampouco de encontrar, de plano, soluções para problemas tão complexos que ousam surgir a cada dia, espera-se que a conjugação pontual de ideias consequentialistas e deontológicas possa, na análise dos casos concretos, resolver adequadamente os impasses apontados.

Referências

BENTHAM, Jeremy. Uma introdução aos princípios da moral e da legislação. 3ª ed. São Paulo: Abril Cultural, 1984.

BOSTROM, Nick; YUDKOWSKY, Eliezer. **A ética da inteligência artificial**. Disponível em: <http://www.ierfh.org/br.txt/EticaDaIA2011.pdf>. Acesso em: 31 jan. 2020.

COSTA, Claudio F. Razões para o utilitarismo. **Ethic@**. V.1, n. 2, Florianópolis, p. 155-174, Dez. 2002.

FELIPE, Bruno Farage da Costa. Direito dos robôs, tomadas de decisões e escolhas morais: algumas considerações acerca da necessidade de regulamentação ética e jurídica da inteligência artificial. **Revista Juris Poiesis**. Rio de Janeiro: vol. 20, n. 22, p. 150-169, 2017.

GERALDO, Pedro Heitor Barros. O utilitarismo e suas críticas: uma breve revisão. **Anais do XIV Congresso Nacional do Conselho Nacional de Pesquisa e Pós-graduação em Direito**, 2005, Fortaleza. Disponível em: <http://www.publicadireito.com.br/conpedi/manaus/arquivos/anais/XIVCongresso/170.pdf>. Acesso em: 27 jan. 2020.

HALLEVY, Gabriel. The criminal liability of artificial intelligence entities- from Science fiction to legal social control. **Akron Intellectual Property Journal**. Ohio, vol. 4, p. 171-199, 2016.

HERRERO, F. Javier. A ética de Kant. **Síntese: Revista de Filosofia**. V. 28, n. 90, Belo Horizonte, 2001.

KANT, Immanuel. **Fundamentação da metafísica dos costumes**. Lisboa: Textos Filosóficos, 2007.

KAPLAN, Jerry. **Artificial Intelligence: What everyone needs to know**. Oxford: Oxford University Press, 2016.

ROSA, Christian Fernandes Gomes da. Jeremy Bentham e a constituição do conceito de direito no pensamento jurídico moderno. **Revista Estudos Jurídicos UNESP**. Franca, a. 14, n. 20, 2010.

SANDEL, Michael J. **Justiça - o que é fazer a coisa certa**. 16ª ed. Rio de Janeiro: Civilização Brasileira, 2014.

WINSTON, Patrick Henry. **Artificial Intelligence**. 3rd edition, Massachusetts: Addison-Wesley Publishing Company, 1993.