

Direitos dos robôs, tomadas de decisões e escolhas morais: algumas considerações acerca da necessidade de regulamentação ética e jurídica da inteligência artificial.

Robots' rights, decision-making and moral choices: some considerations about the need for ethical and legal regulation of artificial intelligence.

Bruno Farage da Costa Felipe¹

RESUMO

O presente artigo trabalha com a hipótese de que a Inteligência Artificial desenvolvida em robôs nos últimos tempos, caso não seja regulamentada, pode se tornar uma ameaça à humanidade. Defende-se a premente necessidade na criação de um “código de condutas” para o desenvolvimento da inteligência artificial, no intuito de se atender uma “moralidade algorítmica”, assim como o estabelecimento de estatuto(s) jurídico(s) que tenha(m) a capacidade de identificar de forma exata a natureza jurídica da AI frente às categorias jurídicas existentes, fixando regras concernentes à responsabilização jurídica dos engenheiros, empresas e indivíduos que desenvolvam, vendam ou utilizem a tecnologia.

PALAVRAS-CHAVE: Robôs; Inteligência Artificial; Direito; Moral; Responsabilização; Regulamentação.

ABSTRACT

This article works with the hypothesis that the Artificial Intelligence (AI) developed in robots in recent times, if not regulated, can become a threat to humanity. The urgent need to create a "code of conduct" for the development of

¹ Mestre em Teoria e Filosofia do Direito pela Universidade do Estado do Rio de Janeiro (UERJ). Graduado em Direito pela Faculdade Metodista Granbery – Juiz de Fora – MG. (2011). Professora da Faculdade Metodista Granbery.

artificial intelligence, in order to attend an "algorithmic morality" is defended, as well as the establishment of legal statute (s) with the ability to accurately identify the legal nature of AI in relation to existing legal categories, establishing rules concerning the legal accountability of engineers, companies and individuals who develop, sell or use the technology.

KEYWORDS: Artificial intelligence; Law; Moral; Responsibility; Regulation.

1. Introdução.

Averiguar se certas condutas praticadas por indivíduos são certas ou erradas, justas ou injustas, sempre foi e talvez sempre será um problema central da filosofia do direito. A problemática entorno dos julgamentos das posições tomadas pelas pessoas parece tangenciar ao infinito, principalmente com a complexidade do mundo contemporâneo e seus inerentes conflitos morais.

Não há novidade em falar nos conflitos e dilemas morais que enfrentamos no cotidiano, em especial na seara do direito. A questão que se propõe analisar neste artigo, todavia, é: e quando o parâmetro não são condutas humanas, mas condutas de robôs dotados de inteligência artificial? É possível dizer que um robô agiu de maneira imoral? Uma máquina pode cometer injustiças? Devem existir padrões morais, quiçá jurídicos, que disciplinem as condutas dos robôs?

Desde o Frankenstein de Mary Shelley², ou do Golem de Praga³, passando pela história do Robô de Karel Čapek⁴, também pela história de Maria, a robô revolucionária

²Frankenstein ou o Moderno Prometeu (*Frankenstein: or the Modern Prometheus*, no original em inglês), romance de terror gótico com inspirações do movimento romântico, da autora de Mary Shelley. O romance relata a história de Victor Frankenstein, um estudante de ciências que constrói um monstro em seu laboratório.

Para mais detalhes, ver SHELLEY, Mary W. Frankenstein.

³ História em que o rabino Judah Loew Ben Betzalel, de Praga, durante o século XVI criou um golem para defender o gueto de Josefov contra ataques anti-semitas. Para mais, ver PASCHELES, Wolf. Sippurim.

⁴Karel Čapek foi um escritor Tcheco, para muitos o criador da palavra Robot (a partir de robota, que, em sua língua e em outras línguas eslavas, pode significar trabalho exercido de forma compulsória, ou escravo). Autor da obra denominada "R.U.R." (*Rosumovi Univerzální Roboti*), a qual narra a história de um cientista chamado Rossum, que desenvolve uma substância para construção de humanoides (robôs), com o intuito de que estes sejam obedientes e realizem todo o trabalho físico. Para mais, ver CAPEK, Karel. R.U.R.: Rosumovi Umeli Roboti.

porta-voz da classe trabalhadora em Metrópolis⁵; assim como pelo famoso R2-D2 de “Guerras nas Estrelas”⁶; até chegarmos no Robocop da cidade de Detroit⁷; no T-1000 de “o Exterminador do Futuro”⁸; e, recentemente, com as séries Blackmirror⁹ e Westworld¹⁰, as pessoas têm fantasiado acerca da possibilidade de construirmos máquinas inteligentes, androides com características humanas. O que até então era obra da ficção científica e campo da nossa imaginação, tem se tornado uma realidade cada vez mais próxima.

Essa divagação inicial pode aparentar uma suposição fictícia, sem embasamento ou compromisso com a realidade, ou uma especulação para um futuro muito distante. Felizmente – ou não - isso não é verdade. O tema é atual e merece urgência nos debates filosóficos e jurídicos.

Antes de prosseguir, indispensável delimitar alguns conceitos iniciais. O que entendemos por Robô e Inteligência Artificial?

Basicamente, robô nada mais é que uma máquina- especialmente programável por um computador - capaz de executar uma série complexa de ações automaticamente (OXFORD, 2017). Por sua vez, a definição do termo “inteligência artificial” –a partir de agora citaremos apenas como AI (*Artificial Intelligence*)–está intrinsecamente ligada à capacidade de desenvolvimento de inteligência nos robôs, a qual alguns denominam *racionalidade* (RUSSELL; NORVIG, 2009). Também pode ser delimitada como “um esforço em tornar computadores em máquinas com mentes, no sentido pleno e literal” (HAUGELAND, 1985); ou “a automação de atividades associadas ao pensamento humano, como a tomada de decisões e resolução de problemas” (BELLMAN, 1978); ou

⁵ Metrópolis (título original: Metropolis) é um filme alemão de ficção científica lançado em 1927, dirigido pelo cineasta austríaco Fritz Lang. Ver METROPOLIS. Direção de Fritz Lang.

⁶ O famoso dróide astromecânico, responsável por manutenção e navegação de astronaves nos filmes da franquia “Guerra nas Estrelas”. Ver STAR Wars: Episode IV – A New Hope.

⁷ RoboCop baseia-se na história de um policial, Alex Murphy assassinado de forma brutal por um grupo de criminosos, sendo revivido pela *Omni Consumer Products (OCP)*, como um ciborgue força da lei conhecido como “RoboCop”. Interpretado pelo ator Peter Weller. Vide ROBOCOP. Direção de Paul Verhoeven.

⁸ T-1000 é um androide mimético fictício do filme Terminator 2: Judgment Day.

⁹ Série de televisão britânica de ficção científica satírica que trabalha com temas relacionados às consequências imprevistas das novas tecnologias. Em um dos seus episódios, denominado “*Be Right Back*”, a série desenvolve a possibilidade de criação de androides dotado de AI para trazer de volta, por meio de banco de dados, pessoas próximas falecidas. Ver BLACK Mirror. Direção de Charlie Brooker

¹⁰ Westworld é uma série de televisão baseada no filme homônimo de 1973. A história se passa em um parque temático que simula o Velho Oeste e é povoado por andróides sintéticos dotados de AI apelidados de “anfitriões”, os quais atendem aos desejos dos ricos visitantes do parque. Ver WESTWORLD. Produção de Jonathan Nolan. Roteiro: Athena Wickham.

ainda: “a arte de criar máquinas que executam funções que exigem inteligência quando realizada por pessoas” (KURZWEIL, 1990).

Recentemente, intelectuais e grandes nomes da comunidade científica têm demonstrado preocupação com a questão do desenvolvimento da Inteligência artificial e suas implicações. Um dos mais consagrados cientistas da atualidade, o físico Stephen Hawking, por exemplo, já se manifestou – num artigo em coautoria com outros pesquisadores, a saber: Max Tegmark, Stuart Russell e Frank Wilczek– acerca da potencialidade de a AI ser uma ameaça. Segundo Hawking et al. (2014), a ameaça real à humanidade não deve ser pensada como relacionada a um agente externo, mas interno. Ou seja, o mal pode ser fruto de nossas próprias criações, como é o caso da AI.

O sucesso em criar inteligência artificial seria o maior evento na história da humanidade. Infelizmente, também pode ser o último, a menos que aprendamos a evitar seus males. Os cientistas acreditam que todos nós deveríamos nos perguntar o que podemos fazer agora para aumentar as chances de potencializar os benefícios da inteligência artificial, evitando os seus riscos (HAWKING et al. 2014).

Em outro artigo recente, Hawking (WHIPPLE, 2017) voltou a afirmar que a tecnologia precisa ser controlada para que não acabe destruindo a espécie humana. Segundo Hawking (GRIFFIN, 2015), o verdadeiro risco da AI não reside em uma “essência maldosa”, mas em sua competência. Um robô dotado de AI pode ser extremamente eficaz em alcançar seus objetivos e, caso estes objetivos não estejam alinhados com os nossos, podemos ter sérios problemas¹¹.

Também apoiam esse discurso, dentre outros, Bill Gates e Elon Musk. Enquanto o primeiro (RAWLINSON, 2015) afirma não entender o fato de as pessoas não estarem ainda preocupadas com a possibilidade de a AI ser uma ameaça, o segundo está convencido de que nós devemos ter muito cuidado a respeito da inteligência artificial, sendo esta a maior ameaça à nossa existência. Para Musk (GIBBS, 2014), deveria haver alguma espécie de regulamentação sobre a questão, talvez em nível nacional e

¹¹ Um exemplo utilizado por Hawking, a título de ilustração, é sobre formigas e formigueiros: o físico alega que nós, humanos, não odiamos formigas, mas se nosso objetivo é implementar um projeto sustentável de energia hidrelétrica – por exemplo- e há um formigueiro na região que será alagada, certamente iremos exterminar essas formigas. Para Hawking, não podemos colocar a humanidade na posição análoga à destas formigas.

internacional, apenas para termos certeza de que não faremos algo que ponha nossa existência em risco, quando da criação de robôs dotados deste tipo de inteligência.

Embora a AI tenha gerado frutos benéficos até agora, como assistentes pessoais para computadores e smartphones, carros autônomos, *drones*, dentre outros, também é perceptível o uso não seguro desse tipo de inteligência no campo militar, como na produção de armas autônomas, as quais podem rastrear e eliminar seus alvos automaticamente.

A construção de armas letais autossuficientes já vem sendo alvo de críticas a ponto de levar instituições e acadêmicos ao redor do mundo a advogarem uma proibição em massa desse tipo de tecnologia. ASARO (2012, p. 687), por exemplo, defende uma teoria para o banimento desses tipos de armas autônomas que seja fundamentada nos direitos humanos e nos princípios humanitários, com suporte não apenas em padrões morais, mas também jurídicos.

No dia 31 de maio de 2016, o Parlamento Europeu – *TheEuropeanParliament* – apresentou uma moção, que contém recomendações à Comissão sobre disposições de Direito Civil sobre Robótica, com conteúdo alarmante: o parlamento considera a possibilidade de, dentro de algumas décadas, a AI ultrapassar o intelecto humano de tal modo que irá criar sérios problemas à capacidade humana em conter a sua própria criação e, por conseguinte, isso pode acarretar em perigos à nossa existência (AFFAIRS, 2016, p. 4).

Os parlamentares (2016) ressaltaram que a AI está prestes a desencadear uma nova revolução industrial, a qual atingirá todos os segmentos da sociedade, sendo, por isso, de vital importância que o direito – com suas legislações – considere essas implicações.

Acerca dessa “nova revolução industrial”, vale considerar o recente relatório do McKinsey Global Institute, segundo o qual quase metade de todo o trabalho que fazemos será automatizado até o ano de 2055, a não ser que uma variedade de fatores, incluindo a política e o sentimento público em relação à tecnologia, possam frear esse avanço ou encontrar soluções para o convívio harmônico entre empregados humanos e robôs (MANYIKA et al., 2017).

Na mesma linha de raciocínio, o co-fundador da Apple e programador de computadores, Steve Wozniak, acredita que se construirmos robôs com inteligência

artificial ilimitada, com a capacidade de se inserirem em qualquer ramo da vida, eventualmente eles pensarão mais rápido do que nós e, fatalmente, irão “eliminar os humanos lentos” para administrar empresas com mais eficiência¹²(WOOLASTON, 2015).

As fontes de preocupação não residem somente no desenvolvimento equivocado deste tipo de inteligência ou na falta de regulamentações legais para suas condutas. Preocupa, também, as possibilidades de falhas no código de um robô, assim como avarias no sistema ou de ataques informáticos (*hacking*) a sistemas robóticos interligados.

No patamar no qual se encontra hoje a sociedade, acontecimentos como estes podem ser catastróficos, pois as aplicações autônomas desenvolvidas estão sendo utilizadas em ascendente escala, sejam elas destinadas a carros e a aeronaves pilotadas à distância (*drones*), a robôs que prestam assistência, ou a robôs utilizados para a manutenção de ordem pública e policiamento (AFFAIRS, 2016).

Outro aspecto a se considerar diz respeito a questões afetivas. Para o parlamento (2016), eventualmente os robôs dotados de AI substituirão os cuidados e as companhias dos humanos e isso fatalmente afetará aspectos relacionados à dignidade humana. Impossível, portanto, que sem regulamentação específica, o desenvolvimento e utilização desta tecnologia não se esbarre e até mesmo colida com princípios de hierarquia constitucional a serem preservados.

A evolução tecnológica observada na última década é tamanha que os robôs com AI além de realizarem atividades que outrora eram exclusivas a seres humanos, também desenvolveram características autônomas e cognitivas. A autonomia de um robô pode ser definida como a capacidade de tomar decisões e de as aplicar no mundo exterior, independentemente do controle ou da influência externa. Esta autonomia possui uma natureza puramente tecnológica e o seu grau depende de como foi concebido o seu nível de sofisticação na interação de um robô com o seu ambiente (AFFAIRS, 2016, p. 5).

O Parlamento Europeu (2016, p. 7) propõe uma definição comum para os robôs autônomos inteligentes, para fins de regulamentação. Em geral, robôs autônomos, dotados de Inteligência artificial, seguem as seguintes características: (1) adquire

¹²Wozniak faz o seguinte questionamento: “Nós seremos os deuses? Nós seremos os animais da família? Ou nós seremos as formigas sendo pisoteadas?”.

autonomia através de sensores e/ou através da troca de dados com o seu ambiente (interconectividade) e troca e analisa dados; (2) aprende por si mesmo (critério opcional); (3) possui um suporte físico; (4) adapta o seu comportamento e as suas ações ao ambiente em que se encontra.

Conforme REIS (2003), características como estas distinguem um agente inteligente – ou agente autônomo - de um mero objeto. Enquanto os objetos possuem autonomia sobre o seu estado (dados), por outro lado, não possuem autonomia sobre o seu comportamento, ou seja, se disponibilizarem um determinado método, sempre que outro objeto o invoque este estará disponível. Pelo contrário, os agentes dotados de inteligência artificial possuem controle sobre o seu comportamento e, como tal, um outro agente terá de solicitar ao agente que execute uma dada ação. Este pode decidir, em cada situação concreta, se irá efetuar ou não a ação solicitada.

De acordo com a moção do parlamento (2016, p. 6) a capacidade dos robôs de aprenderem com a experiência e de tomarem decisões independentes os tornaram cada vez mais similares a agentes que interagem com o seu ambiente e que conseguem alterá-lo de forma significativa.

Essa interação com o ambiente e a capacidade de modificá-lo é bem delimitada por FRANKLIN e GRAESSER (1997), segundo os quais um agente autônomo é um sistema situado e que faz parte de um dado ambiente, que sente esse ambiente e age nele ao longo do tempo, de forma a realizar a sua própria agenda e de forma a afetar o que sentirá no futuro.

Ainda sobre essa capacidade de adaptação e aprendizado, em recente estudo, Bringsjord et al. (2015) aferiram que robôs dotados de AI passaram em testes humanos de autoconsciência. Os pesquisadores apontam a autoconsciência como fenômeno essencial no mundo social, eis que somos moralmente competentes apenas porque sabemos o que devemos fazer. Por exemplo: caso alguém seja coagido, com uma arma apontada para sua cabeça, a roubar um produto em uma loja próxima, este alguém sabe que não possui culpa sobre o ato, sendo que esta deve recair sobre aquele que pratica a coação. Tal diagnóstico depende de autoconsciência.

Para Bringsjord et al. (2015), a culpa é um conceito-chave no discurso moral, e, obviamente, afirmações como “Eu não sou culpado” estão intrinsecamente ligadas à autoconsciência. Os pesquisadores aplicaram testes em robôs e puderam afirmar que

estes são capazes de refletir acerca de uma ação no mundo real para inferir em conclusões verdadeiras, o que demonstra um estágio, ainda que inicial, de consciência¹³.

Nesse contexto, a necessidade de se estabelecer parâmetros para a responsabilidade jurídica das ações lesivas de um robô tem se tornado um problema iminente. Fato é que, quanto maior a autonomia nos robôs, menos estes podem ser percebidos como mero instrumentos ou máquinas destinadas a realizar atividades específicas para os humanos. Por esse motivo, o Parlamento Europeu acredita que as normas ordinárias em matéria de responsabilidade são, hoje, insuficientes. Há a necessidade de se instituir normas centradas no modo como se pode responsabilizar – no todo ou em parte – a máquina pelas suas ações ou omissões.

2. Tomadas de decisões, moralidade algorítmica e responsabilização jurídica dos robôs.

O nível de inserção da AI na sociedade contemporânea apenas cresce. Exemplo próximo está relacionado aos carros autônomos. Carro autônomo consiste em um automóvel dotado de sistema de piloto automático, o que lhe permite mover de um lugar para outro sem o auxílio de um motorista humano. O único papel do ser humano, ao menos no protótipo ideal, seria o de indicar o destino do automóvel.

A implementação desta tecnologia pode, em tese, levar a muitas melhorias no transporte, incluindo a redução nos acidentes de carro, eficiência no trânsito e aumento na capacidade das rodovias (THRUN, 2010).

O problema central, contudo, talvez resida no aspecto da segurança, principalmente porque – em especial ao se falar em um período de transição de tecnologia – estes carros terão que navegar ao lado de carros normais, dirigidos por motoristas humanos.

Dotados de capacidade de escolha, os carros autônomos com implementação AI eventualmente terão que fazer opções, sendo indispensável uma regulamentação específica para apuração dessas condutas. Em maio de 2016, um carro semiautônomo – da empresa Tesla, modelo S- envolveu-se em um acidente considerado o primeiro na

¹³ Para mais detalhes, ver BRINGSJORD, Selmer et al. Real robots that pass human tests of self-consciousness.

história dos semiautônomos a acarretar em uma fatalidade. O acidente aconteceu no dia 7, na cidade de Williston, na Flórida. O motorista morreu após colidir com seu Tesla Model S enquanto utilizava a função de piloto automático (TRANSPORTATION, 2016).

Nos relatórios preliminares, constatou-se que o automóvel não conseguiu frear quando um caminhão fez uma curva à esquerda na frente do veículo em um cruzamento. O departamento de transportes dos Estados Unidos (NHTSA) encerrou as investigações concluindo que não foi encontrada nenhuma evidência de falha no veículo (TRANSPORTATION, 2016).

Na ocasião, a empresa Tesla divulgou uma nota lamentando o ocorrido, mas dando a entender que não se responsabiliza por acidentes causados no piloto automático. Segundo a empresa “O piloto automático está melhorando com o passar do tempo, mas não é perfeito e, portanto, ainda exige que o motorista permaneça alerta. No entanto, quando utilizado em conjunto com a atenção do motorista, os dados mostram de forma inequívoca que o piloto automático reduz a carga de trabalho do motorista e resulta numa melhoria estatisticamente significativa na segurança quando em comparação com a direção totalmente manual” (TEAM, 2016).

Uma das empresas que também vem implementando o sistema de piloto automático em seus veículos é a Volvo. Diferentemente da Tesla, esta já se manifestou no sentido de que irá se responsabilizar por qualquer acidente quando seus carros estiverem operando no modo totalmente autônomo (KOROSEC, 2015).

Mesmo que o acidente recente tenha envolvido um veículo semi-autônomo, ou seja, não dotado completamente de inteligência artificial, a tendência é que se atinja o patamar da autonomia total em poucos anos. A questão que urge diante de situações como esta é: quem é o responsável pelos danos? Como decidir questões judiciais que envolvam condutas de robôs dotados de inteligência artificial? Deve haver um estatuto de responsabilização próprio para os robôs e seus criadores?

Defende-se, neste trabalho, a hipótese de que as legislações das searas cível e trabalhista não estão preparadas, ainda, para lidar com tamanha complexidade. Fundamental, portanto, preparar o direito para questões desta estirpe.

Além disso, engenheiros estão enfrentando dilemas éticos na programação e desenvolvimento dos sistemas de carros autônomos. A medida em que ascendem

patamares de autonomia, maiores serão as situações de escolha da própria máquina, dispensando comandos dos proprietários.

Como o carro deve ser programado para agir em caso de acidente inevitável? Podemos imaginar, por exemplo, uma situação em que um grupo inesperado de transeuntes surge subitamente na rota do veículo, de tal maneira que, inexistindo a possibilidade de parar a máquina e evitar um acidente – devido a fatores como alta velocidade - esta deva fazer uma escolha: atropelar o grupo ou desviar da rota, podendo colidir com uma mureta ou outros veículos? Deveria, o carro, minimizar a perda de vidas, mesmo que isso implique em sacrificar os ocupantes do automóvel, ou deveria proteger os ocupantes a todo custo?

Problema semelhante, corriqueiramente utilizado em experimentos para aferição de teorias éticas, é o famoso *trolley problem*, ou dilema do bonde. Em uma de suas versões, exemplificada por Judith Jarvis Thomson (1985), um bonde desgovernado segue em direção a cinco pessoas. Certo sujeito, posicionado em uma ponte sobre a estrada, poderia parar o bonde atirando um peso na frente dele. Ao lado do indivíduo está apenas um homem muito gordo, de tal modo que a única maneira de parar o veículo seria empurrando este homem da ponte para a pista, matando-o para salvar outras cinco vidas. Qual atitude deveria tomar o sujeito¹⁴?

O ponto onde pretende-se chegar com essas divagações, é: dilemas éticos que nós - seres humanos - enfrentamos hodiernamente já fazem parte da tarefa dos engenheiros ao programarem veículos autônomos, o que torna essencial e urgente pensar em um norte para a “moralidade algorítmica” e a criação de um código de conduta específico para as máquinas dotadas de AI.

Bonnefon, Shariff e Rahwan (2016) contribuíram formidavelmente para essa questão com um estudo denominado “O dilema social dos veículos autônomos”, onde apontaram que a automação completa nos veículos trará à tona a necessidade destes em escolherem, às vezes, entre dois males, como o de atropelar pedestres ou sacrificar a si mesmo e seus passageiros para salvá-los.

¹⁴ Exemplos semelhantes também são utilizados por Michael Sandel (2012), na obra “Justiça, o que é fazer a coisa certa?”.

Segundo os autores do artigo, definir os algoritmos que estabelecerão os parâmetros para a tomada de decisões dos veículos autônomos é um imenso desafio, justamente por envolver, em certos casos, dilemas morais.

Em seus estudos, descobriram que os participantes da pesquisa desenvolvida por eles, em sua maioria, aprovam veículos autônomos utilitaristas, que são aqueles que sacrificam seus passageiros, em menor número, para o bem maior em caso de escolhas morais difíceis¹⁵.

Tal tipo de escolha seria pautada em uma das principais características da teoria moral utilitarista, o Princípio da agregação. Segundo este princípio, deve-se levar em consideração - no cálculo moral em busca da maximização da felicidade dos sujeitos - o saldo líquido do bem-estar de todos os indivíduos afetados pela ação, independentemente da distribuição deste saldo. O que conta é a quantidade global de bem-estar produzida, qualquer que seja a repartição desta quantidade. Nesta linha de raciocínio, considera-se válido "sacrificar uma minoria", cujo bem-estar será diminuído, a fim de aumentar o bem-estar geral.

A possibilidade de sacrifício se baseia na ideia de compensação: a infelicidade ou desprazer de uns é compensada pelo bem-estar dos outros. Se o saldo de compensação for positivo, a ação é julgada moralmente boa.

Segundo Hirose (2014), com base nos "princípios morais interpessoalmente agregados", os ganhos para um grupo de indivíduos podem superar moralmente as perdas para um grupo diferente de indivíduos. O princípio da agregação, denominada por Hirose de "agregação interpessoal", deve ser usada quando existir um conflito de interesses entre diferentes grupos de indivíduos. Aqueles que se opõem à agregação

¹⁵ Veículos utilitários ou utilitaristas seriam aqueles que sacrificarem a própria existência e a vida de seus passageiros para minimizar os danos maiores, como a morte de vários pedestres. De certa forma, seriam os carros que se pautam, em suas escolhas, pela teoria ética do utilitarismo. Conforme Durozoi (1993), ao contrário da concepção kantiana - segundo a qual o valor moral da ação não se organiza a partir de seus resultados, mas a partir da intenção que anima - o utilitarismo, inscrevendo-se numa tradição que remonta a Hobbes e que se inspira em Epicuro, é uma doutrina que coloca a utilidade como critério ou princípio da atividade do ponto de vista moral: a moral utilitarista é a teoria racional que permite determinar as técnicas que garantem o máximo da felicidade individual. Trata-se, em primeiro lugar, da doutrina de Bentham, que se reduz quanto ao essencial a uma aritmética dos prazeres, ou seja, a um cálculo egoísta da maior quantidade possível de felicidade individual, o que leva aliás - por um cálculo inteligente - a visar o máximo de felicidade para o maior número. O utilitarismo altruísta de Stuart Mill leva ademais em consideração a qualidade dos prazeres e conclui que o indivíduo por interesse deve finalmente querer a felicidade de todos.

interpessoal sustentam que os ganhos combinados para um grupo não superam moralmente as perdas combinadas a outro grupo ou indivíduo.

Essa maioria, destacada na pesquisa de Bonnefon, Shariff e Rahwan (2016), gostaria que as pessoas, em geral, comprassem carros deste tipo - que sacrificam sua existência, assim como seus passageiros - em prol da maioria. Contudo, eles próprios admitem que prefeririam comprar os carros autônomos que protejam seus passageiros a todo custo. Desaprovam, portanto, a imposição pelo mercado e fabricantes de carros autônomos utilitaristas e alegam que estariam menos dispostos a comprar tais veículos programados visando as escolhas morais utilitaristas.

Consequentemente, se tão somente a opinião pública for levada em consideração na fabricação e programação de carros autônomos, sem regulamentação ética e jurídica específica, os algoritmos implementados nos carros autônomos que estão por vir podem aumentar paradoxalmente a perda de vidas, adiando a adoção de uma tecnologia mais segura (BONNEFON; SHARIFF; RAHWAN, 2016).

Além disso, a autonomia dos robôs dotados de AI- consistente na capacidade tomar decisões e de aplicá-las no mundo exterior, independentemente do controle ou da influência externa – suscita uma discussão problemática: a da sua natureza frente às categorias jurídicas existentes. Seriam, os robôs dotados de AI, pessoas singulares, pessoas coletivas, animais, objetos ou deve-se estabelecer uma nova categoria com características e implicações próprias no que diz respeito à atribuição de direitos e deveres, incluindo a responsabilidade por danos? Este é mais um ponto que elucida a urgência de regulamentação em estatuto próprio.

Certamente, frente ao cenário real e cada vez mais próximo no qual robôs podem tomar decisões autônomas, as normas tradicionais não serão suficientes para acionar a responsabilidade de um robô, vez que ainda não é possível identificar juridicamente, de forma exata, a parte responsável para prestar a indenização e para exigir a essa parte reparação pelos danos causados.

Na fase de programação ou criação da inteligência artificial, o Parlamento Europeu (2016, p. 4), argumenta em prol do estabelecimento de padrões ou princípios gerais relativos ao desenvolvimento da robótica e da inteligência artificial para utilização civil. Sugere que certos princípios gerais sejam seguidos, a começar pelas

Leis de Asimov¹⁶, segundo as quais: (1) - Um robô não pode magoar um ser humano ou, por inação, permitir que tal aconteça; (2) - Um robô tem de obedecer às ordens dos seres humanos, exceto quando tais ordens entrarem em conflito com a primeira lei; (3) - Um robô tem de proteger a sua própria existência desde que tal proteção não entre em conflito com a primeira ou com a segunda lei¹⁷ (ASIMOV, 2008).

Além disso, o Parlamento (2016, p. 7) propõe a criação de um código de conduta, o qual estabelecerá padrões por meio de um quadro ético orientador baseados em princípios como: (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; (7) a não estigmatização; (8) a autonomia; (9) a responsabilidade individual; (10) o consentimento esclarecido; (11) a privacidade e (12) a responsabilidade social.

Segundo o Parlamento (2016, p. 15), o código de conduta ética proposto no domínio da robótica deve estabelecer as bases para a identificação, supervisão e cumprimento dos princípios éticos fundamentais desde as fases de concepção e desenvolvimento dos robôs dotados de AI.

Este código de conduta deve ser concebido de modo ponderado, a fim de possibilitar ajustamentos individuais numa base casuística para que seja possível avaliar se um dado comportamento é certo ou errado diante de certa situação, bem como a tomada de decisões de acordo com uma hierarquia predefinida de valores.

O código não deve substituir a necessidade de resolver todos os grandes problemas jurídicos neste domínio, devendo, porém, ter uma função de complementaridade. Deverá facilitar a classificação ética da robótica, reforçar os esforços de inovação responsáveis neste domínio e atender às preocupações da população. Daí, portanto, a necessidade de complementação de regulamentação na seara jurídica.

¹⁶ Isaac Asimov, escritor e bioquímico americano nascido na Rússia.

¹⁷ Mais tarde, Asimov teria criado uma quarta regra, denominada “regra zero”: um robô não pode magoar a humanidade ou, por inação, permitir que tal aconteça.

Em especial, tal código de conduta deve dar atenção especial às fases de investigação e desenvolvimento da pertinente trajetória tecnológica (processo de concepção, exame ético, controles de auditoria, etc.). Deve conceder instrumentos para resolução dos pertinentes dilemas éticos e permitir que estes sistemas inteligentes funcionem de um modo eticamente responsável.

No que tange à proteção aos direitos humanos e fundamentais, o Parlamento Europeu (2016, p. 16) sugere que o código de conduta estabeleça que as atividades de desenvolvimento da robótica devam respeitar integralmente tais direitos, sendo que estes devem ser considerados em todos os momentos, em prol do interesse e do bem-estar dos indivíduos e da sociedade, seja na fase concepção dos robôs ou em sua utilização. A dignidade humana – quer na sua dimensão física, quer na sua faceta psíquica – necessariamente deve ser sempre levada em consideração.

A responsabilização também deve ser algo previsto no código, de tal forma que este torne os engenheiros de robótica responsáveis pelos impactos sociais, ambientais e na saúde humana que a robótica possa ter nas gerações presentes ou futuras.

Em relação à segurança e à saúde, o estatuto deve estabelecer que criadores de robótica persigam preceitos como: o bem-estar físico, a segurança e a saúde das pessoas. Nesse sentido, um engenheiro de robótica tem de considerar, na programação, a preservação do bem-estar humano, de tal forma que o robô dotado de AI consiga revelar prontamente fatores que possam colocar em perigo as pessoas ou o ambiente.

A reversibilidade também seria condição necessária, sendo esta um conceito fundamental na programação dos robôs para que estes tenham um comportamento seguro e confiável. Um modelo de reversibilidade indica ao robô quais as ações que são reversíveis e como as reverter caso sejam reversíveis. A capacidade de anular a última ação ou uma sequência de ações permite aos utilizadores anularem ações indesejadas e regressarem à boa fase do seu trabalho ou *status quo ante*.

Outro ponto importante a ser estabelecido pelo código de conduta liga-se à privacidade. O direito à privacidade tem de ser sempre protegido, sendo que um engenheiro de robótica deve garantir que as informações privadas serão mantidas em segurança e utilizadas apenas de forma adequada, respeitando os limites da intimidade. O Parlamento (2016, p. 17) também sugere que os engenheiros devam ser vinculados

pelo código a cumprir qualquer ordem judicial para a destruição de quaisquer dados armazenados pela inteligência artificial.

Ademais, o Parlamento Europeu (2016, p. 17) sugere que um engenheiro, ao programar a inteligência artificial, deve garantir que os indivíduos não possam ser identificados em termos pessoais, salvo em caso de circunstâncias excepcionais por meio de consentimento esclarecido e informado de forma inequívoca. O consentimento humano informado deve ser solicitado e obtido antes de qualquer interação homem-máquina. Nessa medida, os criadores de robótica são responsáveis por desenvolver e seguir procedimentos para o consentimento válido, para a confidencialidade, o anonimato, o tratamento justo e o processo equitativo.

Percebe-se, portanto, que o Parlamento Europeu já se despertou para a relevância e urgência do debate, estabelecendo diretrizes, ainda que gerais, sobre a temática.

No Brasil, o debate ainda é imaturo ou inexistente. Conseqüentemente, nossa legislação interna não possui aparatos para lidar com a complexidade das demandas que começarão a existir em um futuro muito próximo. Um dos intuitos deste artigo é justamente despertar o interesse dos pesquisadores e o legislativo pátrio para a questão, para que, em conjunto, pensem em soluções viáveis para as possíveis demandas.

Assim como o Parlamento Europeu, acreditamos que a regulamentação deva acontecer no ordenamento jurídico interno em, pelo menos, dois níveis: (1) no momento da criação e desenvolvimento da AI por engenheiros, os quais devem seguir uma “moralidade algorítmica”. Nesse caso, um estatuto próprio seria de grande valia para a criação de parâmetros morais a serem seguidos, assim como para a delimitação da natureza jurídica da AI; (2) no que diz respeito à responsabilização jurídica dos robôs dotados de AI, assim como de seus criadores, incluindo engenheiros e empresas que financiam o desenvolvimento. Nesse segundo nível, deve-se discutir a viabilidade de alteração das legislações já existentes – na esfera cível em matéria de responsabilidade, por exemplo – com o incremento da discussão, ou então a criação de legislação específica para a temática.

CONCLUSÃO

A preocupação de grandes intelectuais da tecnologia, assim como de estudiosos e acadêmicos sobre o tema, revela a premente necessidade de incluir o desenvolvimento da inteligência artificial no debate ético e jurídico.

Durante o desenvolvimento deste artigo, tentamos demonstrar que a AI pode ser uma ameaça à humanidade, caso não existam mecanismos de controle que garantam um desenvolvimento saudável da tecnologia. Isso porque, da mesma forma que a AI tem trazido - e ainda trará - inúmeros benefícios para a humanidade, ela também pode ser um fator de risco.

Por isso, defendemos neste trabalho a necessidade da criação de padrões éticos que necessariamente devem ser seguidos por empresas e engenheiros, quando da programação e desenvolvimento de robôs dotados de AI, assim como sugerido pelo Parlamento Europeu em uma moção no ano de 2016.

Nesse sentido, a criação de um “código de condutas” para o desenvolvimento da inteligência artificial é uma medida a ser providenciada com urgência, no intuito de se atender uma “moralidade algorítmica” no desenvolvimento da AI. Com o estabelecimento de padrões éticos, pode-se garantir que a AI não desenvolva comportamentos inesperados e que ameacem, causem mal ou ponham em risco a humanidade, permitindo, portanto, que estes sistemas inteligentes funcionem de um modo eticamente responsável.

Além disso, é elementar e urgente a criação de estatutos jurídicos que possibilitem a identificação exata da natureza jurídica da AI frente às categorias jurídicas existentes. Tal regulamentação jurídica também deve disciplinar matérias de responsabilidade voltadas exclusivamente para à AI, de tal de tal forma que consiga enquadrar, com precisão, condutas dos engenheiros de robótica e das empresas envolvidas no desenvolvimento da tecnologia, os quais devem ser responsáveis pelos impactos sociais e ambientais, além dos relacionados à saúde, bem-estar e existência humana, que a robótica possa ter nas gerações presentes ou futuras.

Certamente, se nos atentarmos para a relevância do assunto e tomarmos os cuidados necessários para o desenvolvimento prudente da tecnologia, será possível que desfrutemos com tranquilidade dos benefícios e facilidades advindos da mesma.

REFERÊNCIAS

AFFAIRS, European Parliament: Committee On Legal. **Draft Report**: with recommendations to the Commission on Civil Law Rules on Robotics. 2016. Disponível em: <<http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+COMPARL+PE-582.443+01+DOC+PDF+V0//EN>>. Acesso em: 27 fev. 2017.

ASIMOV, Isaac. **I, Robot**. New York City: Bantam Spectra, 2008. 256 p.

ASARO, Peter. On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. **International Review Of The Red Cross**, [s.l.], v. 94, n. 886, p.687-709, jun. 2012. Cambridge University Press (CUP). <http://dx.doi.org/10.1017/s1816383112000768>.

BELLMAN, Richard. **Artificial Intelligence: Can Computers Think?**. Boston: Thomson Course Technology, 1978. 146 p.

BLACK Mirror. Direção de Charlie Brooker. Produção de Zeppotron e HouseofTomorrow. 2011-presente.

BONNEFON, J.-f.; SHARIFF, A.; RAHWAN, I.. The social dilemma of autonomous vehicles. **Science**, [s.l.], v. 352, n. 6293, p.1573-1576, 23 jun. 2016. American Association for the Advancement of Science (AAAS). <http://dx.doi.org/10.1126/science.aaf2654>.

BRINGSJORD, Selmer et al. Real robots that pass human tests of self-consciousness. **2015 24th Ieee International Symposium On Robot And Human Interactive Communication (ro-man)**, [s.l.], p.1-7, ago. 2015. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/roman.2015.7333698>.

CAPEK, Karel. **R.U.R.: Rosumo vi Umeli Roboti**. [s.i]: Jiahu Books, 2014. 82 p.

DUROZOI, Gérard; ROUSSEL, André. *Dicionário de Filosofia*. Campinas: Papirus, 1993. 511p. Tradução de: Marina Appenzeller.

FRANKLIN, Stan; GRAESSER, Art. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. In: Intelligent agents III agent theories, architectures, and languages. Springer Berlin Heidelberg, 1997. p. 21-35.

GIBBS, Samuel. **Elon Musk: artificial intelligence is our biggest existential threat**. 2014. Disponível em: <<https://www.theguardian.com/technology/2014/oct/27/elon-musk-artificial-intelligence-ai-biggest-existential-threat>>. Acesso em: 12 mar. 2017.

GRIFFIN, Andrew. **Stephen Hawking: Artificial intelligence could wipe out humanity when it gets too clever as humans will be like ants:** AI is likely to be 'either the best or worst thing ever to happen to humanity,' Hawking said, 'so there's huge value in getting it right'. 2015. Disponível em: <<http://www.independent.co.uk/life-style/gadgets-and-tech/news/stephen-hawking-artificial-intelligence-could-wipe-out-humanity-when-it-gets-too-clever-as-humans-a6686496.html>>. Acesso em: 12 mar. 2017.

HAWKING, Stephen; RUSSELL, Stuart; WILCZEK, Max Tegmark; Frank. **Transcendence looks at the implications of artificial intelligence - but are we taking AI seriously enough?** 2014. Disponível em: <<http://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-9313474.html>>. Acesso em: 26 fev. 2017.

HAUGELAND, John. **Artificial intelligence:** The very idea. Cambridge: Mit, 1985. 287 p.

HIROSE, Iwao. Why Aggregation? **Moral Aggregation**, [s.l.], p.3-18, 4 nov. 2014. Oxford University Press (OUP). <http://dx.doi.org/10.1093/acprof:oso/9780199933686.003.0001>.

KURZWEIL, Ray. **The Age of Intelligent Machines.** Cambridge: Mit Press, 1990. 580 p.

KOROSEC, Kirsten. **Volvo CEO:** We will accept all liability when our cars are in autonomous mode. 2015. Disponível em: <<http://fortune.com/2015/10/07/volvo-liability-self-driving-cars/>>. Acesso em: 03 mar. 2017.

MANYIKA, James et al. **A future that works:** automation, employment and productivity. 2017. Disponível em: <<http://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works>>. Acesso em: 04 mar. 2017.

METROPOLIS. Direção de Fritz Lang. Roteiro: Thea von Harbou. [s.i]: Universum Film (ufa), 1927. P&B.

OXFORD. **English Oxford Living Dictionaries.** Disponível em: <<https://en.oxforddictionaries.com>>. Acesso em: 22 mar. 2017.

PASCHELES, Wolf. **Sippurim.** [s.i]: Nabu Press, 2012. 374 p.

REIS, Luís Paulo. **Coordenação em Sistemas Multi-Agente:** Aplicações na Gestão Universitária e Futebol Robótico. 2003. 451 f. Tese (Doutorado) - Curso de Engenharia, Universidade do Porto, Porto, 2003. Cap. 10.

ROBOCOP. Direção de Paul Verhoeven. Produção de Arne Schmidt. Roteiro: Edward Neumeier; Michael Miner. Música: Basil Poledouris. 1987.

RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence: A Modern Approach**. 3. ed. New York City: Pearson, 2009. 1152 p.

SHELLEY, Mary W. **Frankenstein**; trad. de Éverton Ralph, 5ª ed. Rio de Janeiro: Ediouro, 1996.

SANDEL, Michael. **Justiça: o Que É Fazer a Coisa Certa**. 6. ed. Rio de Janeiro: Civilização Brasileira, 2012. 349 p.

SULLEYMAN, Aatif. **Elon Musk: Humans must become cyborgs to avoid AI domination**: 'It's mostly about the bandwidth, the speed of the connection between your brain and the digital version of yourself'. 2017. Disponível em: <<http://www.independent.co.uk/life-style/gadgets-and-tech/news/elon-musk-humans-cyborgs-ai-domination-robots-artificial-intelligence-ex-machina-a7581036.html>>. Acesso em: 12 mar. 2017.

STAR Wars: Episode IV – A New Hope. Direção de George Lucas. Roteiro: George Lucas. Música: John Williams. 1977.

TEAM, The Tesla. **A Tragic Loss**. 2016. Disponível em: <https://www.tesla.com/pt_PT/blog/tragic-loss?redirect=no>. Acesso em: 03 mar. 2017.

TERMINATOR 2: Judgment Day. Direção de James Cameron. Produção de James Cameron; Stephanie Austin; B.j. Rack; Gale Anne Hurd; Mario Kassar. Roteiro: James Cameron; William Wisher Jr. Música: Brad Fiedel. 1991.

THOMSON, Judith Jarvis. The Trolley Problem. **The Yale Law Journal**, [s.l.], v. 94, n. 6, p.1395-1415, maio 1985. JSTOR. <http://dx.doi.org/10.2307/796133>.

THRUN, Sebastian. Toward robotic cars. **Communications Of The Acm**, [s.l.], v. 53, n. 4, p.99-106, 1 abr. 2010. Association for Computing Machinery (ACM). <http://dx.doi.org/10.1145/1721654.1721679>.

TRANSPORTATION, U.s Department Of; ADMINISTRATION, National Highway Traffic Safety. **ODI Resume**. 2016. Disponível em: <<https://static.nhtsa.gov/odi/inv/2016/INCLA-PE16007-7876.PDF>>. Acesso em: 03 mar. 2017.

WESTWORLD. Produção de Jonathan Nolan. Roteiro: Athena Wickham. 2016-presente.

WHIPPLE, Tom. **Stephen Hawking on humanity (and Jeremy Corbyn): The physicist is hopeful for the future - as long as the Labour leader goes**. 2017. Disponível em: <<http://www.thetimes.co.uk/edition/news/hawking-on-humanity-and-corbyn-jk88zx0w2>>. Acesso em: 12 mar. 2017.

WOOLASTON, Victoria. **Will robots make us their PETS? Apple founder Steve Wozniak has no doubt artificial intelligence will take over the world.** 2015. Disponível em: <<http://www.dailymail.co.uk/sciencetech/article-3007920/Will-robots-make-PETS-Apple-founder-Steve-Wozniak-no-doubt-artificial-intelligence-world.html>>. Acesso em: 23 mar. 2017.

Recebido em: 08 de abril de 2017

Aprovado em: 20 de abril de 2017