

Problemas ético-jurídicos concernentes aos danos causados por robôs autônomos na saúde: algumas definições conceituais

Ethical and legal problems about damages caused by autonomous robots on healthcare: some conceptual definitions

Problemas ético-legales sobre daños causados por robots autónomos en la salud: algunas definiciones conceptuales

Ana Elisabete Farinha Ferreira Dias Pereira¹

Resumo

Objetivo: subsidiar a compreensão filosófica-especulativa dos conceitos de *dano* e *autonomia* no amplo e profícuo pensamento publicado sobre robôs. **Metodologia:** foi realizada uma análise da literatura mais relevante e uma definição conceptual prévia, para depois argumentar e definir noções fundamentais sobre o problema da programação ética de forma integrada. **Resultados:** *dano* e *autonomia* são conceitos acriticamente e recorrentemente utilizados na literatura, porém, o convívio com robôs autônomos exige uma reflexão que tome em conta a agentividade ética de robôs em contraponto com humanos. Foram identificados três princípios como essenciais na procura de um consenso conceptual: o respeito pela recusa em ser tratado por um robô; a proteção da liberdade humana face a um robô; e a manutenção do tecido social. **Conclusão:** a programação ética de robôs autônomos, longe de ser um exercício utilitarista abstrato, exige um consenso transdisciplinar, ético-jurídico, que permita abranger diferentes categorias de dano e de autonomia.

Palavras-chave: Autonomia. Dano ao paciente. Direito da saúde. Ética baseada em princípios. Robótica.

Abstract

Objective: this article presents two of the most prominent normative problems on damage caused by autonomous robots in healthcare: what is considered *autonomy* when referring to the action of a robot, and what is considered *damage* in the activity of an autonomous robot. The aim is to contribute to the philosophical speculative understanding of the concepts of *damage* and *autonomy* in the broad critical thought about robots. **Methodology:** from an analysis of the most relevant literature and a prior conceptual definition, we discussed and defined fundamental notions about the problem of ethical programming in an integrated (ethical and legal) manner. **Results:** *damage* and *autonomy* are concepts frequently and uncritically used in the literature. However, living with autonomous robots requires reflection that considers the ethical agency of robots as opposed to humans. We have identified three principles as essential in the search for a conceptual consensus: respect for the refusal to be treated by a robot; the protection of human freedom in the face of a robot; and the maintenance of the social bond. **Conclusion:** the ethical programming of autonomous robots, far from being an abstract utilitarian exercise, requires a transdisciplinary, ethical-legal consensus that allows different categories of damage and autonomy to be covered.

Keywords: Autonomy. Harm to the patient. Health law. Ethics based on principles. Robotics.

¹ Doutora em Bioética pela Universidade Católica Portuguesa, Porto, Portugal. Professora adjunta convidada no Instituto Politécnico de Leiria, Leiria, Portugal. <https://orcid.org/0000-0002-3845-3166>. E-mail: anaelisferreira@hotmail.com

Resumen

Objetivo: subsidiar la comprensión filosófico-especulativa de los conceptos de daño y autonomía en el pensamiento amplio y fructífero publicado sobre robots. **Metodología:** se realizó un análisis de la literatura más relevante y una definición conceptual previa, para posteriormente argumentar y definir nociones fundamentales sobre el problema de la programación ética de manera integrada. **Resultados:** el daño y la autonomía son conceptos acrílicos y de uso frecuente en la literatura, sin embargo, convivir con robots autónomos requiere una reflexión que tenga en cuenta la agencia ética de los robots en contraste con los humanos. Se identificaron tres principios como fundamentales en la búsqueda de un consenso conceptual: el respeto a la negativa a ser tratado por un robot; la protección de la libertad humana frente a un robot; y el mantenimiento del tejido social. **Conclusión:** la programación ética de robots autónomos, lejos de ser un ejercicio utilitario abstracto, requiere de un consenso ético-legal transdisciplinario que permita cubrir diferentes categorías de daño y autonomía.

Palabras clave: Autonomía. Daño al paciente. Derecho sanitario. Ética basada en principios. Robótica.

Introdução

Em uma situação discursiva ideal, as relações potencialmente fraturantes do tecido social seriam sempre refletidas e aturadas no plano ético, antes da procura de uma resposta especificamente jurídica. O que permitiria pensar, de forma estruturada e sistematizada, que problemas representam as relações com robôs; qual o seu alcance e a probabilidade efetiva de virem a verificar-se; quais deles apresentam relevância jurídica; e se o ponto-de-fuga antropocêntrico, que subsidia uma reflexão baseada nas diferenças, é o pressuposto adequado a uma resposta axiologicamente comprometida e, ao mesmo tempo, socialmente responsável (1).

Lateralmente, surgiriam aqui noções e princípios sobre o *bom* e o *bem* nas relações humanas mediadas, a relação entre a normatividade e a ciência e as suas exigências de coerência, senão mesmo de congruência (2), e finalmente, a opção por um balizamento consequencialista, puramente deontológico, ou principialista.

Tal não acontece sempre, porque a intervenção do Direito é, sobretudo, uma intervenção responsiva, que se traduz depois em certo conforto intelectual que não deixa de subsidiar alguma letargia ética que compromete a problematização crítica das exigências em causa.

Como, porém, não concebemos que o ponto de partida de qualquer construção normativa possa assentar em apartamento cognitivo entre ética e direito (3), cremos que a reflexão vale sempre a pena, ainda que funcionalmente extemporânea e sem qualquer

preocupação com a utilidade, porque a pergunta pela utilidade no pensamento valorativo é, fundamentalmente, uma pergunta errada.

É verdade que estamos perante um tema relativamente ao qual já muito foi dito. As relações entre a moral e o direito são, dizia Reale (4), a questão mais ampla, mais densa e mais complexa da teoria do direito. E o modo como olhamos para essa relação afeta decisivamente as respostas normativas: transdisciplinaridade é muito diferente de interdisciplinaridade. Um pensamento ético-jurídico é aquele que admite que, por exemplo, a autonomia de que fala a moral e a autonomia de que fala o direito devem ser a mesma coisa, que a conceptualização do prejuízo pessoal pode ser materialmente idêntica na ética e no direito; o que equivale a afirmar que o direito possa ter também uma dimensão interna, e que a ética possa ser, pelo menos intencionalmente, empírica ou coerciva.

Diante desse contexto, o artigo tem por objetivo a compreensão filosófico-especulativa dos conceitos de *autonomia* e *dano* no amplo e profícuo pensamento publicado sobre robôs.

Metodologia

Para alcançar o desiderato, efetuámos uma revisão crítica das fontes contemporâneas mais relevantes, observando diferentes contextos e horizontes de sentido em que os referidos conceitos são colocados. Posteriormente, e constatando que a problemática da convivência com robôs nos convoca a pensar diversas categorias especulativas de autonomia e de dano, concluímos pela necessidade de um encontro transdisciplinar, que se fará em respeito pelos mais relevantes princípios que devem reger a relação entre humanos e robôs.

Resultados e discussão

O primeiro problema ético-jurídico fundamental: o dano no âmbito da atuação de um robô

O dano enquanto prejuízo ou ofensa sofridos por um sujeito tem manifestações evidentes e facilmente discerníveis; tem, porém, formas de manifestação menos óbvias que nem por isso deixam de merecer uma atenção normativa bastante exigente. É evidente o dano da privacidade que pode resultar imediatamente da mera presença de um robô. É também evidente o dano potencial para a integridade física das pessoas, o dano corporal propriamente dito que o contacto com um robô pode provocar, e todas as ofensas

decorrentes de uma violação daquela integridade, como a perda estética, a perda de qualidade de vida, as dores e os transtornos da vida pessoal e da relação.

Mas o convívio com máquinas autônomas convoca-nos a uma outra categoria especulativa do dano, especificamente ao dano da dignidade, tal qual como coloca Rodotà (5): i) o dano da renúncia à presença do outro-pessoa; ii) os danos da expropriação, da agressão, da dispersão, da submissão, da aceitação acrítica, da decisão arbitrária; iii) o dano da ingratidão.

Quererá o direito ter resposta para o que aqui está em causa? E se sim, quererá o direito estabelecer uma linha limítrofe de intervenção, espaços de direito e de não direito, que salvaguardem os efeitos nefastos de uma *macro-juridificação* (6) do corpo e da relação? Neste ponto, lembremos especificamente que as relações com robôs – muitas delas – se encontram em espaços que habitualmente pensamos em termos de *interpessoalidade*, de vínculo entre duas pessoas, ligações que se encontram, em grande extensão, fora do âmbito típico de intervenção do direito, como é o caso das relações de cuidado, afetivas e sexuais.

Da leitura pregressa de Rodotà (5) são-nos sugeridas algumas breves reflexões: a renúncia à presença do outro-pessoa é hoje uma realidade largamente sedimentada: o outro foi substituído por formas de manifestação mediatas dessoutro, como aquelas que se apreendem por meio de aparelhos virtuais e audiovisuais. Já não é o outro, mas apenas a sua manifestação – incompleta, vazia de corpo e mediata – aquilo que se nos apresenta como interlocutor.

Contudo, não há razões para supor que, à medida que os robôs proliferem, as relações humanas sejam mais afetadas do que já são, efetivamente. A preocupação com a deterioração do tecido social humano deve afligir-nos, independentemente do avanço e progresso das máquinas (7). Há muitas perspectivas sobre como a tecnologia afeta – positiva e negativamente – as relações humanas, mas há, pelo menos, dois pontos de consenso: a tecnologia deve ser usada em benefício das relações humanas; as relações de cuidado mútuo entre humanos são infungíveis na sua plenitude.

Os danos (potenciais) da expropriação, da agressão, da dispersão, da submissão, da aceitação acrítica, e da decisão arbitrária, são os mais evidentes no uso de tecnologia avançada: a fronteira entre o que é nosso e o que não é (com particular destaque, a informação pessoal e a imagem); a dificuldade de relacionamento com robôs; a dispersão social dos humanos e mesmo a dispersão íntima e a constante sensação de estar fora de

um sistema humano; a submissão aos desígnios científicos e a sua aceitação acrítica, operando por uma suposta autoridade científica que nos é transmitida como um definitivo e alimentada pelo marketing, que a todo o momento nos *revela* o que é necessário ou útil nas nossas vidas; e, finalmente, a decisão arbitrária – a nossa, quando referida aos robôs, e a dos robôs, quando referida aos humanos, que poderá não ser perceptível, rastreável ou interpretável.

Finalmente, podemos falar num dano da ingratidão. É sobejamente conhecida a linha que tem separado as Humanidades e as Ciências e, dentro daquelas e em particular, a linha entre a Religião e o progresso tecnológico. O intuito dissipatório grave e gratuito da espiritualidade humana, por vezes promovido pelo pensamento tecnológico, onde inexistente lugar de pensamento para a metafísica e para o transcendente, tende a ter como resposta uma certa repugnância face ao progresso e uma visão do mesmo como amoral ou mesmo imoral. Destarte, fala-se em ingratidão face ao Criador, à Natureza, à Biologia, à Evolução, aos nossos antepassados, etc. Contudo, há uma visão diametralmente oposta dessa fenda bioética, que se traduz em uma questão concreta: recusar o progresso não revela também, precisamente, ingratidão à Ciência e ao próprio homem que a produz? O ser humano cria porque tem a capacidade de criar, e inova porque é capaz de o fazer. Negar o progresso é coartar o homem na sua plenitude, e negar o valor da ciência e da investigação, que são hoje quase unanimemente considerados *bens* juridicamente relevantes em si mesmos.

Podemos perguntar se é lícito ou moralmente adequado alterar características das pessoas e das suas relações intersubjetivas por meio de tecnologia (genética, robótica, farmacológica); mas também devemos, ao mesmo passo, perguntar se é lícito ou moralmente adequado não o fazer quando podemos fazê-lo, no sentido da beneficência.

Que dizer, depois, dos danos inconhecíveis? Estarão, pela sua própria natureza, excluídos da resposta normativa? Hogg (8), designadamente, entende que não, e que os riscos não conhecidos merecem integrar-se na resposta da responsabilidade civil, ainda que tal concorra para uma total diluição do paradigma da culpa.

O modo como o direito *performa* a dignidade, não como uma supernorma acima de todas as outras, mas como uma agregação de princípios normativos, diz-nos Rodotà (5), fez do direito uma antropologia, no seio do qual a dignidade convoca uma memória coletiva e uma matriz cultural ímpar, translucidamente simplificada na formulação da Declaração da ONU, segundo a qual “o reconhecimento da dignidade inerente a todos os membros da

família humana e seus direitos, iguais e inalienáveis, são o fundamento da liberdade, da justiça e da paz no mundo” (9).

O drama é que essa dignidade como antropologia encontra sempre novas formas de ser violada.

Princípios mediadores nas relações com robôs

Na Resolução do Parlamento Europeu, de 16 de fevereiro de 2017 (10), que contém recomendações à Comissão sobre disposições de Direito Civil sobre Robótica (2015/2103(INL)) estão vertidos alguns dos princípios que resultaram do denso trabalho de prospeção e supervisão da Comissão de Liberdades Cívicas, Justiça e Assuntos Internos nesta matéria, dos quais merecem ser destacados quando convocamos os conceitos de autonomia e de dano.

O princípio do respeito pela recusa em ser tratado por um robô

Princípio que se insere sem esforço na linha autonomista dominante, de acordo com a qual é garantido ao paciente a liberdade de escolher os cuidados e tratamentos médico-cirúrgicos que lhe sejam dirigidos, independentemente da razoabilidade da recusa e enquanto manifestação de uma vontade que se autovalida de *per si*, quase sempre sem integração hermenêutica do risco, da segurança, da adequação ou da proporcionalidade.

É a garantia de liberdade que vai implicar uma densidade acrescida ao nível da informação prestada e do consentimento, e do tempo próprio em que deve ser recolhido; a mexer com a autonomia prospetiva, por um lado, e com a livre gestão e administração de serviços de saúde, por outro.

Há, contudo, que ter em conta que a possibilidade de recusa em ser tratado por um robô não é apenas uma manifestação do princípio da autonomia. Ela expressa, ainda que sub-repticiamente, a hipótese estética do *uncanny valley*, que a ética e o direito não têm combatido: numa palavra, a ideia de que as relações entre humanos e robôs são necessariamente mediadas pela repulsa, que se acentua à medida que as semelhanças se evidenciam, o que pode e deve ser desconstruído (11).

O princípio da proteção da liberdade humana face a um robô

É outro princípio que se insere bem na linha autonomista e no respeito pela liberdade fundamental, A saúde é uma área em que a liberdade é fortemente constringida pelos

fármacos; sedações; analgesias; imobilizações; pela própria situação de internamento, ainda que voluntário; pelas restrições físicas da mobilidade que são por vezes necessárias e que se traduzem, na prática, numa situação de prisão.

Onde reside exatamente a diferença consoante essas práticas sejam levadas a cabo por seres humanos ou por robôs? Fica diminuída a nossa dignidade ao sermos coartados por um robô, mais do que se o fôssemos por um ser humano?

O princípio da manutenção do tecido social

Pelo menos desde a invenção da máquina a vapor que a colonização do mundo por máquinas se apresenta como a maior das ameaças à humanidade do homem e ao tecido social criado por relações de comoção, quentes e emocionais, como são por natureza as nossas (12). O perigo de dilaceração do tecido social baseado no calor humano, *lato sensu*, tem criado em diversas esferas um medo denso que, todavia, vem sucessivamente negligenciando a observação empírica e a probabilidade.

A questão metodológica que nos parece mais evidente é a de que esse suposto princípio desafia a própria noção de princípio. É que nos princípios normativos, o vocábulo *normativos* não é um adjetivo. O laço social é, por certo, um bem normativamente relevante, mas em um contexto de intersubjetividade fortemente marcado pela cultura do indivíduo, que na contemporaneidade não pode deixar de ver-se como parte da ideação de como as coisas poderiam ser de outra maneira. O espectro dessa ideação desafia, como bem notou Lacan, o mito de totem e tabu (13), subvertendo o proibido e o perigo e sobrevalorizando a novidade, a expectativa, o nunca antes. Pode bem dizer-se que o homem é hoje, mais do que antes, um caminhante que nunca chega, que – como nota E. O. Wilson (14) – jamais alcançará a paz social e a abundância de recursos, porque jamais se contentará em viver como uma formiga. Se a afinidade cultural, as relações de afeto e os laços de responsabilidade são certamente fundamento do laço social, e boas razões para o preservar, não pode dizer-se, ao mesmo passo, que esse laço não se demarque cada vez mais pela insatisfação e por uma abertura quase integral a um modo de ser *tecno-lógico*.

Essa tensão desafia o princípio, mas não impede que haja um reduto de consenso alargado: devemos preservar as relações de cuidado mútuo, que novamente nos remetem para um direito como antropologia, ainda na senda de Rodotà (5), onde é a dignidade – quer nas suas dimensões circunstancialmente radicadas, quer na sua dimensão autoevidente – a medida do adequado. Assim, o dano é configurado sempre na perspectiva do lesado, de

como ele concretamente se sente lesado, mas também da sua configuração hipotética: como eu me sentiria lesado e como posso preveni-lo, sem a ânsia de restringir encadeamentos que só casuisticamente e não universalmente podem ser considerados danosos.

O segundo problema ético-jurídico fundamental: a autonomia no âmbito da atuação de um robô

Afirmámos que eram duas as questões mais flagrantemente ético-jurídicas dessa temática, e que a segunda seria a da autonomia, não por referência ao ser humano da relação, mas concernente ao robô.

Costumou falar-se acriticamente em inteligência artificial autónoma, robôs autónomos, veículos autónomos, mas nessa composição é sobretudo o adjetivo que tem relevância normativa.

O problema pode ser tão só um problema de grau de habilidade, apenas uma questão de perguntar até que ponto a máquina, robô ou *software* executa tarefas sem a mediação, a presença ou intervenção de um ser humano. Nessa perspetiva, a autonomia maquinal é habitualmente dividida em graus, na senda de Sheridan e Verplank, que já em 1978 (15) configuravam níveis 4 e 5 de automação: *high automation* e *full automation*. A automação total exprime os requisitos de independência técnica bem sintetizados por Andresen e Andersen (9): autónomo é o robô que pode identificar um dado problema como problema, estabelecer critérios de decisão, pesar os critérios de decisão, gerar alternativas, avaliar as alternativas no caso concreto, escolher uma das alternativas, implementar a decisão e justificar, a jusante, a opção que tomou.

Na área da saúde, falamos concretamente de três tipologias de materiais robóticos: em primeiro lugar, os de auxílio (níveis 1 e 2 de automação), designadamente os assistentes médicos robóticos que monitorizam estatísticas vitais, os carrinhos robóticos que se deslocam pelos corredores do hospital carregando suprimentos ou os robôs que funcionam como instrumentos de cirurgia, v.g., permitindo incisões menores e/ou mais precisas. Em segundo lugar, os materiais robóticos de avanço (níveis 1, 2 e 3 de automação), que operam, por exemplo, a desinfecção de blocos operatórios, a colocação de substâncias em salas de quarentena ou de isolamento, a recolha e armazenamento de análises clínicas ou montagem de materiais e equipamentos. Finalmente, os materiais de efetiva substituição, nos níveis 4 e 5 de automação, como modelos computacionais *in silico* na investigação clínica, os

modelos computacionais de diagnóstico, os auxiliares de saúde que supervisionam e assistem dos doentes e, finalmente, os materiais robóticos de melhoramento, como os exoesqueletos, os nanorobôs farmacológicos, os robôs de estimulação cerebral e, de uma certa perspectiva, até os *pacemakers* de última geração (16).

Essa autonomia como habilidade diz-nos, porém, pouco sobre a capacidade valorativa e o alcance da intervenção desses robôs nas relações humanas. O que é preciso levar a sério é a agentividade e a ética aplicada à programação.

A agentividade traduz-se, na computação, na capacidade de uma entidade de analisar um ambiente e agir dentro dele. Dentro dessa ampla definição, a maior parte das máquinas, ainda que com parca autonomia, é considerada um agente, o que é amplamente aceito, a partir do trabalho magistral de Russel e Norvig (17).

Na filosofia, contudo, a agentividade apresenta uma configuração diferente. Olhando apenas teorias mais recentes sobre o problema (18), ser considerado agente aparece relacionado com certos estados dos sujeitos que são, na realidade estados mentais: a percepção da individualidade; a utilização de uma energia própria ou autopoietica; e a normatividade do movimento. Encarando assim a questão, parece claro que aquilo que a computação considera um agente não corresponde aos pressupostos filosóficos da agentividade.

Olhando o direito como um projeto cultural e axiológico autónomo, poderá invocar-se que a agentividade juridicamente perspectivada não terá de corresponder a qualquer uma dessas definições. Saber quem é um agente para o direito é uma questão suficientemente rica para acarretar os seus próprios problemas na área da inteligência artificial. Problemas que perguntam pela suscetibilidade de ser titular autónomo de direitos e deveres, pela capacidade de culpa, pela possibilidade de estar por si mesmo em juízo, de ser autor próprio de um ato ilícito. Demos um exemplo concreto: um produto robótico que desempenhe funções de auxiliar de serviços gerais numa unidade de saúde poderia ser considerado, para certos efeitos, um agente jurídico?

Sabemos que a ausência de personalidade jurídica não nos oferece uma resposta completa, pois condomínios ou associações sem personalidade jurídica são tidos por agentes em certos contextos. Não ganharia em operacionalidade a nossa resposta jurídica se configurássemos aqui, por analogia, uma relação comitente – comissário? Porque as soluções de que dispomos, e que se referem a coisas, animais e atividades perigosas,

poderão ser insuficientes para responder aos problemas suscitados nessa área. E sempre terá de se colocar a questão interpretativa, que é a mais pertinente, sobre a relevância normativa dessas normas e a abrangência teleológica das mesmas. É possível, tendo isso em conta, que as relações com robôs convoquem mesmo um estatuto jurídico alternativo.

Esse estatuto jurídico alternativo tem de fazer duas perguntas: o agente autônomo é um agente ético? O agente autônomo pode tornar-se um agente ético?

Quanto à primeira questão, surgem distinções importantes. James Moor (19) distinguiu quatro perspectivas a partir das quais a reflexão sobre os valores pode ser relacionada com as máquinas:

De um primeiro ponto de vista, algumas aplicações da inteligência artificial – sobretudo na computação ordinária – podem ser perspectivadas como agentes normativos, mas não necessariamente como agentes éticos, porque performam apenas ações programadas (19).

De um segundo ponto de vista, algumas aplicações da inteligência artificial podem configurar *agentes de impacto ético*, na medida em que a sua criação e aplicação altera uma situação de facto eticamente relevante; por exemplo, situações em que a utilização de robôs em determinados labores anula a utilização de mão de obra escrava (19).

De uma terceira perspectiva, ainda, existem na inteligência artificial, e na robótica em particular, *agentes éticos implícitos*, isto é, agentes que não foram programados para performar ações éticas, mas tão só para evitar ações antiéticas como, paradigmaticamente, não ofender a integridade de seres humanos (19).

Em uma quarta vertente, porém, podemos configurar *agentes éticos explícitos*: trata-se aqui já de agentes que poderão escolher entre ações éticas e antiéticas, e de entre uma pluralidade de opções de relevância ética (19).

Muitos pensadores poderão argumentar que, ainda que se criem robôs capazes de tomar, autonomamente, decisões éticas, ainda aí não estaremos a criar agentes éticos no sentido em que, tradicionalmente, nos referimos aos seres humanos. Não pretendemos contestar essa posição previsível, apenas fazer notar que aí se tratará, tão só, de uma discussão conceptual, uma vez que, materialmente, os robôs irão, de facto, operar escolhas com relevância ética e, nessa escolha, determinar uma ação num determinado sentido (20).

Isso leva-nos ao problema de saber qual o padrão de princípios que irão basear essa programação (21), e essa é uma questão de suma importância. O padrão do *homem médio*

– ancorado na resposta que a maioria dos seres humanos daria perante determinada situação – é, para já, o padrão em vigor. As máquinas são atualmente programadas segundo um modelo abduativo, para tomar decisões morais apropriadas. Nas palavras de Moniz Pereira (22), “by appropriate moral decisions we mean the ones that conform with those the majority of people make, based on empirical results”.

Mas será esse o padrão mais adequado para a programação de robôs autônomos?

Em todos os tempos, em todos os lugares, alguém teve de escolher os critérios que justificam cada uma das mais importantes regras de convivência social. Esses critérios dos padrões de conduta parecem-nos, muitas vezes, intuitivos, baseados no bom senso; outras vezes, porém, parecem não estar de acordo com aquilo que a maioria das pessoas entende individualmente, mas justificam-se, ainda assim, em nome do bem da comunidade como um todo.

Essa percepção é mais intrincada, porém, se deixarmos de pensar em robôs – máquinas mais ou menos antropomórficas – e considerarmos outras aplicações da inteligência artificial, que também configuram *agentes éticos explícitos*, mas de uma forma menos evidente. Dois exemplos imediatos serão os veículos automatizados, e os nanorrobôs cognitivos.

Relativamente ao primeiro exemplo, Goodall (23) coloca a tónica nesse dilema: os veículos que se autoconduzem, ainda que perfeitamente programados, vão muitas vezes confrontar-se com situações em que a *melhor opção* é moral e/ou juridicamente ambígua. Muito trabalho intelectual tem sido aplicado a essa matéria.

O primeiro grande desafio é programar articulando os valores da sociedade numa variedade de cenários. Isto é especialmente difícil dado que a investigação sobre os padrões de conduta trabalha habitualmente com exemplos que se concentram em escolhas individuais com resultados conhecidos, ao passo que, na realidade, há várias camadas de escolhas e os resultados de cada uma são incertos. O segundo desafio é traduzir essa moralidade para uma *linguagem* que um computador possa entender, sem a capacidade humana de discernir e analogizar.

Relativamente aos nanorrobôs (24) cognitivos que poderão vir a povoar os nossos cérebros, para modificar ou melhorar as suas funções de base, são assunto superlativamente difícil, sobretudo por se encontrarem ainda num domínio conjectural. Interferirão com os processos neuronais que subjazem, por exemplo, à memória e poderão,

eventualmente, tomar decisões autônomas e até reproduzir-se, quando fundidos com elementos biológicos.

Esses pequenos *cyborgs* terão de ser ensinados a comportar-se eticamente e enfrentam os mesmos dilemas que qualquer agente ético explícito.

Conclusão

São muitos, pois, os desafios que nos cabem, encetando com as noções de dano e de autonomia, que vêm sendo utilizadas em sentidos muito distintos, sem que se vislumbre pensamento crítico e sistematizado que envide um consenso ético-jurídico. Essa é uma responsabilidade que bioeticistas e juristas de hoje devem assumir na prevenção de maiores danos àqueles que identificamos como bastiões das relações entre humanos e robôs, particularmente na área da saúde: o princípio do respeito pela recusa em ser tratado por um robô, o princípio da proteção da liberdade humana face a um robô e o princípio da manutenção do tecido social. Observamos que as noções fundamentais de *autonomia* e *dano* não se encontram suficientemente aturadas em ordem a permitir um consenso ético-jurídico operativo, o que é primordial definir antes de avançar com alternativas à programação maquinal atual, que é fundamentalmente utilitarista. E é preciso pensá-lo agora e tomar decisões agora, antes que os danos se tornem evidentes.

Referências

1. Bensamoun A *et al.* Droit de l'Intelligence Artificielle. Issy-les-Moulineux: LGDJ Editions; 2019. p. 25.
2. Albert H. Treatise on Critical Reason. Princeton: Princeton University Press; 1985. p. 99.
3. Füller L. Morality of Law. Yale: Yale University Press; 1964. p. 186.
4. Reale M. Lições Preliminares de Direito. São Paulo: Editora Saraiva; 2001. p. 38.
5. Rodotà S. Il diritto di avere diritti. Milano: Laterza; 2012. p. 200.
6. Rodotà S. Il Corpo Giuridificato - Il Governo del Corpo. Milano: Gioffrè; 2011. p. 51-76.
7. Nie NH, Erbring L. Internet and society: A preliminary report. Stanford, CA: Stanford Institute for the Quantitative Study of Society. [Internet]. 2000 [acesso em 26-08-2020]. Disponível em:
http://www.nomads.usp.br/documentos/textos/cultura_digital/tics_arq_urb/internet_society%20report.pdf.

8. Hogg M. Liability of Unknown Risks: A Common Law Perspective. 15th Annual Conference on Tort Law. Viena: ECTIL – European Centre of Tort and Insurance Law; 2016. p. 26-28.
9. Organização das Nações Unidas. Declaração Universal dos Direitos Humanos. Disponível em <https://www.ohchr.org/EN/UDHR/Pages/Language.aspx?LangID=por>.
10. Estrasburgo. Disposições de Direito Civil sobre Robótica. Resolução do Parlamento Europeu, de 16 de fevereiro de 2017, que contém recomendações à Comissão sobre disposições de Direito Civil sobre Robótica (2015/2103(INL)). Disponível em: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_PT.html.
11. Mori M *et al.* The Uncanny Valley. IEEE Robotics & Automation Magazine. 2012; (19): 98-100.
12. Denton D. As Emoções Primordiais: A Emergência da Consciência. Lisboa: Instituto Piaget; 2010. p. 50.
13. Caudill DS. Lacan and Legal Language: Meanings in the Gaps, Gaps in the Meanings. Law and Critique. Liverpool: Deborah Charles Publications; 1992. p. 169.
14. Wilson EO. O Sentido da Vida Humana (Prémio Pulitzer). Lisboa: Clube do Autor; 2014. p. 107.
15. Sheridan T, Verplank WL. Human and Computer Control of Undersea Teleoperators. [Internet]. 1978 [acesso em 26-08-2020]. p. 8-15. Disponível em: https://www.researchgate.net/publication/23882567_Human_and_Computer_Control_of_Undersea_Teleoperators.
16. Yadav J *et al.* Intelligent Heart Rate Controller for Cardiac Pacemaker. [Internet]. 2013 [acesso em 26-08-2020]. p. 36. Disponível em: https://www.researchgate.net/publication/266010590_Intelligent_Heart_Rate_Controller_for_Cardiac_Pacemaker.
17. Russel S, Norvig P. Artificial Intelligence: A Modern Approach. 3.^a edição. New Jersey: Pearson; 2010. p. 610.
18. Hakli R, Seibt J. Sociality and Normativity for Robots: Philosophical Inquiries into Human-Robots Interaction. Switzerland: Springer; 2017.
19. Andersen M, Andersen SL. Machine Ethics. Cambridge: Cambridge University Press; 2011. p. 83.
20. Moor J. The Nature, Importance, and Difficulty of Machine Ethics. IEEE Intelligent Systems. 2006 (4): 18-21.
21. Andersen M, Andersen SL. Machine Ethics. Cambridge: Cambridge University Press; 2011. p. 7.

22. Moniz Pereira L, Saptawijaya A. Programming Machine Ethics. Switzerland: Springer; 2016. p. 109.
23. Goodall N. Machine Ethics and Automated Vehicles. Road Vehicle Automation. Switzerland: Springer; 2014. p. 93-102.
24. Swan M. Machine Ethics Interfaces: An Ethics of Perception of Nanocognition. In: Rethinking Machine Ethics in the Age of Ubiquitous Technology Information Science Reference; 2017. p. 97-150

Submetido em: 05/12/19
Aprovado em: 28/07/20

Como citar este artigo:

Pereira AEFDD. Problemas ético-jurídicos concernentes aos danos causados por robôs autônomos na saúde: algumas definições conceituais. Cadernos Ibero-Americanos de Direito Sanitário. 2020 jul./set.; 9(3): 12-25.

<http://dx.doi.org/10.17566/ciads.v9i3.613>