ARTIGO

# Responsabilidade civil pelo uso de sistemas de inteligência artificial: em busca de um novo paradigma

**Enrico Roberto**

Pesquisador do InternetLab e do Lawgorithm. Doutorando em direito pela Universidade de São Paulo.

N. 1 / V. 1 / FEVEREIRO DE 2020
PÁGINAS 121 A 143



# Responsabilidade civil pelo uso de sistemas de inteligência artificial: em busca de um novo paradigma

**Palavras-chave**
responsabilidade civil / inteligência artificial / aprendizado de máquina / risco

**Resumo**
No presente artigo, buscamos endereçar o problema de como o direito civil poderia responder a casos de danos causados por sistemas de inteligência artificial. Para tal, traçamos uma definição de sistema de inteligência artificial com foco em uma de suas técnicas de implementação, o aprendizado de máquina. Tais sistemas são definidos, portanto, por sua capacidade de autoaprendizado e de tomar decisões autônomas, sendo ainda desenvolvidos de forma difusa, i.e., por autores diversos e potencialmente anônimos, e em uma "*black* box", i.e., de forma que seu funcionamento interno não possa ser satisfatoriamente esclarecido. A partir do enfoque da "interação homem-máquina", ou seja, da constatação de que sistemas de inteligência artificial são utilizados e desenvolvidos em complemento à ação humana, realizamos breve ensaio sobre a subsunção de regras brasileiras de responsabilização subjetiva e objetiva a tais sistemas, ressaltando os desafios que sua aplicabilidade encontra em diferentes situações. Por fim, realizamos exposição não exaustiva a respeito das iniciativas legislativas no mundo sobre o tema.

Artigo desenvolvido em programa de pós-graduação mediante bolsa de estudos oferecida pela CAPES

N. 1 / V. 1 / JANEIRO DE 2020
PÁGINAS 121 A 143



# Civil liability for the use of artificial intelligence systems: towards a new paradigm

**Keywords**
civil liability / artificial intelligence / machine learning / risk

**Abstract**
In this article, we seek to address the problem of how civil law could respond to cases of damage caused by artificial intelligence systems. For this, we draw a definition of artificial intelligence systems focusing on one of its implementation techniques, machine learning. Such systems are defined, therefore, by their capacity for self-learning and autonomous decision-making. Besides that, we note how their development takes place diffusely, i.e., by diverse and potentially anonymous authors, and in the scope of a black box, i.e., so that their internal functioning can not be satisfactorily clarified. From an approach which we will call the "man-machine interaction", that is, from the observation that artificial intelligence systems are used and developed in addition to human action, we conduct a brief essay on the applicability of Brazilian rules on subjective and objective liability to such systems, stressing the challenges that are posed to such applicability in different situations. Finally, we make a non-exhaustive exposition of the legislative initiatives on the subject taking place in the world.

## 1. Responsabilidade civil pelo uso de sistemas de inteligência artificial: em busca de um novo paradigma

No dia 19 de março de 2018, no Arizona, Estados Unidos, um carro autônomo da Uber atropelou e feriu fatalmente uma mulher de 49 anos, Elaine Herzberg. Logo antes do acidente, segundo vídeo gravado pelo próprio veículo, Elaine atravessou abruptamente a rua mal iluminada, fora da faixa de pedestres, e carregava na mão uma bicicleta com sacos de compras (Levin, 2018). O veículo, um Volvo modificado para dirigir de maneira autônoma, parece não ter feito qualquer manobra ou desacelerado para evitar a colisão, e a motorista "reserva" que estava no carro – com o intuito exatamente de intervir em casos de emergência – não o fez. O caso, a primeira vez em que um carro autônomo levou um pedestre a óbito, é chocante e imediatamente levanta a questão: quem é responsável por essa morte?

Os exemplos de danos causados por sistemas de inteligência artificial (conceito que definiremos adiante) não se limitam a carros autônomos. Outro exemplo interessante ocorreu recentemente em Hong Kong, onde a empresa *Tyndaris Investments*, proprietária da plataforma de investimento autônomo K1, está sendo processada pela perda, por um investidor, de 20 milhões de dólares. O valor foi perdido em vista de uma má decisão de investimento tomada por esse sistema autônomo (Beardsworth, 2019).

É fácil perceber como o uso crescente de sistemas de inteligência artificial pode levar a danos a seus usuários ou outras pessoas. Por mais que os exemplos concretos ainda sejam relativamente limitados, os problemas dessa realidade vêm se impondo pouco a pouco. Assim, diante disso, nos deparamos com um problema: *como o direito civil poderia responder a casos de danos causados por sistemas de inteligência artificial?*

Como apontado, as situações fáticas citadas acima para contextualizar o debate não fornecem, ainda, material suficiente para a elaboração de estudos de caso substanciais. Trata-se de questão ainda incipiente, sem informações suficientes para apresentar uma análise robusta empiricamente sustentada. Porém, com base no que já sabemos, e sem a pretensão de apresentar respostas ou soluções estanques nesse momento, é possível levantar questões que circundam o problema e pensar nas possibilidades hoje postas de resposta jurídica.

Para tanto, propomos, em primeiro lugar, uma revisão bibliográfica introdutória de estudos sobre o funcionamento da inteligência artificial, como forma de compreender exatamente com que tipo de mecanismo estamos lidando, assim como construir o objeto de análise do presente artigo. Neste ponto, definiremos "sistema de inteligência artificial" e "interação homem-máquina", assim como apontaremos duas características de seu desenvolvimento com importante relevância jurídica: sua produção difusa e a opacidade de seu funcionamento (a *black box* da inteligência artificial).

Em seguida, apresentamos um pequeno ensaio sobre como o direito brasileiro, partindo das normas hoje existentes quanto às responsabilidades subjetiva e objetiva, poderia responder a esse tipo de problema real – identificamos possíveis respostas, limites e desafios a esse exercício de subsunção normativa.

Por fim, oferecemos um levantamento não exaustivo das iniciativas legais atualmente existentes, que têm como objetivo regular as atividades envolvendo sistemas de inteligência artificial. Trata-se de estudo que não possui a pretensão de esgotar o tema, mas tão somente introduzir o debate e apontar para possíveis caminhos.

Para o desenvolvimento das ideias do artigo que se segue, em especial o mencionado exercício de subsunção normativa, utilizaremos como ponto de partida, para maior clareza e concretude nos raciocínios apresentados, o primeiro caso apontado acima, o do atropelamento de Elaine Herzberg por um carro autônomo da Uber no Arizona. No entanto, dada a mencionada insuficiência de tal material para estudos empíricos aprofundados e a proposta desse artigo de apresentar caminhos pretensamente aplicáveis a sistemas de inteligência artificial no geral, conforme descrito na formulação de nosso problema acima, procuraremos a todo momento "universalizar" as argumentações trazidas no contexto desse caso específico.

## 2. Inteligência artificial: introdução ao objeto

### 2.1. Sistemas de inteligência artificial

Embora uma definição precisa e abrangente de inteligência artificial possa desempenhar um papel essencial em muitas questões jurídicas e éticas, não é particularmente necessária ou desejável para os fins deste artigo. A despeito disso, importante notar, a título de esclarecimento, que "inteligência artificial" é um termo que, desde sua concepção, nos anos 1950, pressupõe diferentes definições e abordagens, cada qual em seus diferentes contextos. Tais definições e abordagens, muitas vezes focadas na capacidade de emular uma ou outra capacidade humana, foram celebremente estruturadas por Russel e Norvig (2016, p. 2), que identificaram quatro pontos focais diferentes a partir dos quais se pode pensar em inteligência artificial. Especificamente, o termo pode referir-se a máquinas capazes de: (i) pensar como humanos; (ii) agir como humanos; (iii) pensar racionalmente ou (iv) agir racionalmente. Na esteira da possibilidade de "agir racionalmente", elaboraram a hoje frequentemente utilizada definição de *agente racional*: "aquele que age de forma a alcançar o melhor resultado ou, quando há incerteza, o melhor resultado esperado" (Russel & Norvig, 2016, p. 2, tradução livre).

Trata-se, como se vê, de uma definição centrada na ideia de "resultado" e na possibilidade de alcançá-lo de uma forma ou de outra. No entanto, como bem apontado por Scherer (2015, p. 361), a dificuldade de se definir concretamente um sistema a partir de tais noções, em vista da amplitude de significados possíveis para "resultado" ou "melhor resultado esperado", tem por corolário a dificuldade de fixação do termo em uma tecnologia objetivamente delimitada – e, por consequência, sua limitada aplicabilidade a questões regulatórias. Em verdade, conforme passamos a elucidar em seguida, buscamos nos focar para os fins deste artigo em um dos tipos específicos de inteligência artificial: os algoritmos de *machine learning*, ou aprendizado de máquina. Trata-se de vertente da inteligência artificial que, principalmente com o aumento na quantidade de bases de dados disponíveis e na capacidade computacional dos *microchips* de silício, viu impressionante desenvolvimento nos últimos anos (Alpaydin, 2016, p. 1).

Assim, no âmbito desta exposição, um "sistema de inteligência artificial" é definido como *um software que possui capacidades de autoaprendizagem e pode, portanto, tomar decisões autônomas independentes*. Como todo *software*, deve encontrar-se armazenado em algum *hardware*, importância do qual variará entre os diferentes sistemas de inteligência artificial (por exemplo, a importância do *hardware* para delimitar o que constitui um "carro autônomo" é maior do que para delimitar o que constitui um "assistente

de voz", tecnologia normalmente armazenada nos servidores da empresa que a disponibiliza). Com efeito, para muitos autores, é exatamente essa capacidade de "autoaprendizagem" o que caracteriza determinado sistema como imbuído de "inteligência artificial" (Čerka, Grigienėa & Sirbikytėb, 2015, p. 4; Scherer, 2015, p. 365; Calo, 2015, p. 538).[1] Neste item, trataremos brevemente, portanto, dos dois aspectos trazidos por essa definição: (i) capacidade de autoaprendizagem e (ii) decisões autônomas ou independentes.

Como mencionado, as capacidades de autoaprendizagem são possibilitadas e delimitadas por técnicas que se enquadram no conceito de aprendizado de máquina. Trata-se, de maneira geral, de um processo que permite que um sistema aprenda novos fatos a partir de dados sem algoritmos explícitos, bem como adaptar tais fatos aprendidos a novas situações (Alpaydin, 2016, p. 17).

Algumas abordagens comuns à autoaprendizagem incluem os campos da aprendizagem em árvore de decisão, por regras de associação, redes bayesianas, aprendizagem de reforço, *deep learning*, entre outras (Čerka, Grigienėa & Sirbikytėb, 2015, p. 4; Alpaydin, 2016, p. 20; Ertel, 2013, pp. 203-226). Embora estas sejam frequentemente citadas como áreas específicas da inteligência artificial, elas não passam de processos diferentes para o que se apontou acima: em outras palavras, a percepção de padrões em dados para sua conformação em novos cenários, de forma a permitir conclusões não explicitamente buscadas por seus programadores. As diferentes técnicas mencionadas diferem em seus algoritmos e, principalmente, na forma como os dados são fornecidos e como novos resultados surgem a partir desses dados. A capacidade de autoaprendizagem, portanto, refere-se exatamente *à possibilidade que tais sistemas têm de realizar tais inferências não esperadas e não pré-programadas a partir de um conjunto de dados.*

Precisamente porque os sistemas de inteligência artificial não são integralmente limitados por regras humanas predeterminadas, eles podem encontrar soluções que as pessoas não haviam considerado, ou que, mesmo que pareçam a princípio menos intuitivas, são mais eficientes (e.g. do ponto de vista de gasto energético) para atingir os objetivos para o qual os sistemas em questão foram criados (Balkin, 2015, p. 52). É precisamente esta capacidade de criar soluções inesperadas que torna a utilização de sistemas de inteligência artificial cada vez mais atrativa numa variedade de áreas. Às soluções, ou ao "*output*" dos sistemas de inteligência artificial, por envolverem a escolha de determinada solução em detrimento de outras, e como forma de ressaltar seu caráter independente de algoritmos pré-determinados, damos o nome de "*decisão*".

O fato central a se atentar aqui, assim, é que tais decisões não são diretamente decorrentes da programação original de seus desenvolvedores e são, portanto, até certo ponto, incontroláveis, como bem apontado por Scherer (2015, p. 366):

> Pode ser difícil para os seres humanos manter o controle de máquinas que são programadas para agir com considerável autonomia. Há um grande número de formas pelas quais uma perda do controle pode ocorrer: um mau funcionamento, tal como um arquivo corrompido ou dano físico ao equipamento de *input*; uma brecha de segurança; o tempo de resposta superior dos computadores comparados aos seres humanos; ou programação defeituosa. Essa última possibilidade levanta os desafios mais interessantes, porque cria a possibilidade de que uma perda do controle pode ser consequência direta, mas involuntária, de uma escolha de design consciente. O controle, uma vez perdido, pode ser difícil de se recuperar se o sistema de IA for projetado com

> recursos que lhe permitam aprender e se adaptar. Estas são as características que fazem da IA uma potencial fonte de risco público numa escala que excede em muito as formas mais familiares de risco público que são apenas o resultado do comportamento humano. (Tradução livre)

Assim, por tais características, fala-se que as decisões tomadas por sistemas de inteligência artificial são *independentes* ou *autônomas*. Adiante, utilizaremos os termos "decisão autônoma" e "decisão independente" de forma intercambiável.

Por fim, é importante notar aqui que a capacidade de tomar decisões independentes e aprender com a própria experiência é justamente o que torna a inteligência artificial atrativa; não se trata meramente de característica inafastável de uma tecnologia qualquer, mas também sua própria vantagem frente a outras formas de solução de problemas.[2]

## 2.2. Produção difusa e black box

Além disso, há outras duas características da inteligência artificial que têm importante reflexo jurídico e que merecem menção aqui. Muitas vezes, decisões independentes serão, assim, (i) *ininteligíveis* ou *opacas*, decorrência direta da estrutura que baseia seu próprio funcionamento, sujeito a uma inexplicabilidade a que comumente se dá o nome de "*black box*" da inteligência artificial; e (ii) criadas de forma *difusa*, sem possibilidade clara de se especificar a contribuição de um ou outro autor para o resultado final do desenvolvimento do sistema. Explicamos.

Muito se fala da *black box* da inteligência artificial, essa "caixa preta", cujo interior não pode ser visualizado, onde ocorre o processamento do sistema (Knight, 2018). Em determinadas maneiras de aplicação do *machine learning*, especialmente em *deep learning*, as informações externas que são alimentadas ao sistema – os *inputs* – são direcionadas a uma rede de "neurônios artificiais" ou "nodos" que processam os dados e, em seguida, distribuem os comandos necessários – os *outputs* – para operar o sistema no mundo físico ou virtual. No entanto, na maior parte dos casos, ainda não é possível, tecnicamente, refazer o caminho lógico tomado pelos nodos do sistema para saber o porquê de tal operação.

O funcionamento interno de sistemas de inteligência artificial é tão intrincado que até mesmo os engenheiros que os projetam não são tecnicamente capazes de apontar motivos específicos que os levem a tomar determinada decisão (Knight, 2018). E, da mesma forma, não há ainda nenhuma maneira óbvia de projetar tais sistemas para que passem a ser capazes de fornecer tal explicação, por mais que pesquisas nesse sentido tenham sido realizadas nos últimos tempos (Snow, 2017).

Interessante notar que uma das soluções que vem sendo defendida pela academia (Bird, Barocas, Crawford, Diaz, & Wallach, 2016), e que é inclusive objeto do Projeto de Lei nº 2018/49 da cidade de Nova Iorque, Estados Unidos, consiste em estabelecer diretrizes para a criação de mecanismos de transparência em tais sistemas (The New York City Council, 2018), exatamente para facilitar a percepção de elos causais e permitir *accountability* pela sua implementação e uso. Com efeito, a importância da utilização de sistemas de inteligência artificial inteligíveis, inclusive para o direito, vem sendo defendida pela academia (Maranhão, 2019).

Fora isso, a "produção difusa" dos sistemas de inteligência artificial também apresenta desafios. Trata-se de fenômeno que encontra suas bases no movimento do *software* livre, que surgiu na década de 1980 como reação à lógica proprietária das grandes empresas de

desenvolvimento de *softwares*. Os *softwares* livres podem ser definidos, então, como "programas de computador cujo código-fonte é aberto e permite que qualquer um o estude, o copie, o modifique e o redistribua" (Torres, 2013, p. 12).

Nessa esteira, o sucesso do *software* livre como forma de assegurar o acesso público a códigos de programação abriu espaço para a disponibilização, em bibliotecas abertas, de algoritmos ou protótipos de algoritmos, a partir dos quais programadores podem desenvolver livremente seus próprios sistemas – e.g., sistemas de inteligência artificial. Em muitos casos, dada a enorme quantidade de pessoas e empresas, frequentemente anônimas e espalhadas por dezenas de países, que participam na criação de um sistema de inteligência artificial, torna-se tarefa impossível saber quem que contribuiu com o que para determinado projeto.[3] Se se permite que tais contribuições sejam acessadas e utilizadas livremente, ainda por cima, como, por exemplo, por meio de bibliotecas abertas, como a *sci-kit learn*, disponibilizada no GitHub, ou a TensorFlow, do Google, os problemas de responsabilização se multiplicam, por mais que uma tal *open robotics*, tal como defendida por Calo (2010), por exemplo, seja sob muitos vieses desejável.[4]

Temos construído com isso, portanto, nosso objeto. Ao falarmos de sistemas de inteligência artificial, estamos falando de sistemas opacos, desenvolvidos de forma difusa, com a capacidade de autoaprendizado e de tomarem decisões independentes. Tais características, no entanto, por mais que representem, para nossas finalidades, a própria delimitação do que constitui um sistema de inteligência artificial, não expressam a realidade de sua inserção social e uso por seres humanos de maneira absoluta. Para trazê-las à análise jurídica, faz-se necessária a apresentação de outro conceito: o da "interação homem-máquina".

## 2.3. A interação homem-máquina

Sob a alcunha de "interação homem-máquina", buscamos identificar a realidade de que todo sistema de inteligência artificial (ou, com efeito, qualquer máquina) estará inevitavelmente em contato com algum ser humano, seja ele seu desenvolvedor, usuário ou a própria coletividade. Essa realidade toma diferentes nomes em diferentes locais: enquanto profissionais de Tecnologia da Informação preocupam-se em desenvolver "Interfaces Homem-Máquina" (*HMI – Human Machine Interfaces*), a própria base técnica da realidade que buscamos descrever, outros falam de "ciborgues" (Haraway, 2006).

Um exemplo contemporâneo dessa interação se mostra em classificação criada pela *Society of Autonomous Engineers* (SAE), sociedade de padronização de *standards* baseada nos Estados Unidos, e depois utilizada pela *National Highway Traffic Safety Administration* (NHTSA), autoridade federal de trânsito deste país, para ordenar os níveis de autonomia de carros autônomos. A taxonomia proposta classifica o nível de autonomia de veículos em um número de 0 a 5, sendo do nível 5 o carro capaz de trafegar sem qualquer interferência humana e em qualquer condição climática, e do nível 0 o que apresenta somente capacidades automáticas básicas, como a emissão de avisos sonoros em casos de risco (SAE International's On-Road Automated Vehicle Standards Committee, 2013).

A existência dessa taxonomia, assim, revela o aspecto central da interação homem máquina: o fato de que a inteligência artificial serve, em grande parte, como *complemento à ação de seu usuário*, e que somente em poucos casos poderemos falar de decisões finais completamente independentes por parte da máquina. Paralelamente, como é claro, em nenhum caso poderemos falar de um sistema de inteligência artificial *criado* de forma independente de seres

humanos. Com isso, podemos concluir que, na maior parte do tempo, ao se falar de ações pretensamente autônomas, estará ocorrendo, em realidade, uma interação homem-máquina.

Essa observação é de suma importância jurídica. Nesses casos, será sempre importante estabelecer os limites e possibilidades de tal interação: é claro que, em muitas situações, *o humano responderá nos limites de sua esfera de controle e das ações que tomou ou deixou de tomar.* A automação de parte de suas ações não deve afastar o simples fato de que o homem deve responder, subjetivamente, nos limites de sua imputabilidade. Assim, na terminologia utilizada por este artigo, fica claro que o humano responderá no limite de sua atuação no "polo humano" da interação homem-máquina. Nesse ponto, relevante notar que grande parte da legislação atualmente em discussão no mundo para administrar os problemas resultantes da utilização de carros autônomos procura, mesmo que inadvertidamente, criar limites e deveres para o "polo humano" da interação, conforme veremos adiante.

Se de um dos lados da régua da "interação homem-máquina" se encontra a atuação humana, deve-se considerar, juridicamente, analisá-la exatamente sob as regras da responsabilidade subjetiva, aquela efetivamente focada no *sujeito* de direito. A ela contraporemos, nos itens seguintes, a responsabilidade *objetiva*, aquela cujo foco é o objeto, como tentativa de abordagem do outro polo da interação: a do "objeto máquina".

## 3. O polo humano da interação: responsabilidade subjetiva?

Mesmo sendo, em alguns casos, menos propensos a acidentes do que humanos (Smith, 2017, p. 16), sistemas de inteligência artificial também causam – e causaram – danos. O caso de Elaine Herzberg é paradigmático, mas, *do ponto de vista da responsabilidade subjetiva*, relativamente simples. Seria necessário averiguar, na prática, se houve ação ou omissão voluntária, negligência ou imprudência nos termos do Código Civil – por parte de algum ser humano.[5] Nesse caso, o motorista "reserva", um dos polos humanos da interação homem-máquina, poderia ser culpado: será que poderia ter intervindo antes do acidente, e só não o fez por negligência ou imperícia? O teste de subsunção nesse caso já é conhecido do direito há séculos, por mais que os fatos sejam novos, e não cabe aos nossos propósitos delimitá-lo aqui.

A responsabilidade subjetiva também poderia recair sobre os próprios produtores ou outros envolvidos na cadeia de produção do veículo, naturalmente. Seria o caso de peças montadas com imperícia, falta de vistorias legal ou tecnicamente necessárias, etc. Pertinente notar que a produtora do LiDAR (radar de reconhecimento de imagens do carro autônomo) instalado no Volvo do caso do Arizona já alegou sua falta de culpa pela tragédia (Felton, 2018). Nesse caso, a averiguação de responsabilidade subjetiva dos produtores, dada a complexidade do produto e da cadeia produtiva, seria evidentemente bastante dificultosa (não é à toa que o Código de Defesa do Consumidor estabelece a responsabilidade solidária entre os participantes da cadeia de produção). Nos casos de carros autônomos e de outros sistemas de inteligência artificial, as dificuldades multiplicam-se principalmente em vista da difusão de seus produtores e pela *black box* inerente a esse tipo de tecnologia, conforme apontado acima.

A discussão sobre responsabilidade subjetiva de produtores de sistemas de inteligência artificial tem, inclusive, interessante intersecção com as polêmicas sobre viés algorítmico: a discussão sobre vieses inerentes às decisões de algoritmos de *machine learning*, advindos

principalmente da subjetividade na escolha dos dados utilizados para seu treinamento (Giannandrea, 2017). É o caso, por exemplo, de *softwares* de reconhecimento facial que não reconhecem pessoas negras, em vista de não haver representatividade desta população nos dados utilizados para seu treinamento (Breland, 2017). Ou, em caso ainda mais preocupante, *software* utilizado para previsão de crimes nos Estados Unidos, o COMPAS, que concluiu que pessoas negras são mais propensas a cometê-los, resultado esse proveniente do fato de que tal algoritmo foi alimentado com dados de pessoas efetivamente presas – e sujeitas, com isso, aos vieses aos quais os policiais e outros operadores do sistema carcerário americano estão submetidos (Giannandrea, 2017). Se, com a experiência da indústria, passar a ser previsível esse tipo de viés, havendo inclusive formas simples ou boas práticas para evitá-lo, e se há culpa por parte dos desenvolvedores ao criarem *software* propenso a danos em vista da não observação de tais boas práticas, poderia caber a discussão de sua responsabilização subjetiva caso tais danos efetivamente se consumassem.

Claro que, no limite, com a gradativa conscientização pública a respeito dos riscos impostos por essas tecnologias, normas de conduta mais claras para os polos humanos da interação, inclusive usuários do sistema, poderão ser elaboradas. Avisos públicos que informem sobre a implementação de tais ferramentas, locais exclusivos para seu uso (e.g. faixas exclusivas para carros autônomos ou altitudes reservadas para *drones* de entrega), manuais de instruções mais claros e iniciativas similares deverão fazer parte do arcabouço de normas de conduta a serem esperadas dos que interagem com a inteligência artificial.

## 4. Responsabilidade por decisões autônomas independentes

### 4.1. Em busca de um novo paradigma

O caso que se apresenta do outro lado da régua da interação homem-máquina, nesse momento, merece atenção: e as decisões tomadas de *forma efetivamente independente*? Conforme apontado, carros autônomos e outros sistemas de inteligência artificial são "sistemas de autoaprendizagem": imbuídos de algoritmos de *machine learning*, aprendem a tomar decisões a partir de padrões em conjuntos de dados. As decisões que tomam podem ser, portanto, *independentes*, i.e., independem da vontade tanto do fabricante quanto do usuário do sistema. Essas decisões, por estarem fora da esfera de atuação tanto dos fabricantes quanto do usuário do sistema, em regra não lhes poderiam ser atribuídas.

Temos, com isso, sistemas capazes de tomar decisões a partir de experiências e dados, com pouca ou nenhuma interferência humana, cujo processo de tomada de decisão é invisível aos olhos humanos, e que muitas vezes serão produzidos por tantas pessoas concomitantes que apontar responsáveis se tornaria praticamente impossível. Exemplo valioso para ilustrar os limites testados aqui é o do robô Gaak, projeto de pesquisa de uma universidade sueca realizado no ano de 2002 (Higgens, 2002). Nesse projeto, diversos animais-robô foram treinados para agir como "presas" e "caçadores"; as presas procurando por pontos de luz que eram interpretados como "comida" e os caçadores tentando capturar as presas. O intuito era testar

a hipótese evolutiva da sobrevivência do mais forte, e liberar os robôs para desenvolverem estratégias de sobrevivência por si próprios. Ocorre que uma das presas, por razões desconhecidas, começou a circundar a grade do espaço de testes, encontrou uma lacuna, escapou, atravessou uma rodovia nas proximidades e quase foi atropelada por um motorista que dirigia por lá. Trata-se de ilustrativo exemplo de uma decisão autônoma.

Do ponto de vista da responsabilidade subjetiva, dificilmente se poderia falar em negligência ou em omissão nos termos do Código Civil por uma decisão autônoma tomada nesses moldes. Em vista de não haver possibilidade de atuação por parte dos desenvolvedores ou usuários, ou mesmo de estabelecimento de uma relação causal entre suas ações e os danos, em vista da produção difusa e opaca desse tipo de sistema, não se vislumbra a possibilidade de configuração da responsabilidade subjetiva. Da mesma forma, conforme veremos no item seguinte, os institutos atualmente existentes de responsabilidade objetiva apresentam algumas importantes lacunas.

Assim, de forma similar a *bugs* de *software*,[6] que até certo ponto são inevitáveis, decisões autônomas apresentam um risco inerente e que não pode ser completamente extinto: não se pode afastar o fato de que sistemas de inteligência artificial, dada sua relativa autonomia, nunca venham a causar danos. Diversos autores já vêm apontando a existência desse "risco da autonomia",[7] o risco inerente à implementação e uso de sistemas autônomos, propondo diferentes maneiras de administrá-lo. As leis atuais não foram pensadas para a implementação desse tipo de tecnologia, e deve haver profunda discussão pela sociedade e pelas autoridades reguladoras para entender em que medida o direito deve responder a esses desafios. Entre outros, a verdadeira extensão e o desenho legal dos limites da "interação homem-máquina" ganham relevante importância sob este viés.

## 4.2. Responsabilidade objetiva

Se a atuação humana é objeto do direito há milênios, danos causados por objetos independentemente de culpa por parte de seres humanos é matéria relativamente mais recente, mas ainda assim realidade jurídica há muito conhecida (Bittar, 2005, p. 46). Com efeito, teoria corrente da responsabilidade objetiva mostra que esta passa a existir exatamente para fazer frente a um *risco* social não facilmente endereçado pela responsabilidade subjetiva (Marques, Benjamin, & Miragem, 2013, p. 381).

Assim, dada a existência do "risco da autonomia" a que se aludiu acima, parecem tratar-se os danos causados por decisões autônomas de caso claro de aplicação das normas de responsabilização objetiva, como as do Código de Defesa do Consumidor (CDC) (Brasil, 1990) ou a do parágrafo único do Art. 926 do Código Civil (Brasil, 2002), que passaremos a explorar perfunctoriamente neste item.

### 4.2.1. Código de Defesa do Consumidor

Por um lado, danos causados por sistemas de inteligência artificial serão, muitas vezes, ocasionados por defeitos de fabricação ou de programação, o que poderia ensejar, no Brasil, a *responsabilidade objetiva do produtor* do sistema por defeito no produto, nos termos do CDC, caso as exigências dessa lei se apliquem ao caso concreto. No caso do Arizona, poderíamos argumentar, por exemplo, que o fato de o carro não ter brecado, ou não ter reconhecido com seus radares a presença da pedestre atropelada,

constituiria um *defeito*? O Art. 12 do CDC estabelece que "o produto é defeituoso quando não oferece a segurança que dele legitimamente se espera".

A dualidade de argumentos imediatamente se apresenta. Por um lado, contra a aplicação do CDC pode-se dizer, numa perspectiva macroscópica, que o nível de segurança esperado de sistemas de inteligência artificial jamais será absoluto; sempre haverá a chance de acidentes, mesmo que em quantidade menor do que a esperada de condutores humanos. E, com efeito, é difícil afirmar que uma decisão autônoma por parte de um sistema de inteligência artificial constitui um "erro". Em realidade, tomar decisões autônomas *com um certo grau de risco* é um efeito esperado e *desejado* desse tipo de sistema, sendo a existência de danos potenciais em tais decisões amplamente reconhecida e tecnicamente impossível de se afastar, conforme vimos acima. Por outro lado, pode-se dizer que o CDC, na verdade, faz referência à perspectiva microscópica: refere-se à expectativa de segurança de um produto individualmente considerado. E, é claro, espera-se que carros autônomos não atropelem pedestres, ou, de forma geral, que sistemas de inteligência artificial não causem danos.

A depender da maneira como se interpreta a finalidade dessas normas de responsabilização objetiva, pode-se decidir contra ou a favor da aplicação do CDC para tais casos. Nas palavras de Marques, Benjamin e Miragem (2013, p. 381):

> Mister perguntar inicialmente qual seria o fundamento dessa responsabilidade [por defeito no produto]. Seria a culpa do fornecedor ao não agir com a diligência necessária (...)? Seria o risco criado pela atividade dos fornecedores (...)? Ou teria esta responsabilidade como base o resultado objetivo da ação do fornecedor, de ter introduzido um produto com defeito e este defeito ter causado dano ao consumidor (...)?"

Responde aos questionamentos, depois, afirmando haver um "sistema misto" no Brasil, onde todos os fundamentos se misturam. Ressalta, no entanto, que o dever de segurança é "de todos os fornecedores que ajudam a introduzir (atividade de risco) o produto no mercado", mas que "só haverá violação deste dever, nascendo a responsabilidade de reparar os danos, quando existir um defeito no produto. (...) No sistema do CDC, pode haver o dano e o nexo causal entre o dano e produto (...), mas se não existir o defeito, não haverá obrigação de reparar.

Se por um lado se pode dizer que houve a criação de um risco com a introdução do sistema no mercado, seria difícil afirmar, em diversos casos, que uma decisão autônoma tomada por uma máquina deve ser considerada um defeito, exatamente por se tratar de característica desejada e esperada desse tipo de tecnologia. Essa argumentação é tão mais forte quanto mais autônoma e independente de interferência humana for a ação tomada pelo sistema.

De qualquer maneira, se entendermos que o CDC é aplicável ao caso,[8] a discussão precisaria se voltar, neste momento, à apuração da conduta da própria pedestre, para averiguar se, nos termos dessa lei, a culpa foi "exclusivamente do consumidor ou de terceiro". Tratar-se-ia de excludente de responsabilidade objetiva sob o CDC, nos termos de seu Art. 12, §3º, III. Percebe-se aqui que essa apuração de responsabilidade não é a mesma que caberia caso o veículo fosse conduzido por um ser humano, sujeito às regras de responsabilização subjetiva do Código Civil, conforme visto acima. Se fosse esse o caso, a inexistência de negligência ou imperícia por parte do motorista seria suficiente para o afastamento da responsabilidade;

no caso do CDC, somente a culpa exclusiva da pedestre (ou outro terceiro) afastaria a responsabilidade daquele que introduziu o sistema de inteligência artificial no mercado.

Com isso, a lei atual, quando aplicada aos carros autônomos, parece tender a responsabilizar a empresa – mesmo que tal aplicabilidade seja em alguns aspectos questionável. A responsabilização da empresa por acidentes causados por carros autônomos, de fato, é um efeito que vem sendo largamente esperado e explorado: conforme diversos autores vêm notando, a implementação em massa de tecnologias autônomas provavelmente resultará, na prática, em um "*deslocamento de responsabilidade*" dos motoristas, que passam a ser inexistentes ou com esfera de atuação reduzida, aos produtores, que responderão, objetivamente, sob regras normalmente voltadas à proteção consumerista. Nesse sentido, por exemplo, argumentam Bodungen e Hoffmann (2016, p. 503); Smith (2017, p. 1777); Beiker e Calo (2010); Boeglin (2015, p. 172); Horner e Kaulartz (2016, p. 22); e Jänich, Schräder e Reck (2015, p. 313). Os efeitos socioeconômicos desse deslocamento são desconhecidos. Por um lado, argumenta-se, o aumento dos custos e riscos que devem ser assumidos pelas empresas para fazer frente a danos inevitáveis pode inibir a inovação na área e evitar a entrada de concorrentes menores no mercado. Por outro, seria inadmissível que *não houvesse responsáveis* por tais acidentes, e parece justo que seja a empresa – em consonância com o fato de ser decisão comercial sua a implementação do sistema – a encarregada de indenizar civilmente os danos a que deu causa.

Além do desafio imposto pela definição de "defeito" e sua aplicabilidade à inteligência artificial, dado o fato de que sua autonomia é desejada, outro importante desafio se impõe à aplicação do CDC nesses casos: especificamente, o fato de que, em muitos casos, sequer se poderá falar em relação de consumo – é o caso do robô Gaak mencionado acima, por exemplo. Caso se concretize um mundo de *open robotics* e de produção robótica difusa, espera-se a criação de riscos sociais que vão muito além daqueles endereçados pela proteção consumerista.

A *black box* e a produção difusa trazem ainda outros desafios nesse contexto. O primeiro e mais patente é a questão da produção de prova. Para os produtores do sistema de inteligência artificial, caso se encontrem sujeitos ao CDC e, com isso, à inversão do ônus da prova, a *black box* pode levar involuntariamente a um aumento desmedido de sua responsabilidade, já que eles próprios não poderiam provar que a ação tomada pelo sistema não se tratou de um defeito ou vício, por mais que seja esperado que algumas ações autônomas possam causar dano. E, além disso, de forma geral, a limitação do grau de culpa subjetiva dos envolvidos em qualquer tal questão envolvendo sistemas de inteligência artificial ficaria seriamente prejudicada, o que potencializaria o deslocamento da responsabilidade aos produtores conforme apontado anteriormente. Os tribunais e o poder público, quando depararem-se com tais questões, deverão levar em consideração tal impossibilidade técnica no sopesamento de suas decisões.

## 4.3. Código Civil ou um novo tipo de responsabilidade objetiva

Nos casos não abrangidos pela responsabilidade consumerista, poderíamos argumentar pela aplicabilidade do parágrafo único do Art. 927 do Código Civil, que estabelece que "haverá obrigação de reparar o dano, independentemente de culpa, (...) quando a atividade normalmente desenvolvida pelo autor do dano implicar, por sua natureza, risco para os

direitos de outrem." De fato, parece que, na falta de leis específicas para a inteligência artificial, essa normativa encontraria aplicabilidade em diversas situações.

Mesmo em sua simplicidade normativa, no entanto, não se afastam todos os desafios – nesse caso, especificamente, pela delimitação em diversos momentos de quem seria o autor da atividade desenvolvida. Como vimos, o sistema não somente agirá de forma independente, mas também, muitas vezes, será impossível estabelecer quem especificamente agiu de forma a resultar em determinado dano. Isso não só por conta da *black box* da inteligência artificial, mas também pela absoluta *difusão de seus desenvolvedores*.

A realidade é que será impossível apontar uma única pessoa que tenha dado causa a determinado dano, e que caberá ao legislador ou à jurisprudência delimitar o conceito de "atividade normalmente desenvolvida" para o caso de implementação de sistemas de inteligência artificial. Portanto, postos tais desafios, parece-nos que a responsabilidade da empresa sob o CDC ou do "autor" sob o Art. 927 do Código Civil parece ser solução meramente temporária – não pode ser considerada remédio final para os riscos criados por decisões autônomas tomadas por sistemas de inteligência artificial no geral.

Para administrar essa nova categoria de risco social, alguns autores têm defendido, por exemplo, um *novo tipo de responsabilidade objetiva*, baseada primordialmente na noção de "criação de um perigo" ou de "implementação de um robô" (Spindler, 2015, p. 766). Inspiram-se e usam como analogia, por exemplo, a responsabilidade civil pelo comportamento de animais, a responsabilidade de mandantes pelos atos dos mandatários e até mesmo, em referência ao direito romano antigo, a responsabilidade por atos de escravos (Wilzig, 1981, p. 442). A exploração minuciosa dessas propostas e seus possíveis resultados demandaria uma tese por si só. Até mesmo a criação de um novo tipo de capacidade ou personalidade jurídica para os próprios sistemas de inteligência artificial vem sendo defendida, sob diferentes moldes (Teubner, 2018).

## 5. Legislação existente

Sob a ótica da "interação homem-máquina", é possível captar o que muitas leis em discussão vêm fazendo, mesmo que inadvertidamente: estabelecendo limites e obrigações de atuação para empresas e motoristas de carros autônomos, o "polo humano" da interação. De forma geral, deve-se ressaltar, o que se observa das iniciativas regulatórias para a inteligência artificial ao redor do mundo é que vêm se focando na questão dos carros autônomos, mesmo que com algumas importantes exceções.

A legislação do estado da Califórnia, por exemplo, é digna de nota: obriga a empresa produtora de um carro autônomo a garantir o respeito às regras de trânsito por seus veículos, assim como regras de conduta para motoristas "reserva" e motoristas remotos, entre outras. Assim, mesmo que deixe aberta para a empresa a forma de cumprir com a regulamentação, cria deveres de conduta para os humanos envolvidos na atividade. O Arizona, por outro lado, exige mera licença veicular por parte das empresas, não estabelecendo nenhuma outra obrigação. É provavelmente por esse posicionamento, situado em um dos extremos da (pretensa) régua "pró-inovação vs. segurança pública", que seu território vem sendo extensivamente utilizado para esse tipo de teste (Hawkins, 2018). Recomendações do órgão de trânsito alemão, o *Bundesministerium für Verkehr und digitale Infrastruktur* (Redaktion beck-aktuell, 2017), tomam linhas similares: estabelecem regras de atuação e de responsabilização de motoristas reserva e das empresas que

comercializam tais sistemas.

Nesse contexto, vale mencionar também projeto de lei atualmente em discussão no congresso alemão (Deutscher Bundestag, 2017)9. Nele, busca-se a resolução de dilema ético frequentemente apresentado quando se discutem carros autônomos: quando confrontado com a decisão de matar uma pessoa ou outra, ou uma pessoa ou várias outras, como deve o robô proceder? Assemelha-se a versão atualizada do célebre problema ético do "dilema do bonde". O projeto de lei alemão, baseando-se no fato de que toda vida é igual sob a lei, princípio também existente no Brasil, responde com o seguinte preceito: não se escolherá entre vidas ou quantidades de vidas, mas sim a favor da situação que causará "menos dano". A forma de averiguar qual situação causaria menos dano, no entanto, fica também a cargo da empresa.

Fora essas, outra solução prática e com vários precedentes históricos[9] vem sendo proposta pela academia e pelos poderes públicos: seguros obrigatórios pelas empresas que comercializam os carros autônomos. Considerando-se a inevitabilidade em larga escala de danos e a dificuldade de prevê-los ou de determinar individualmente sua causa, conforme vimos acima, a imposição de um seguro obrigatório poderia, também, ao menos em parte, ter efeitos positivos. Com efeito, o seguro obrigatório é exatamente o que se propõe na *Automated and Electric Vehicles Bill* (United Kindgom Parliament, 2017), projeto de lei atualmente em discussão no parlamento britânico, e nas alterações ao código de trânsito do estado da Califórnia, com vigor a partir de 2 de abril de 2018 (State of California, 2018), que obrigam empresas a serem capazes de indenizar até 5 milhões de dólares em danos que seus veículos causem.

De qualquer maneira, regras universais para a responsabilidade civil de sistemas de inteligência artificial no geral não parecem estar sendo discutidas. É questionável, inclusive, se uma única lei seria capaz de regular de forma abrangente o tema, considerando-se os diferentes níveis de risco envolvidos na implementação de diferentes sistemas e a capacidade financeira de seus desenvolvedores, assim como o nível de interferência humana nos comportamentos ditos autônomos. Caberá à sociedade e aos tribunais determinar os limites das soluções que a academia vem gradativamente apresentando, assim como até que ponto elas seriam de fato juridicamente necessárias ou socialmente relevantes.

Nota-se que a atividade estatal em torno da tecnologia, no que *não* se refere aos carros autônomos, parece focar-se na criação de planos nacionais para o seu desenvolvimento – não tanto centrados em sua regulação (por mais que muitos sejam conscientes de seus desafios, como os apontados nesse artigo), mas sim em políticas públicas para sua maior adoção e estímulo ao investimento na área (Lawgorithm, 2019). Duas importantes regras sendo discutidas no âmbito da inteligência artificial, que mencionamos aqui a título de completude, em especial em vista da menção aos problemas da *black box* e do enviesamento algorítmico feita no decorrer deste artigo, são o mencionado Projeto de Lei nº 2018/49, da cidade de Nova Iorque, que estabelece diretrizes para a criação de mecanismos de transparência em tais sistemas, assim como o *Algorithmic Accountability Act*, projeto de lei do Senado estado-unidense que delegaria à Comissão Federal do Comércio desse país (FTC – *Federal Trade Commission*) a criação de regras para avaliação de sistemas automatizados "altamente sensíveis", obrigando empresas que fizessem uso deles a avaliar se os dados e algoritmos que alimentam estas ferramentas são enviesados ou discriminatórios, bem como se representam um risco para a privacidade ou a segurança dos seus usuários (United States Senate, 2019).

Finalmente, essencial mencionar a crescente importância das legislações de proteção de dados nesse contexto, em especial a Lei

Geral de Proteção de Dados brasileira (LGPD) (Brasil, 2018) e a Regulação Geral de Proteção de Dados da União Europeia (RGPD) (European Union, Parliament & Council, 2016), na qual aquela se inspirou. Por se tratar a inteligência artificial de uma (importante) técnica de processamento de dados, como apontamos extensivamente aqui, natural que tais normas encontrem aplicabilidade a diversos aspectos dessa tecnologia. Assim, mesmo que a regulação da proteção de dados não pretenda referir-se exclusivamente à inteligência artificial, seus princípios e direitos têm importante repercussão no desenvolvimento e uso desse tipo de tecnologia. Especificamente, fazemos questão aqui de mencionar o (i) "direito à revisão humana" sobre decisões tomadas unicamente com base em tratamento automatizado de dados, direito estabelecido pelo RGPD e originalmente também pela LGPD (limitado neste caso, no entanto, após um veto presidencial), e (ii) o "direito à explicação", segundo o qual o titular de dados tem o direito de obter informações a respeito da forma como determinada decisão automatizada foi tomada, por exemplo. As dificuldades e desafios da aplicação das regulações de proteção de dados à inteligência artificial vêm sendo bem exploradas pela literatura especializada, tal como nos trabalhos desenvolvidos por Goodman e Flaxman (2017), Selbst e Powles (2017), e Edwards e Veale (2017).

## 6. Conclusão

O acidente fatal em que um carro autônomo da Uber se envolveu foi, ao que tudo indica, o primeiro de muitos casos similares. O aumento do uso desse tipo de tecnologia forçará o direito a encontrar respostas satisfatórias às questões de responsabilização que se levantarão. Nesse artigo, buscamos trabalhar o problema "*como o direito civil poderia responder a casos de danos causados por sistemas de inteligência artificial?*" sob os princípios de responsabilidade civil do direito romano germânico e com foco na lei brasileira. A partir de observações concernentes ao caso do mencionado atropelamento, expandimos seus raciocínios para expor, também, as principais questões apresentadas pela discussão sobre a responsabilidade de sistemas de inteligência artificial no geral.

Para tal, partiu-se de uma definição de sistema de inteligência artificial focada em sua capacidade de autoaprendizado. Para nossos fins, portanto, fala-se desse tipo de sistema quando este apresentar a possibilidade de realizar inferências não esperadas e não pré-programadas a partir de um conjunto de dados. Esta capacidade é possibilitada tecnicamente pelas técnicas de *machine learning*, ou aprendizado de máquina, e dela decorre a capacidade de encontrar soluções – ou decisões – de forma não previsível e não controlada pelos programadores ou usuários do sistema; decisões, portanto, independentes ou autônomas.

Fora isso, ressaltamos duas características importantes da inteligência artificial: o fato de que sua criação se dá frequentemente de forma difusa, i.e., por diversos atores e em diversos locais, de forma muitas vezes anônima e não reconstruível, e de que seu funcionamento ocorre de maneira inexplicável e opaca, dentro de uma *black box* inacessível não somente aos usuários do sistema, mas também a seus desenvolvedores.

Finalmente, apontamos como o uso e desenvolvimento dos sistemas de inteligência artificial não deve ser considerado, para fins jurídicos, como um objeto em si mesmo, devendo nosso foco pousar em realidade sobre a "interação homem-máquina". Trata-se do fato de que todo sistema é utilizado, de alguma forma, como complemento à ação humana, de maneira a se visualizar uma "régua" onde um dos extremos é a mencionada decisão autônoma com mínima ou nenhuma interferência

humana e a outra é o comportamento sob a esfera de ação e controle humanos.

O "polo humano" dessa interação vem sendo estudado pelo Direito há milênios, por mais que os fatos trazidos aqui sejam novos. Do ponto de vista da responsabilidade civil, poderíamos exatamente pensar na responsabilidade do *sujeito* que se encontra em tal polo, na responsabilidade subjetiva.

Assim, nesse caso, deve-se averiguar, para o teste de subsunção do caso de um dano causado com o uso de um sistema de inteligência artificial, se houve negligência, imperícia ou dolo por parte do usuário do sistema – tal como a motorista reserva que se encontrava no veículo no momento do aludido acidente.

Fora isso, a depender da situação, e por mais que a constatação prática disso seja bastante dificultosa, poder-se-ia falar também da responsabilização subjetiva dos produtores do veículo ou de partes dele, como seu próprio LiDAR (radar de reconhecimento de imagens do carro autônomo). Essa averiguação é dificultada também pela *black box* da inteligência artificial, e nos casos de sistemas de inteligência artificial produzidos de forma difusa, pela dificuldade acentuada em se estabelecerem elos causais individualizados entre seus desenvolvedores e o dano causado.

Por mais que a responsabilização subjetiva encontre alguns desafios nesse caso concreto, ela não parece insuficiente, *a priori*, para endereçar os danos causados dentro da esfera de atuação humana. No entanto, quanto mais autônoma for a ação danosa, i.e., quão mais perto do "polo máquina" da interação homem-máquina ela estiver, mais se acentuam determinados desafios. Especificamente, conforme vimos, dificilmente se poderia falar em negligência ou em omissão nos termos do Código Civil por uma decisão completamente autônoma tomada por uma inteligência artificial. Em vista de não haver possibilidade de atuação por parte dos desenvolvedores ou usuários, ou mesmo de estabelecimento de uma relação causal entre suas ações e os danos, e em vista da produção difusa e opaca desse tipo de sistema, não se vislumbra a possibilidade de configuração da responsabilidade subjetiva.

O que se observa, assim, é que o uso de sistemas de inteligência artificial pressupõe um certo "risco da autonomia" que não pode ser facilmente endereçado pela responsabilidade subjetiva. Seria o caso, então, de se pensar em uma responsabilidade objetiva, modalidade exatamente desenhada para fazer frente a determinados riscos impostos à coletividade?

Do ponto de vista da responsabilidade objetiva, poderíamos a princípio nos indagar sobre a aplicação do CDC. As discussões nesse caso são diversas: há diversos argumentos possíveis. Em especial, tratamos da dificuldade do uso do conceito de "defeito" para as decisões autônomas, já que são elementos desejados e esperados desse tipo de sistema. Fora isso, especialmente num mundo de *open robotics* e de produção difusa de sistemas, pressupostos básicos para aplicação do CDC não se aplicariam: não estaríamos falando de produtos colocados no mercado de consumo. Paradigma dessa situação é o caso do robô Gaak, sistema desenvolvido para fins acadêmicos que, por conta de uma decisão autônoma, poderia ter custado a um motorista que passava pelas redondezas a sua vida.

No caso da não-aplicabilidade do CDC, poderíamos aludir ainda ao parágrafo único do Art. 927 do Código Civil. De fato, a amplitude dessa norma, que estabelece a responsabilização "quando a atividade normalmente desenvolvida pelo autor do dano implicar, por sua natureza, risco para os direitos de outrem", permitiria sua aplicação a diversos casos envolvendo danos causados pela inteligência artificial, em especial quando se considera a existência de um "risco da autonomia", conforme vimos. No entanto, a dificuldade de se apontar atores específicos para o dano ou de se delimitar qual

atividade efetivamente dá origem ao risco dificulta também a aplicabilidade dessa norma. Para dar frente a esse risco, alguns autores vêm defendendo, portanto, exatamente a criação de um novo tipo de responsabilidade objetiva, baseada, por exemplo, na "criação de um perigo" ou de "implementação de um robô".

Por fim, passamos rápida e não exaustivamente por algumas das iniciativas estatais que têm sido tomadas nesse contexto ao redor do mundo. Em especial, muitas leis e projetos de lei, tais como os de alguns estados dos Estados Unidos e na Alemanha, tentam criar regras específicas para o desenvolvimento e uso de carros autônomos, criando com isso normas de conduta que deverão, por consequência, facilitar a aplicabilidade prática das regras de responsabilidade objetiva. Fora isso, iniciativas como seguros obrigatórios vêm também sendo discutidas, além de, de forma geral, planos nacionais e outras leis focadas na transparência e não discriminação por sistemas de inteligência artificial.

Com o presente artigo, buscou-se construir um objeto delimitado, o do sistema de inteligência artificial e suas decisões autônomas, assim como uma maneira de abordá-lo metodologicamente, o foco na "interação homem-máquina", para guiar as discussões em torno da responsabilidade civil por danos causados pela inteligência artificial. Como vimos, o pouco material prático e teórico a respeito do assunto, ainda de certa forma reservado a um futuro de curto a médio prazo, não permite conclusões fechadas e subsunções jurídicas claras. Mesmo assim, buscamos oferecer pontos de partida para o debate no assunto, de forma a guiar as atividades dos operadores do direito nos desafios que pouco a pouco se impõem.

## Referências

Agravo em Recurso Especial, AREsp 263077/RJ, Superior Tribunal de Justiça. (2014, 11 de julho). Rel. Min. Raúl Araújo.

*Algorithmic Accountability Act of 2019*. (United States Senate) (EUA). Disponível em https://www.wyden.senate.gov/imo/media/doc/Algorithmic%20Accountability%20Act%20of%202019%20Bill%20Text.pdf?utm_campaign=the_algorithm.unpaid.engagement&utm_source=hs_email&utm_medium=email&_hsenc=p2ANqtz-___QLmnG4HQ1A-IfP95UcTpIXuMGTCsRP6yF2OjyXHH-66cuuwpXO5teWKx1dOdk-xB0b9

Alpaydin, E. (2016). *Machine Learning: The New AI*. MIT Press.

Balkin, J. M. (2015). The path of robotics law. *California Law Review*.

Beardsworth, T., & Kumar, N. (2019, 5 de maio). Who to Sue When a Robot Loses Your Fortune. *Bloomberg*. Future Finance. Disponível em https://www.bloomberg.com/news/articles/2019-05-06/who-to-sue-when-a-robot-loses-your-fortune

Beiker, S., & Calo, R. (2010). Legal aspects of autonomous driving. *SSRN Electronic Journal*. Disponível em http://ssrn.com/abstract=2767899 doi: 10.2139/ssrn.2767899

Bird, S., Barocas, S., Crawford, K., Diaz, F. & Wallach, H. (2016, outubro). Exploring or Exploiting? Social and Ethical Implications of Autonomous Experimentation in AI. *Workshop on Fairness, Accountability, and Transparency in Machine Learning*.

Bittar, C. A. (2005). *Responsabilidade civil: teoria e prática*.

Bodungen, B. V.; Hoffmann, M. (2016). Autonomes Fahren – Haftungsverschiebung entlang der Supply Chain? *NZV-Neue Zeitschrift für Verkehrsrecht*, 29(10&11), (pp. 449-454).

Boeglin, J. (2015). The Costs of Self-Driving Cars: Reconciling Freedom and Privacy with Tort Liability in Autonomous Vehicle Regulation. *Yale JL & Tech.*, 17, 171.

Breland, A. (2017, 4 de dezembro). How white engineers built racist code – and why it's dangerous for black people. *The Guardian*. News. Tech. Disponível em https://www.theguardian.com/technology/2017/dec/04/racist-facial-recognition-white-coders-black-people-police

Calo, R. (2010). Open Robotics. *Md L. Rev.* 70, 571.

Calo, R. (2015). Robotics and the Lessons of Cyberlaw. *California Law Review* (pp. 513-563).

Čerka, P; Grigienėa, J.; & Sirbikytėb, G. (2015). Liability for damages caused by artificial intelligence. *Computer Law Security Review*, 31(3), (pp. 376-389).

Dent, S. (2017, 20 de junho). Tesla driver in fatal Autopilot crash ignored safety warnings. *Engadget*. Transportation. Disponível em https://www.engadget.com/2017/06/20/tesla-driver-in-fatal-autopilot-crash-ignored-safety-warnings/

Deutscher Bundestag. (2017, 30 de março). Straßenverkehrsgesetz für automatisiertes Fahren geändert. Disponível em https://www.bundestag.de/dokumente/textarchiv/2017/kw13-de-automatisiertes-fahren/499928

Giannandrea, J. (2017). Forget Killer Robots—Bias Is the Real AI Danger. W. Knight, Interviewer.

Ertel, W. *(2013). Grundkurs Künstliche Intelligenz: Eine praxisorientierte Einführung*. Springer-Verlag

Edwards, L., & Veale, M. (2017). Slave to the algorithm: Why a right to an explanation is probably not the remedy you are looking for. *Duke L. & Tech.* Rev., 16, 18.

European Union, Parliament, and Council. (2016, 5 de abril). General Data Protection Regulation. *Official Journal of the European Union*, L 119/1.

Felton, R. (2018, 23 de março). LIDAR maker velodyne shifts away blame in fatal uber self-driving crash. *Jalopnik*. Car Technology. Disponível em https://jalopnik.com/lidar-maker-velodyne-blame-to-uber-in-fatal-self-drivin-1824027977.

Goodman, B., & Flaxman, S. (2017). European Union regulations on algorithmic decision making and a "right to explanation". *AI Magazine*, Fall 2017, (pp. 50-57).

Günther, J. P. (2016). *Roboter und Rechtliche Verantwortung: Eine Untersuchung der Benutzer- und Herstellerhaftung* (Vol. 814). Herbert Utz Verlag.

Haraway, D. (2006). A Cyborg Manifesto: Science, Technology and socialist-feminism in the late twentieth century. In The international handbook of virtual learning environments (pp. 117-158). Springer, Dordrecht.

Hawkins, A. J. (2018, 2 de março). The self-driving car war between Arizona and California is heating up. *The Verge*. Google. Tech. Transportation. Disponível em https://www.theverge.com/2018/3/2/17071284/arizona-self-driving-car-governor-executive-order

Higgens, D. (2002, 20 de junho). Robot learns how to escape from exhibition. *Independent*. UK. Home news. Disponível em https://www.independent.co.uk/news/uk/home-news/robot-learns-how-to-escape-from-exhibition-180874.html

Horner, S., Kaulartz, M. (2016). Rechtliche Herausforderungen im Kontext der Industrie 4.0. *Zeitschrift zum Innovations- und Technikrecht*, vol. 1.

Hotchkiss, H. G. (1939). Changing Standards of Liability towards Passengers for Owners and Operators of Aircraft. *Va. L. Rev.*, 25, 796.

Jänich, V. M., Schräder P. T., & Reck, V. (2015). Rechtsprobleme des autonomen Fahrens. NZV-*Neue Zeitschrift für Verkehrsrecht*, 28(7).

Knight, W. (2018, 25 de março). The Dark Secret at the Heart of AI. *MIT Technology Review*. Artificial Intelligence/ Machine Learning. Disponível em https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/

Lawgorithm. (2019, 12 de setembro). Estratégias nacionais de inteligência artificial. Disponível em https://www.lawgorithm.com.br/estrategias-ia/

Lehman-Wilzig, S. N. (1981) Frankenstein unbound: Towards a legal definition of artificial intelligence. *Futures*, 13(6), (pp. 442-457).

*Lei nº 13.709, de 14 de agosto de 2018.* (Brasil). Disponível em http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm

*Lei nº 8.078, de 11 de setembro de 1990*. (Brasil). Disponível em http://www.planalto.gov.br/ccivil_03/leis/l8078.htm

*Lei nº 10.406, de 10 de janeiro de 2002*. (Brasil). Disponível em http://www.planalto.gov.br/ccivil_03/leis/2002/l10406.htm

Levin, S. (2018). Video released of Uber self-driving crash that killed woman in Arizona. *The Guardian*. News. Tech. Disponível em https://www.theguardian.com/technology/2018/mar/22/video-released-of-uber-self-driving-crash-that-killed-woman-in-arizona

Maranhão, J. S. A. (2019, 22 de fevereiro). A importância da inteligência artificial inteligível no direito. *Jota*. Tecnologia. Disponível em https://www.jota.info/paywall?redirect_to=//www.jota.info/opiniao-e-analise/artigos/a-importancia-da-inteligencia-artificial-inteligivel-no-direito-22022019

Marques, C. L., Benjamin, A. H. & Miragem, B. (2013). *Comentários ao Código de Defesa do Consumidor*. Revista dos Tribunais.

Redaktion beck-aktuell, Kabinett beschließt Maßnahmenplan zum automatisierten Fahren. (2017, 23 de agosto). Disponível em https://rsw.beck.de/aktuell/meldung/kabinett-beschliesst-massnahmenplan-zum-automatisierten-fahren.

Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited.

Scherer, M. U. (2015). Regulating artificial intelligence systems: risks, challenges, competencies, and strategies. *Harv. JL & Tech.*, 29, 353.

Selbst, A. D., & Powles, J. (2017). Meaningful information and the right to explanation. *International Data Privacy Law*, 7(4), (pp. 233-242).

Smith, B. W. (2017). Automated Driving And Product Liability. *Mich. St. L. Rev.*, 1.

Snow, J. (2017, 7 de novembro). New Research Aims to Solve the Problem of AI Bias in "Black Box" Algorithms. *MIT Technology Review*. Artificial Intelligence. Disponível em https://www.technologyreview.com/s/609338/new-research-aims-to-solve-the-problem-of-ai-bias-in-black-box-algorithms/

Spindler, G. (2015). Roboter, Automation, künstliche Intelligenz, selbststeuernde Kfz: Braucht das Recht neue Haftungskategorien? *Computer und Recht*, 31(12), (pp. 766-776).

SAE International's On-Road Automated Vehicle Standards Committee. (2013, dezembro). Summary of Levels of Driving Automation for On-Road Vehicles. *Stanford Center for Internet and Society*. Disponível em http://cyberlaw.stanford.edu/loda

*Order to Adopt. Title 13, Division 1, Chapter 1. Article 3.7—Testing of Autonomous Vehicles*. (State of California) (EUA). Disponível em https://www.dmv.ca.gov/portal/wcm/connect/a6ea01e0-072f-4f93-aa6c-e12b844443cc/DriverlessAV_Adopted_Regulatory_Text.pdf?MOD=AJPERES

Teubner, G. (2018). Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagenten. *Archiv für die civilistische Praxis*, 218(2), (pp. 155-205). Disponível em https://www.jura.uni-frankfurt.de/69768539/TeubnerDigitale-RechtssubjekteAcP-18Dez17.pdf

*Law No. 2018/049—A Local Law in relation to automated decision systems used by agencies, 2018*. (The New York City Council) (EUA). Disponível em http://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0

Torres, A. L. (2013). *A tecnutopia do software livre: uma história do projeto técnico e político do GNU*. (Tese de Doutorado, Universidade de São Paulo).

*Bill 112 of 18th October 2017—Automated and Electric Vehicles Bill*. (United Kingdom Parliament) (Reino Unido). Disponível em https://publications.parliament.uk/pa/bills/cbill/2017-2019/0112/18112.pdf

## Notas finais

*1* Vide Alpaydin (2016, p.17), por exemplo: "Machine learning is not just a database or programming problem; it is also a requirement for artificial intelligence. A system that is in a changing environment should have the ability to learn; otherwise, we would hardly call it intelligent."

*2* Vide Scherer (2015, p. 366), por exemplo: "The experiences of a learning AI system could be viewed as a superseding cause — that is, "an intervening force or act that is deemed sufficient to prevent liability for an actor whose tortious conduct was a factual cause of harm" — of any harm that such systems cause. This is because the behavior of a learning AI system depends in part on its post-design experience, and even the most careful designers, programmers, and manufacturers will not be able to control or predict what an AI system will experience after it leaves their care. Thus, a learning AI's designers will not be able to foresee how it will act after it is sent out into the world — but again, such unforeseeable behavior was intended by the AI's designers, even if a specific unforeseen act was not."

*3* Os apontamentos de Scherer (2015, p. 370) são valiosos: "The participants in an AI-related venture may also be remarkably diffuse by public risk standards. Participants in an AI-related project need not be part of the same organization — or, indeed, any organization at all. Already, there are a number of open-source machine- learning libraries; widely dispersed individuals can make dozens of modifications to such libraries on a daily basis. Those modifications may even be made anonymously, in the sense that the identity in the physical world of individuals making the modifications is not readily discernible. The AI program itself may have software components taken from multiple such libraries, each of which is built and developed discretely from the others. An individual who participates in the building of an open-source library often has no way of knowing beforehand what other individuals or entities might use the library in the future. Components taken from such libraries can then be incorporated into the programming of an AI system that is being developed by an entity that did not participate in assembling the underlying machine-learning library."

*4* Vide, em especial, Calo (2010, p. 118): "The widespread availability of robotic platforms capable of running nonproprietary software is more likely to lead to a global robot software industry. Such an industry could take many forms. Anyone could write and share code, or only trusted partners of the platform could be entrusted to do so. Consumers could buy task-specific software permanently or rent it for the day. Importantly, however, the purpose of at least some software would be to enable consumer innovation—that is, to allow consumers to put their robots to new uses."

*5* É exatamente o que se constatou, por exemplo, em um acidente ocorrido com um carro autônomo da Tesla: conforme dados obtidos após o acidente, o motorista atuou durante apenas 25 segundos dos 37 minutos em que o veículo exigiu sua intervenção (Dent, 2017).

*6* Fora isso, importante notar que, por mais que seja corrente, a comparação entre danos causados por inteligência artificial e bugs de software não deve ser levada a suas últimas consequências: conforme vimos, a existência de um campo de atuação e de tomada de decisões imprevisíveis e criativas pela inteligência artificial é economicamente vantajosa

e uma característica desejada por seus desenvolvedores. O bug, por outro lado, é inevitável, mas não é desejado.

7 Günther (2016) fala, por exemplo, de um "Potencial de Perigo" (Gefahrenpotential), conceito intimamente associado à fundamentação da responsabilidade objetiva na doutrina civilista alemã. Scherer (2015, p. 365), por sua vez, fala dos "Riscos criados pela autonomia da IA" (Risks created by the autonomy of AI). Já Teubner (2017), fala diretamente do "Risco da Autonomia" (Autonomierisiko).

8 Nesse artigo, estamos ignorando a querela, há muito discutida pela doutrina e jurisprudência, se o dano causado a terceiro (no caso, a pedestre) deve ser indenizado sob o Art. 12 do CDC, que se refere exclusivamente ao consumidor. Seria necessária a discussão sobre se a pedestre poderia ser considerada consumidora por equiparação, nos termos do Art. 2, parágrafo único, desta lei. Vide por exemplo STJ, AREsp 263077, Rel. Min. Raúl Araújo, publicado em 07/11/2014.

9 Vide a própria existência do APP e do DPVAT no Brasil, assim como a obrigatoriedade de seguros de aeronaves, estabelecida nos Estados Unidos em 1938 (Hotchkiss, 1939, p. 796).

Recebido em 30/06/2019
Aceito em 18/12/2019