

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Faculdade de Direito

Programa de Pós-Graduação em Direito

Giovana Figueiredo Peluso Lopes

**INTELIGÊNCIA ARTIFICIAL (IA): Considerações sobre personalidade, imputação e
responsabilidade**

Belo Horizonte

2020

Giovana Figueiredo Peluso Lopes

INTELIGÊNCIA ARTIFICIAL (IA): Considerações sobre personalidade, imputação e responsabilidade

Versão final

Dissertação apresentada ao Programa de Pós-Graduação em Direito da Universidade Federal de Minas Gerais como requisito parcial para obtenção do título de Mestre em Direito.

Orientador: Prof. Dr. Brunello Souza Stancioli

Coorientador: Prof. Dr. Renato César Cardoso

Belo Horizonte

2020

FICHA CATALOGRÁFICA

L864i Lopes, Giovana Figueiredo Peluso.
Inteligência artificial (IA): considerações sobre personalidade, imputação e responsabilidade [manuscrito] / Giovana Figueiredo Peluso Lopes. – 2020.

Orientador: Brunello Souza Stancioli.
Coorientador: Renato César Cardoso.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Faculdade de Direito.

1. Inteligência artificial – Teses. 2. Responsabilidade civil – Teses. 3. Pessoa jurídica - Teses. I. Stancioli, Brunello Souza. II. Cardoso, Renato César. III. Universidade Federal de Minas Gerais. Faculdade de Direito. IV. Título.

CDU 34:007

DEFESA DE DISSERTAÇÃO DE MESTRADO
ÁREA DE CONCENTRAÇÃO: DIREITO E JUSTIÇA
BEL^a. GIOVANA FIGUEIREDO PELUSO LOPES

Aos sete dias do mês de agosto de 2020, às 14:00 horas, via plataforma virtual - Google Meet, reuniu-se, em sessão pública, a banca examinadora constituída de acordo com o art. 73 do Regulamento do Programa de Pós-Graduação em Direito da Universidade Federal de Minas Gerais, e das Normas Gerais de Pós-Graduação da Universidade Federal de Minas Gerais, integrada pelos seguintes professores: Prof. Dr. Brunello Souza Stancioli (orientador da candidata/UFMG); Prof. Dr. Renato César Cardoso (UFMG); Profa. Dra. Mariana Alves Lara (UFMG) e Prof. Dr. Daniel Mendes Ribeiro (UFJF), designados pelo Colegiado do Programa de Pós-Graduação em Direito da Universidade Federal de Minas Gerais, para a defesa de Dissertação de Mestrado da **Bel^a. GIOVANA FIGUEIREDO PELUSO LOPES**, matrícula nº **2018663342**, intitulada: **"INTELIGÊNCIA ARTIFICIAL (IA): CONSIDERAÇÕES SOBRE PERSONALIDADE, IMPUTAÇÃO E RESPONSABILIDADE"**. Os trabalhos foram iniciados pelo orientador da candidata, Prof. Dr. Brunello Souza Stancioli, que, após breve saudação, concedeu a candidata o prazo máximo de 30 (trinta) minutos para fins de exposição sobre o trabalho apresentado. Em seguida, passou a palavra ao Prof. Dr. Renato César Cardoso, para o início da arguição, nos termos do Regulamento. A arguição foi iniciada, desta forma, pelo Prof. Dr. Renato César Cardoso, seguindo-se-lhe, pela ordem, os Professores Doutores: Mariana Alves Lara, Daniel Mendes Ribeiro e Brunello Souza Stancioli. Cada examinador arguiu a candidata pelo prazo máximo de 30 (trinta) minutos, assegurando a mesma, igual prazo para responder às objeções cabíveis. Cada examinador atribuiu conceito a candidata, em cartão individual, depositando-o em envelope próprio. Recolhidos os envelopes, procedeu-se a apuração, tendo se verificado o seguinte resultado:

Prof. Dr. Brunello Souza Stancioli (orientador da candidata/UFMG)
Conceito: 100

Prof. Dr. Renato Cesar Cardoso (UFMG)
Conceito: 100

Profa. Dra. Mariana Alves Lara (UFMG)
Conceito: 100

Prof. Dr. Daniel Mendes Ribeiro (UFJF)
Conceito: 100

A banca examinadora considerou a candidata aprovada, com nota 100. Nada mais havendo a tratar, o Professor Doutor Brunello Souza Stancioli, orientador da candidata, agradecendo a presença de todos, declarou encerrada a sessão. De tudo, para constar, eu, Fernanda Bueno de Oliveira, Servidora Pública Federal lotada no Programa de Pós-Graduação em Direito da UFMG, mandei lavrar a presente ata, que vai assinada pela banca examinadora e com o visto da candidata.

BANCA EXAMINADORA:

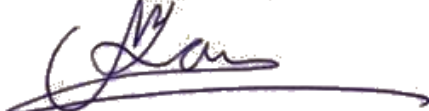
Prof. Dr. Brunello Souza Stancioli (orientador da candidata/UFMG)



Prof. Dr. Renato César Cardoso (UFMG)



Profa. Dra. Mariana Alves Lara (UFMG)



Prof. Dr. Daniel Mendes Ribeiro (UFJF)



- CIENTE: Giovana Figueiredo Peluso Lopes (Mestranda)



Agradecimentos

Embora o processo de escrita de uma dissertação possa ser solitário, tive a sorte de contar ao longo do caminho com o apoio essencial de algumas pessoas, às quais dedico este espaço. Primeiramente, agradeço aos meus pais, Flávia e Oswaldo, e à minha irmã, Maria Eduarda, por terem me formado em valores e caráter que me permitiram trilhar a rota que escolhi para mim. De vocês sempre obtive o apoio necessário para seguir em frente, e por isso serei eternamente grata.

Agradeço aos meus amigos, Ana Clara, Marcela, Samille, Carol, Lorena, Érica e Alexandre, com quem compartilhei minhas incertezas e angústias, e dos quais sempre obtive, em troca, palavras de incentivo. Em especial, agradeço ao Marcos e à Andressa, que ainda de longe compartilharam diretamente dessa jornada comigo, e com quem espero seguir conjuntamente como colegas de profissão.

Agradeço aos meus queridos orientadores, Brunello e Renato, pela extrema empatia e compreensão no período do mestrado em que mais precisei. Obrigada por compartilharem comigo o seu conhecimento e suas ideias, e obrigada pela confiança que sempre depositaram em mim.

Sou imensamente grata, ainda, pela oportunidade de dar continuidade aos meus estudos em uma Universidade Pública, cujas diversas vivências que abriga seguem contribuindo para a minha formação enquanto pessoa. Comprometo-me com a defesa de uma educação gratuita e de qualidade, para que outros possam desfrutar da mesma oportunidade que tive.

Por fim, dedico este trabalho ao Denis, principal responsável por despertar em mim o interesse pela pesquisa acadêmica, e que sempre acreditou no meu potencial para seguir neste caminho.

Resumo

Nas interações sociais e econômicas da atualidade, nota-se com uma frequência crescente a presença de uma nova categoria de ente: o agente artificial. Ele é responsável pela compra e venda de mercadorias online, por determinar a elegibilidade para benefícios como a obtenção de crédito ou assistência médica, por negociações em bolsas de valores, e inúmeras outras aplicações cotidianas. Entretanto, a grande maioria desses agentes artificiais não possui, ainda, uma característica essencial: a autonomia para agir independentemente de inputs diretos de pessoas humanas. Eles são, no máximo, semiautônomos, mas pesquisas no campo da inteligência artificial indicam que este não será o caso em um futuro próximo. Na medida em que agentes artificiais se tornam capazes de agir de maneira verdadeiramente autônoma, isto é, independentemente de orientações humanas diretas, e baseando-se nas informações por eles próprios adquiridas e analisadas, será possível que eles realizem ações não antecipadas por seres humanos. Como resultado, surgem questionamento acerca do tratamento jurídico a ser conferido a estes agentes autônomos de inteligência artificial, sobretudo no que diz respeito à imputação responsabilidade civil em caso de danos, e à possibilidade de alteração de seu status para o de pessoas, em oposição a meros objetos. Assim, a presente pesquisa ocupa-se das seguintes questões: a possibilidade de se personificar agentes artificiais autônomos, e os regimes de responsabilidade civil aplicáveis para os casos em que tal personificação não ocorra. A análise da personificação de agentes artificiais é realizada, portanto, dentro do contexto de existência de uma lacuna de responsabilidade, visando solucioná-la em casos em que a ocorrência de um dano causado pelo agente não possa ser diretamente atribuível a uma pessoa humana. Para tanto, inicialmente, serão apresentadas as principais características capazes de definir uma inteligência artificial como um agente autônomo, e os principais problemas ocasionados, sobretudo para o campo do direito, por esta autonomia e possibilidade de atuação independente de seus programadores e usuários. Em seguida, serão investigadas as principais razões que informam a personificação de entidades não humanas, sobretudo as chamadas pessoas jurídicas, como corporações, a fim de verificar se estas também se fazem presente em agentes artificiais autônomos. Os principais empecilhos atualmente existentes à personificação de agentes autônomos de inteligência artificial serão também apresentados. Finalmente, serão avaliados os diferentes regimes de responsabilidade civil passíveis de serem aplicados em duas hipóteses diversas, quais sejam, danos causados por uma inteligência artificial que podem ser rastreados à atuação de seus programadores, ao design da máquina, aos dados que lhe foram imputados, ou ao comportamento de seus usuários; e danos não diretamente atribuíveis a indivíduos humanos, levantando-se a questão de quem deverá arcar com os custos.

Palavras-chave: inteligência artificial; personalidade; pessoa jurídica; responsabilidade civil.

Abstract

In today's social and economic interactions, a new category of being is noticed with increasing frequency: the artificial agent. It is responsible for buying and selling goods online, for determining eligibility for benefits such as obtaining credit or medical assistance, for trading on the stock market, among numerous different everyday applications. However, the vast majority of these artificial agents do not yet have an essential feature: the autonomy to act independently from the direct inputs of human persons. They are, at most, semi-autonomous, but research in the field of artificial intelligence indicates that this will not be the case in the near future. To the extent that artificial agents become able to act in a truly autonomous way, that is, regardless of direct human orientations, and based on information the agents themselves acquire and analyze, it will be possible for them to perform actions not anticipated by human beings. As a result, questions arise as to how the legal system should treat these autonomous artificial intelligence agents, especially in regards to the attribution of liability in the occurrence of damages, and the possibility of changing their legal status to that of persons, as opposed to mere objects. Thus, this research focuses on the following questions: the possibility of personifying autonomous artificial agents and the applicable liability regime for cases in which such personification does not occur. The analysis of the personification of artificial agents is carried out, therefore, within the context of the existence of a liability gap, and thus aiming to solve it in cases where the occurrence of damages caused by the agent cannot be directly attributable to a human person. In order to do so, initially, the main characteristics capable of defining an artificial intelligence as an autonomous agent will be presented, as well as the central problems caused by its autonomy and the possibility of independent action regarding its programmers and users, especially in the field of law. Then, the main reasons that inform the personification of non-human entities, particularly legal entities such as corporations, will be investigated in order to verify whether these reasons are also present in autonomous artificial agents. The main obstacles currently facing the personification of autonomous agents of artificial intelligence will also be presented. Finally, the different liability regimes that could be applied to them will be evaluated under two different contexts, namely, damages caused by artificial intelligence that can be traced back to the performance of its programmers, to the design of the machine, to the data used to train it, or to the behavior of its users; and damages not directly attributable to human individuals, raising the ultimate question of who should bear the costs.

Keywords: artificial intelligence; personality; legal person; civil liability.

SUMÁRIO

INTRODUÇÃO	7
INTELIGÊNCIA ARTIFICIAL	11
1. Conceito e espécies.....	11
2. Agentes artificiais autônomos	23
3. Os problemas decorrentes da autonomia de agentes artificiais.....	30
3.1. Imprevisibilidade	31
3.2. Ininteligibilidade	41
4. Personalidade jurídica (eletrônica): solução à lacuna de responsabilidade?	50
AGENTES ARTIFICIAIS AUTÔNOMOS COMO PESSOAS JURÍDICAS (ELETRÔNICAS)	56
1. O debate acerca da personificação de agentes artificiais autônomos	56
2. As razões da pessoa jurídica	66
3. Agentes autônomos de IA como pessoas jurídicas	79
3.1. Argumentos contrários à personificação	82
A RESPONSABILIDADE CIVIL POR DANOS CAUSADOS POR UM AGENTE ARTIFICIAL AUTÔNOMO.....	105
1. Danos diretamente atribuíveis a seres humanos.....	107
2. Danos não diretamente atribuíveis a seres humanos	113
3. A contratação de seguros para danos causados por um agente autônomo de IA	123
CONSIDERAÇÕES FINAIS	130
REFERÊNCIAS	133

INTRODUÇÃO

A ideia de que seres humanos conseguiriam, em determinado momento histórico, desenvolver máquinas capazes de tomar decisões por si próprias e obter autonomia relativamente a seus criadores, é algo que permeia o imaginário coletivo desde tempos primordiais, servindo como pano de fundo para contos, mitos, obras literárias, teatrais e cinematográficas. Cita-se, por exemplo, o mito judaico do golem; o clássico romance “Frankenstein ou o Moderno Prometeu” de Mary Shelly; as obras distópicas de Isaac Asimov como “Eu, Robô”; o popular desenho animado The Jetsons, produzido pela empresa Hanna-Barbera; e a personagem HAL 9000, principal antagonista de 2001: Uma Odisseia no Espaço, de Arthur C. Clarke.

Entretanto, o que antes se apresentava como mera mitologia ou ficção científica, nas últimas décadas passou a configurar uma real possibilidade, cada vez mais presente no cotidiano do mundo globalizado, e trazendo consigo a promessa de revolucionar aspectos fundamentais da vida em sociedade.

Atualmente, sistemas baseados em inteligência artificial (ou apenas IA) se fazem presentes nas mais diversas áreas, tendo deixado os laboratórios e alcançado o dia-a-dia de grande parte da população global. Essa (não tão nova) tecnologia encontrou no mundo digital e na economia movida a dados o ambiente propício para alcançar projetos antes impensáveis, com aplicações tanto benéficas quanto ameaçadoras.

Hoje, a pervasividade de sistemas de IA passa muitas das vezes despercebidas por aqueles que dela fazem uso: eles estão presentes em provedores de aplicação do comércio eletrônico, nas principais plataformas de busca e compartilhamento de dados, nas sugestões adaptadas ao perfil de consumidores em plataformas de música e streaming, e nas redes sociais. É crescente o número de consumidores que adotam em seus lares assistentes pessoais.

Além disso, previsões de especialistas estimam que carros totalmente autônomos possivelmente entrarão no mercado de consumo entre 2030 e 2035, robôs são utilizados em linhas de montagem, embalagem e delivery de bens de consumo, e objetiva-se, no futuro, empregá-los como assistentes terapêuticos de indivíduos com algum tipo de trauma, deficiência ou distúrbio.

O sistema financeiro global, por sua vez, tem cada vez mais feito uso desse tipo de sistema, capaz de identificar padrões imperceptíveis ao olho humano e, consequentemente,

antecipar as movimentações do mercado e executar complexas operações financeiras em questão de segundos.

Os avanços em sofisticação técnica e funcionamento autônomo de sistemas de inteligência artificial representam uma continuação lógica da adoção, por seres humanos, de tecnologias de automação, constituindo uma nova extremidade de um *continuum* de tecnologias que visam automatizar e aperfeiçoar capacidades humanas, ampliando seu aparato cognitivo e, em última instância, tornando-se modelos aprimorados de si mesmos.

Cumprе ressaltar, porém, que até o presente momento, por mais sofisticados e inovadores que sejam tais sistemas de inteligência artificial, pode-se afirmar que eles são, no máximo, semiautônomos. Significa dizer que é possível identificar em suas ações as diretrizes humanas responsáveis por determinado funcionamento ou tomada de decisão. Em outras palavras, há um input ou supervisão diretos por parte do programador, operador, ou usuário do sistema. Disso resulta que é possível rastrear a origem de determinada ação, na maioria das vezes, à forma como o software foi programa, ao design da máquina, aos dados que foram alimentados ao sistema ao longo do processo de aprendizagem ou então à supervisão adequada ou não por parte de seu responsável.

Sistemas de IA verdadeiramente autônomos, por outro lado, trazem consigo novos desafios, uma vez que se elimina, ao menos diretamente, o elemento humano do processo de tomada de decisão. Daí resulta a possibilidade de que eles executem determinadas ações não antecipadas, que poderão, eventualmente, causar danos aos seus usuários ou terceiros.

A noção de previsibilidade constitui um elemento fundamental em sistemas de responsabilidade civil, sendo comum a análise da possibilidade de se antever uma ação danosa quando da alocação da responsabilidade por um dano. No campo da inteligência artificial, entretanto, a imprevisibilidade pode constituir um elemento crucial para o sucesso de determinada empreitada, por viabilizar o surgimento de resultados e alternativas não cogitadas por seres humanos e, por conseguinte, inovadoras.

Na hipótese de uma inteligência artificial autônoma provocar um dano não diretamente rastreável a uma pessoa humana envolvido em seu processo de criação ou utilização, ter-se-ia a situação de uma lacuna de responsabilidade. Uma vez que a IA desse experimento não é considerada como um sujeito pelo ordenamento jurídico, mas como mera coisa, não seria possível, *a priori*, imputar-lhe a responsabilidade pelo fato danoso.

Diante do cenário que se apresenta, no presente trabalho buscou-se investigar a possibilidade de se atribuir o status jurídico de pessoa a um agente autônomo de inteligência

artificial. E, como decorrência deste problema principal, foram também analisadas as alternativas existentes no âmbito da responsabilidade civil para lidar com a imputação de danos decorrentes de uma IA, levando-se em conta o seu status atual de mero objeto.

Devido à amplitude das bases teóricas sobre as quais se fundou a pesquisa, foi necessário o seu desenvolvimento a partir de uma perspectiva interdisciplinar. Assim, além dos aspectos estritamente jurídicos do problema colocado, foram também abordados aspectos relativos à ciência da computação, aspectos técnicos do campo da inteligência artificial, e análises filosóficas e históricas dos temas objetos da pesquisa.

A metodologia utilizada no trabalho é de natureza essencialmente bibliográfica, valendo-se de fontes documentais diretas e indiretas para uma revisão compreensiva do tema. Trata-se, ainda, de pesquisa de caráter teórico-propositivo, consistindo numa abordagem qualitativa por traços de significação, por meio do procedimento da análise de conteúdo.

No primeiro capítulo do trabalho, intitulado “Inteligência Artificial”, buscou-se inicialmente apresentar a evolução histórica da IA como um campo de pesquisa independente da ciência da computação, e apresentar as principais espécies desta tecnologia, separando-as em dois grandes blocos: aquelas baseadas em procedimentos de programação clássica, através da manipulação formal de símbolos, e aquelas baseadas em processos de aprendizado de máquina, que fazem uso de grandes volumes de dados para gerar inferências.

Apresentar um conceito exato de inteligência artificial revelou-se uma tarefa difícil, tendo em vista as divergências existentes sobre o tema que levam em conta variáveis: comparações à parâmetros humanos de inteligência, amplitude do domínio de aplicações a ser executado pela IA, variedade de técnicas existentes, dentre outros.

Optou-se, portanto, por focar na definição de um “agente artificial autônomo”, sendo analisados os pressupostos que constituem a noção de agência, a influência exercida pelo ambiente sobre determinado agente, a conceituação de autonomia e sua oposição à mera automação. Desse modo, limitou-se o objeto de investigação da pesquisa de todo o universo de aplicações da inteligência artificial a apenas agentes capazes de agir independentemente de instrução humana direta, com base nas informações que ele próprio adquire e analisa, direcionando suas ações em um ambiente imprevisível ou em constante modificação.

A partir dessa definição, foram apresentados os principais problemas decorrentes da autonomia de agentes artificiais, a saber: a sua imprevisibilidade e a sua ininteligibilidade, bem como as implicações de tais características para o ordenamento jurídico, introduzindo-se,

desse modo, a temática da personificação de tais agentes como possível resposta à lacuna de responsabilidade.

Assim, em “Agentes Artificiais Autônomos como Pessoas Jurídicas (Eletrônicas)”, capítulo de número dois deste trabalho, foi analisada a possibilidade de atribuição de personalidade a esses agentes. Tal proposta não se pauta em uma crença, até o momento pouco provável, na semelhança entre um agente artificial e uma pessoa humana (natural), com a consequente atribuição de direitos fundamentais como a proteção à vida ou à integridade física, mas sim em analogia às pessoas jurídicas, como centros de imputação de alguns direitos e deveres apropriados à sua maneira de ser.

Para tanto, foram analisados os principais argumentos no debate ocorrido, sobretudo na década de 1990, a respeito da personificação de inteligências artificiais (à época infinitamente inferiores), sobretudo em dois campos distintos: a formação de contratos eletrônicos, e a responsabilidade civil. Os argumentos apresentados em tal debate justificaram a posterior investigação das razões que influenciaram a atribuição de personalidade a entes não humanos, especialmente corporações, em quase todos os ordenamentos jurídicos ocidentais modernos. Ao final do capítulo, são apresentados argumentos contrários e favoráveis à criação de um novo status jurídico de pessoas (eletrônicas, em menção à sugestão proposta pelo Parlamento Europeu na Resolução de 16 de fevereiro de 2017) para agentes autônomos de inteligência artificial.

Finalmente, no terceiro e último capítulo do presente trabalho, há um retorno ao problema inicialmente apresentado: o da imputação de responsabilidade por danos causados por uma inteligência artificial. Assim, em “A responsabilidade civil por danos causados por um agente artificial autônomo”, são avaliados os possíveis regimes de responsabilidade civil passíveis de serem aplicados a dois casos distintos.

Em primeiro lugar, hipóteses em que é a atribuição direta de um dano a uma pessoa humana, seja ela programadora, operadora, usurária. E em seguida, o caso de danos não atribuíveis a seres humanos diretamente, em virtude do aspecto imprevisível do agente artificial. Nesta segunda hipótese, foco é dado à responsabilidade objetiva, ou seja, que prescinde da análise do elemento de culpa, e à escolha política de quem deve arcar com os danos.

Por fim, serão feitas as considerações finais sobre os principais resultados obtidos através da pesquisa realizada.

CAPÍTULO I

INTELIGÊNCIA ARTIFICIAL

1. Conceito e espécies

No verão de 1956, na *Dartmouth College*, um grupo de dez cientistas que compartilhava o interesse em redes neurais, na teoria dos autômatos e no estudo da inteligência, reuniu-se para a realização de um workshop de seis semanas, idealizado por John McCarthy, Marvin Minsky, Claude Shannon e Nathan Rochester. O *Dartmouth Summer Research Project on Artificial Intelligence*, como foi chamado, é comumente considerado o marco inaugural da inteligência artificial como campo de pesquisa.

Conforme a proposta apresentada à Fundação Rockefeller, patrocinadora do evento em questão:

O estudo será realizado com base na hipótese de que todos os aspectos da aprendizagem ou de qualquer outra característica da inteligência podem, em princípio, ser descritos de forma tão precisa que uma máquina poderia simulá-los. Será feita uma tentativa de descobrir como produzir máquinas que usem linguagem, formem conceitos e abstrações, resolvam determinados tipos de problemas até agora reservados aos humanos e que sejam capazes de se aperfeiçoar. Acreditamos que um avanço significativo possa ser feito em um ou mais desses problemas se um grupo de cientistas cuidadosamente selecionados trabalhar em conjunto no projeto durante um verão (tradução nossa).¹

McCarthy e muitos de seus colegas eram aficionados pela lógica simbólica, ramo da matemática que lida com a representação de conceitos e afirmações em símbolos, definindo as transformações necessárias para manipulá-los a raciocinar dedutiva ou indutivamente. Na época, os matemáticos estudiosos do tema não estavam preocupados tanto com a sua aplicação a problemas específicos, mas sim com a investigação das propriedades teóricas de sistemas lógicos².

Entretanto, o advento de dispositivos de computação eletrônica levantou a possibilidade da aplicação prática das teorias desenvolvidas, sobretudo após a Segunda Guerra Mundial. Nesse contexto, pode-se avaliar a conferência em Dartmouth como uma

¹ MCARTHY, John; MINSKY, Leon; ROCHESTER, Nathanael; SHANNON, Claude. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence. 1955, p. 1-13. Disponível em: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. Acesso em 29 out. 2019.

² KAPLAN, Jerry. *Artificial Intelligence: What everyone needs to know*. New York: Oxford University Press, 2016, p. 14.

tentativa de expandir o uso de computadores para além do mero processamento de números e dados para a manipulação simbólica³.

Foi também na proposta para o projeto de pesquisa em que se utilizou, pela primeira, o termo “inteligência artificial” para designar o intuito de fazer com que uma máquina se comporte de maneira que, caso se tratasse de um ser humano, seria considerada inteligente⁴.

De fato, embora exista uma miríade de definições propostas para a inteligência artificial, boa parte delas encontra-se alinhada a essa ideia central, como a proposição quase tautológica de Kurzweil de que a inteligência artificial consistiria na arte de criar máquinas que executem funções que demandam inteligência quando realizadas por pessoas,⁵ ou no desenvolvimento de soluções automatizadas por sistemas ou agentes para problemas que requereriam a intervenção da inteligência se executados por humanos.⁶

Alterando-se o foco da agência para a modelagem cognitiva, é comum ainda encontrar definições estruturadas em termos da tentativa de reprodução, em sistemas artificiais, da cognição humana e seus mais variados componentes, como o aprendizado, a memória e o processo de tomada de decisões. Nesse sentido, Boden elabora resumidamente o desafio da inteligência artificial como sendo o de fazer computadores realizarem tarefas típicas da mente humana⁷.

Note-se que, antes mesmo do surgimento da inteligência artificial como campo independente de pesquisa em 1956, é possível apontar trabalhos anteriores que abordaram a temática da replicabilidade de aspectos cognitivos e comportamentais humanos em sistemas artificiais.

Por exemplo, a observação feita por McCulloch e Pitts, já em 1943, de que uma rede de neurônios cerebrais poderia ser modelada por expressões lógicas, uma vez que a sinalização, no cérebro humano, seria binária, similarmente ao que ocorre em um mecanismo digital⁸ – e que futuramente serviria de base ao desenvolvimento de redes neurais artificiais. Ou ainda o influente trabalho de Alan Turing “*Computing Machinery and Intelligence*”, o

³ Id., p. 14.

⁴ MCCARTHY, John et al. Op. cit., p. 11.

⁵ Veja-se: KURZWEIL, Ray. *The Singularity is Near: When humans transcend biology*. Nova York: Penguin Books, 2005.

⁶ NEGNEVITSKY, Michael. *Artificial Intelligence: A guide to intelligent systems*, 2 Ed. Harlow: Addison-Wesley, 2004, p. 2.

⁷ BODEN, Margaret. *AI: Its nature and future*. Oxford: Oxford University Press, 2016.

⁸ No sentido oposto de que o cérebro humano também computaria utilizando-se de componentes analógicos, veja-se: NICOLELIS, Miguel A. L.; CICUREL, Ronald. *O Cérebro Relativístico: Como ele funciona e por que ele não pode ser simulado por uma máquina de Turing*. São Paulo: Kios Press, 2015.

qual introduziu conceitos fundamentais da área, como o jogo da imitação, aprendizado de máquina, algoritmos genéticos e aprendizado por reforço⁹.

Atualmente, pouco mais do que mera especulação vincula o trabalho real em IA aos funcionamentos da mente humana, sendo uma disciplina cuja relação com organismos biológicos é, ao menos no estágio presente, principalmente metafórica – questão que será discutida adiante. Entretanto, à época de seu surgimento, não era raro que profissionais da área fizessem promessas e projeções excessivamente otimistas sobre o que seria alcançado pela inteligência artificial e quanto tempo levaria a iniciativa para atingir seus objetivos¹⁰.

Dentre as afirmações otimistas frequentes nos anos 1960, pode-se citar a de Minsky, segundo o qual o problema da criação da inteligência artificial seria substancialmente resolvido no intervalo de uma geração¹¹, ou a previsão de Simon de que, em vinte anos, as máquinas seriam capazes de fazer qualquer trabalho que um homem fosse capaz de fazer.¹²

Entretanto, nas décadas subsequentes à conferência em Dartmouth, o recém criado campo da inteligência artificial passou por períodos de expansão, pautados por grande entusiasmo e financiamentos abundantes tanto do setor público quanto do privado, alternados com períodos de retração (estes comumente denominados “*AI winters*”), na medida em que a resolução de alguns dos problemas propostos revelou-se mais difícil do que o previsto inicialmente, e outros, simplesmente insuperáveis com as tecnologias disponíveis à época.

A princípio, os pesquisadores da área, motivados pela teoria de que muitos avanços científicos ocorrem quando suposições podem ser simplificadas, ou fenômenos estudados isoladamente, direcionaram seus esforços a domínios bem definidos e limitados (“micromundos”, conforme o jargão utilizado) – como jogos e lógica¹³.

Isto possibilitava uma versão reduzida da performance a ser demonstrada, comprovando-se o conceito e mostrando que, em princípio, determinada atividade poderia ser

⁹ TURING, Alan. Computing Machinery and Intelligence. *Mind*, v. 49, n. 236-460, 1950. Disponível em: <https://phil415.pbworks.com/f/TuringComputing.pdf>. Acesso em: 29 nov. 2019.

¹⁰ Para uma revisão sistemática das previsões sobre IA, ver ARMSTRONG, Stuart; SOTALA, Kaj. How We’re Predicting AI – or Failing To. In: ROMPORTL, Jan; ZACKOVA, Eva; IRCING, Pavel; POLAK, Michal; SCHUSTER, Radek (Eds.). *Beyond AI: Artificial Dreams*. Pilsen: University of West Bohemia, 2012, p. 52–75.

¹¹ MINSKY, Marvin. *Computation: Finite and infinite machines*. Englewood Cliffs: Prentice-Hall, 1967, p. 2.

¹² SIMON, Herbert. *The Shape of Automation for Men and Management*. New York: Harper & Row, 1965, p. 96.

¹³ Por exemplo, o *Logic Theorist*, programa de computador para inferência lógica criado por Newell, Shaw e Simon e 1957, e que validou o trabalho teórico de Russell e Whitehead na obra seminal *Principia Mathematica*; e a linguagem de programação LISP (*list processing*) desenvolvida por McCarthy em 1958, que se tornaria a linguagem dominante de programação em IA pelos próximos 30 anos.

feita por uma máquina¹⁴ - em resposta aos chamados “*arguments from various disabilities*”, formulado em termos de uma ‘uma máquina nunca poderia fazer X’¹⁵.

Tal abordagem baseada em métodos de raciocínio e lógica para o alcance de determinado objetivo também foi motivada por necessidade, vez que, à época, eram escassas as informações disponíveis ou passíveis de armazenamento em formato digital, bem como existiam severas limitações de hardware, tanto em capacidade de memória quanto em velocidade de processamento.

Tais métodos que produziram sucessos nos primeiros sistemas de demonstração posteriormente apresentaram dificuldades de extensão para uma variedade mais ampla ou para versões mais complexas dos problemas abordados, razão pela qual foram denominados métodos fracos.

Uma razão para tanto é a chamada explosão combinatorial de possibilidades, que devem ser exploradas por meio de busca exaustiva. Esta, embora funcione bem para instâncias simplificadas de um problema, tem seu desempenho reduzido na medida em que a complexidade aumenta¹⁶. Para solucionar a questão, seria necessária a adoção de uma estratégia de pesquisa informada (heurística), que faz uso de conhecimentos prévios acerca do problema, para além de sua mera definição, sendo apta a encontrar soluções com mais eficiência do que uma estratégia desinformada¹⁷.

Entretanto, tais capacidades eram ainda pouco desenvolvidas nos sistemas iniciais de inteligência artificial, e a crescente constatação de que muitos dos projetos desenvolvidos jamais poderiam realizar suas promessas iniciais levou ao primeiro período de retração do campo, com a diminuição de financiamentos e aumento do ceticismo.

Essa situação foi revertida por volta dos anos 1980, com a proliferação dos chamados sistemas especialistas. Projetados como ferramentas de suporte para tomadas de decisão, tais sistemas eram programas baseados em regras que faziam inferências simples a partir de uma

¹⁴ BOSTROM, Nick. *Superinteligência: Caminhos, perigos e estratégias para um novo mundo*. Rio de Janeiro: DarkSide, 2014, p. 29.

¹⁵ Várias habilidades são elencadas por Turing para preencher o valor de X, como: “ser gentil (...) possuir senso de humor, distinguir o certo do errado (...) aprender com a experiência, usar as palavras corretamente, estar sujeito a seus próprios pensamentos, possuir tanta diversidade comportamental quanto o homem (...)” (TURING, Alan. Op. cit., p. 447, tradução nossa).

¹⁶ Por exemplo, para provar um teorema que tenha uma prova de cinco linhas num sistema de dedução com uma regra de inferência e cinco axiomas, seria possível enumerar as 3.125 combinações possíveis e conferir cada uma delas a fim de checar sua correspondência com a conclusão pretendida. Entretanto, tão logo as tarefas se tornem mais complexas, o método de busca exaustiva torna-se computacionalmente inviável até mesmo para supercomputadores mais rápidos (BOSTROM, Nick. Op. cit., p. 30)

¹⁷ RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 3. ed. New Jersey: Prentice-Hall, 2010, p. 92.

base de conhecimento composta de fatos extraídos de especialistas na área e programados manualmente em linguagem formal¹⁸.

Os sistemas especialistas lidavam com questões típicas de áreas restritas de expertise (*domain specific*), desconstruindo tarefas em dois componentes: a “base de conhecimento” - uma coleção de fatos, regras e relacionamentos sobre um domínio de interesse específico representado em forma simbólica - e um “mecanismo de inferência” de uso geral que descrevia como manipular e combinar esses símbolos¹⁹.

Na época, foram construídos centenas de sistemas do tipo²⁰, e a indústria de IA cresceu de alguns milhões de dólares em 1980 para bilhões de dólares em 1988²¹. No entanto, sua especificidade logo se revelou problemática, na medida em que alguns programas se mostraram caros demais para serem desenvolvidos, validados e atualizados, além de serem de difícil atualização²².

Era impraticável adquirir um computador apenas para rodar um único programa, e extremamente trabalhoso codificar minuciosa e manualmente o conhecimento de experts. Assim, na medida em que o fim da década se aproximava, uma outra temporada de crescimento no campo da inteligência artificial também encerrou seu ciclo, levando a afirmações de que a história da pesquisa em IA, até o momento, havia sempre consistido em “sucessos muito limitados em áreas específicas, seguidos imediatamente pelo fracasso em alcançar os objetivos maiores sugeridos, à primeira vista, por esses sucessos iniciais”²³.

Todavia, aumentos incrementais em energia, armazenamento e poder computacional logo levariam a uma explosão de dados prontamente acessíveis em formato eletrônico, o que possibilitaria o surgimento de uma abordagem completamente diferente para incorporar conhecimentos em programas de computador, elevando a inteligência artificial a patamares sem precedentes.

Desde os primórdios da pesquisa em IA, estudiosos reconheceram que a capacidade de aprender constitui um importante aspecto da inteligência humana. O aprendizado, presume-se, dá-se principalmente a partir de experiências, práticas e treinamentos, não estando limitado apenas ao raciocínio.

¹⁸ BOSTROM, Nick. Op. cit., p. 31.

¹⁹ KAPLAN, Jerry. Op. cit., p. 23.

²⁰ Por exemplo, a linguagem de programação lógica PROLOG desenvolvida por Colmerauer e Roussel em 1972; e o programa para compreensão de linguagem natural SHRDLU de Winograd, desenvolvido entre 1968 e 1970.

²¹ RUSSELL, Stuart J.; NORVIG, Peter. Op. cit., p. 24.

²² BOSTROM, Nick. Op. cit., p. 31.

²³ SCHARTZ, Jacob T. Limits of Artificial Intelligence. In: SHAPIRO, Stuart; ECKROTH, David. *Encyclopedia of Artificial Intelligence*. New York: Wiley, p. 488-503, 1987.

Por exemplo, em sua supracitada obra seminal, Alan Turing introduz o conceito de aprendizado como um ponto essencial à inteligência: uma máquina teria que possuir a capacidade de aprender e, conseqüentemente, de se auto aprimorar, para que pudesse ser considerada inteligente. No processo de tentar imitar uma mente humana adulta, deve-se refletir sobre o processo que a levou ao estado em que se encontra, podendo ser destacados três componentes: o estado inicial da mente (como no nascimento), a educação a que foi submetida, e outras experiências, não necessariamente enquadradas como educacionais, às quais foi submetida²⁴. Com isso em mente, o autor propõe o conceito de “máquina-criança”:

Em vez de tentar produzir um programa para simular a mente do adulto, por que não tentar produzir um programa que simule a mente de uma criança? Se este fosse submetido a um curso apropriado de educação, seria, então, possível obter o cérebro adulto (tradução nossa).²⁵

E aponta, ainda, a possibilidade de comportamentos não antecipados por parte da máquina, afastando-a do paradigma tradicional elaborado por Ada Lovelace de que uma máquina só poderia fazer aquilo que seus programadores soubessem como programar²⁶:

Uma característica importante de uma máquina capaz de aprender é que seu professor [programador] frequentemente desconhece muito do que se passa lá dentro, embora ele ainda possa, até certa medida, prever o comportamento de seu aluno (...) Isso contrasta claramente com o procedimento normal de utilização de uma máquina para realizar cálculos: o objetivo, então, é ter uma imagem mental clara do estado da máquina em cada momento da computação (tradução nossa).²⁷

Afirmar que algo é aprendido implica mais do que a simples captura e armazenamento de dados. Pelo contrário, esses dados, que podem assumir uma variedade infinita de formas, consistindo em praticamente qualquer coisa que possa ser capturada, quantificada ou representada em formato digital ou analógico, devem ser representados de alguma maneira que permita a sua utilização²⁸.

Em linhas gerais, tem-se que programas de computador capazes de aprender são, na verdade, capazes de extrair padrões estatísticos e correlações ocultas em seus (volumosos)

²⁴ TURING, Alan. Op. cit., p. 455.

²⁵ Idem. No original: “Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain”.

²⁶ LOVELACE apud TURING, Alan. Id., p. 454.

²⁷ Na versão original: “An important feature of a learning machine is that its teacher will often be very largely ignorant of quite what is going on inside, although he may still be able to some extent to predict his pupil's behavior. (...) This is in clear contrast with normal procedure when using a machine to do computations one's object is then to have a clear mental picture of the state of the machine at each moment in the computation” (Id., p. 456).

²⁸ KAPLAN, Jerry. Op. cit., p. 27-28.

dados de entrada. Essa nova abordagem da inteligência artificial, centrada na análise de dados, recebeu o nome de aprendizado de máquina (*machine learning*), que pode ser definido, em linhas gerais, como a ciência de fazer com que computadores desempenhem determinada tarefa sem serem explicitamente programados²⁹.

O aprendizado em questão pode ser supervisionado, quando a máquina é treinada a partir de dados previamente rotulados, ou não supervisionado, quando a máquina aprende por si só a agrupar e estruturar dados³⁰. No primeiro caso, por exemplo, para treinar uma IA a identificar um cão, ela é alimentada por milhares de fotos já classificadas corretamente, ou seja, indicando-se aquelas que contêm ou não cachorros na imagem³¹. Em ambientes dinâmicos, são necessárias várias interações iniciais de treinamento do sistema por um humano com domínio naquele contexto específico até que ele consiga gerar resultados mais precisos e minimamente satisfatórios³².

Entretanto, uma característica notável do aprendizado de máquina diz respeito ao fato de que é também possível alcançar o mesmo objetivo apenas apresentando ao sistema imagens que contenham cães, sem qualquer informação adicional (aprendizado não supervisionado).

Isto porque tais imagens possuem padrões - faces caninas, patas, rabos e assim por diante, em uma infindável variedade de poses, cores e ângulos. O que a inteligência artificial realmente detecta são correlações incrivelmente sofisticadas e complexas entre as imagens, na maior parte das vezes imperceptíveis ao olho humano. Após treinar com milhões de imagens, ela desenvolve a capacidade de detectar padrões semelhantes em imagens não apresentadas anteriormente, aprendendo, em outras palavras, a identificar imagens de cães sozinha³³.

Existem, ainda, abordagens que representam um meio termo entre o aprendizado supervisionado e o não supervisionado. No aprendizado semi supervisionado, por exemplo, as classificações prévias dos dados são mínimas, enquanto que no aprendizado por reforço o

²⁹ SAMUEL, Arthur. L. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, v. 3, n. 3, p. 210-229, 1959.

³⁰ LOUKIDES, Mike; LORICA, Ben. *What is Artificial Intelligence?* Sebastopol: O'Reilly Media, Inc, 2016, p. 4-5.

³¹ Para a aplicação prática do treinamento de um algoritmo de *machine learning* em distinguir lobos de cães da raça Husky Siberiano, veja-se: SINGH, Sameer; RIBEIRO, Marco Túlio; GUESTIN, Carlos. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, 2016. Disponível em: <https://arxiv.org/pdf/1602.04938.pdf>. Acesso em: 30 out. 2019.

³² GUTIERREZ, Andrei. É Possível Confiar em um Sistema de Inteligência Artificial? Práticas em torno da melhoria da sua confiança, segurança e evidências de accountability. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 83-98.

³³ KAPLAN, Jerry. Op. cit., p. 30.

sistema recebe algum valor representativo de uma recompensa. Estas, por vez, podem ser devolvidas ao sistema e usadas para maximizar a probabilidade de sucesso³⁴.

É importante notar que, embora os princípios básicos por trás do aprendizado de máquina não sejam necessariamente recentes, algumas alterações ocorridas nos últimos anos fizeram com que o campo de pesquisa experimentasse uma espécie de renascimento, estando por trás de muitas das tecnologias de inteligência artificial hoje em destaque.

Por exemplo, no que diz respeito às redes neurais, embora modelos simples das mesmas existam desde o final dos anos 1950, a introdução do algoritmo de retropropagação (*backpropagation*) criou a possibilidade do treinamento multicamadas. Essas redes multicamadas, ou seja, que possuem uma ou mais camadas intermediárias (“ocultas”) de neurônios (“nós”) entre aquelas de entrada e saída de dados podem aprender uma gama muito mais variada de funções do que suas antecessoras mais simples. De fato, a área denominada “aprendizado profundo” (*deep learning*), por vezes utilizada como sinônimo de aprendizado de máquina, refere-se ao uso de redes neurais artificiais que possuem um elevado número de camadas ocultas.

Note-se, ainda, que o termo aprendizado de máquina abrange uma ampla variedade de algoritmos³⁵ para além das redes neurais, como árvores de decisão, algoritmos de regressão, SVM (*support vector machine*) ou *ensemble methods*³⁶. Eles se distinguem pelo tempo de processamento e espaço de memória, pelos vieses indutivos que pressupõem, pela facilidade com a qual conseguem incorporar conteúdo produzido externamente e por quão transparente o seu funcionamento interno se revela para um analista humano.³⁷

Assim, segundo Nick Bostrom, por trás do alvoroço envolvendo as diversas áreas de aprendizado de máquina,

encontra-se uma série de escolhas com perdas e ganhos matematicamente bem definidos. O ideal em questão é o do agente bayesiano perfeito, que faz uso probabilisticamente ótimo das informações disponíveis. Esse ideal é inalcançável, pois tem uma demanda computacional muito grande para ser implementado em qualquer computador físico. Assim, é possível ver a inteligência artificial como uma procura por atalhos: meios de se aproximar do ideal bayesiano sacrificando

³⁴ LOUKIDES, Mike; LORICA, Ben. Op. cit., p. 5.

³⁵ Em termos gerais, um algoritmo é qualquer processo que possa ser realizado automaticamente (EDWARDS, Lilian; VEALE, Michael. Slave to the Algorithm? Why a ‘right to an explanation’ is probably not the remedy you are looking for. *Duke Law & Technology Review*, v. 16, n. 01, p. 18-84, 2016. Disponível em: <https://ssrn.com/abstract=2972855>. Acesso em: 26 ago. 2019).

³⁶ DIOP, Lamine; CUPE, Jean. Explainable AI: The data scientists’ new challenge. *Towards Data Science*, jun. 2018. Disponível em: <https://towardsdatascience.com/explainable-ai-the-data-scientists-new-challenge-f7cac935a5b4>. Acesso em: 30 out. 2019.

³⁷ BOSTROM, Nick. Op. cit., p. 35.

*alguma otimização ou generalização, enquanto se preserva o suficiente para atingir uma alta performance nos reais domínios de interesse (grifo nosso).*³⁸

Não se deve inferir, dos fatos narrados anteriormente, que o aprendizado de máquina substituiu por completo o paradigma lógico tradicional (comumente chamado, agora, de *good old fashioned artificial intelligence*), concentrado na manipulação de símbolos de alto nível. Na verdade, o raciocínio símbolo ainda se faz bastante presente no subcampo da inteligência artificial denominado planejamento automático, cujo foco é o desenvolvimento de técnicas para a resolução de problemas que envolvam a formulação de uma série de etapas para atingir um objetivo desejado. Exemplos incluem: fornecer instruções de direção, empacotar caixas de tamanhos variados em um caminhão, provar teoremas matemáticos ou analisar contratos e regulamentos legais.

Em geral, o raciocínio simbólico é mais apropriado para problemas que exigem raciocínio abstrato, mas, conforme abordado, o conhecimento e a experiência humanas podem, muitas vezes, assumir formas que resistem à codificação na linguagem humana ou em qualquer forma explicitamente simbólica (basta imaginar o quão seria difícil construir um robô que pudesse andar de bicicleta com base apenas em testemunhos de especialistas).

Por outro lado, o aprendizado de máquina é mais adequado para situações que exigem percepção sensorial ou a extração de padrões de grande volume de dados, sendo inútil, por óbvio, em situações nas quais não se possui um número suficiente de dados acerca do problema, apenas algumas condições iniciais, várias restrições e uma única tentativa de acerto³⁹.

A inteligência artificial é utilizada, nos dias atuais, em mais áreas do que faria sentido revisar neste trabalho, mas a menção a algumas delas pode ser útil para ilustrar a abrangência de suas aplicações. Por exemplo, no campo médico, existem sistemas de apoio a profissionais que os auxiliam a diagnosticar doenças como o câncer de mama e a recomendar planos de tratamento⁴⁰.

³⁸ Id. p. 35.

³⁹ KAPLAN, Jerry. Op. cit., p. 37.

⁴⁰ Ver: CONNER-SIMONS, Adam; GORDON, Rachel. Using AI to predict breast cancer and personalize care. *MIT News*, mai. 2019. Disponível em: <http://news.mit.edu/2019/using-ai-predict-breast-cancer-and-personalize-care-0507>. Acesso em: 30 out. 2019; FORSTER, Victoria. AI Designs Radiation Therapy Treatment Plan For Cancer In Twenty Minutes. *Forbes*, ago. 2019. Disponível em: <https://www.forbes.com/sites/victoriaforster/2018/08/01/ai-can-design-radiation-therapy-treatment-plan-for-cancer-in-twenty-minutes/#2f13b8497a8c>. Acesso em: 30 out. 2019.

Nos últimos anos, programas de reconhecimento facial passaram a ser utilizados na automatização do controle de fronteiras e para o processamento de vistos⁴¹. Organizações militares e de inteligência, sobretudo norte-americanas, têm implementado em larga escala robôs desarmadores de bombas, drones de vigilância e ataque e outros veículos não tripulados⁴².

Ainda, programas de computador que fazem uso de inteligência artificial são responsáveis por aprovar ou reprovar automaticamente transações realizadas com cartões de crédito, monitorando de maneira contínua atividades bancárias em busca de sinais de fraudes. No mercado financeiro global, sistemas de negociação de ações automatizados são amplamente usados por grandes corretoras de investimentos. Enquanto alguns deles são simplesmente utilizados para automatizar a execução de uma ordem de compra ou venda emitida por um administrador de fundos humano, outros seguem complicadas estratégias de compra e venda que se adaptam às mudanças de condições do mercado, ou se especializam em negociações de alta frequência (*high frequency trading*) que buscam lucrar com pequenos movimentos dos preços que ocorrem em questão de milissegundos⁴³.

Assistentes digitais, como Siri e Alexa, desenvolvidas pela Apple e Amazon, respectivamente, reagem a comandos falados e são capazes de responder a perguntas simples e executar comandos de seus usuários, e sistemas de recomendação são utilizados para recomendar livros, filmes ou álbuns de música a usuários com base em suas compras e avaliações prévias⁴⁴.

Softwares como o *COMPAS* ou *Loomis* são utilizados em alguns estados norte-americanos para auxiliar o judiciário na análise da possibilidade de reincidência criminal, para fins de liberar indivíduos encarcerados mediante fiança,⁴⁵ e outros, como o *PredPol*, foram adotados por algumas forças policiais para antecipar, prevenir e responder mais efetivamente

⁴¹ Veja-se: KENDRICK, Molly. The border guards you can't win over with a smile. *BBC Future*, abr. 2019. Disponível em: <https://www.bbc.com/future/article/20190416-the-ai-border-guards-you-cant-reason-with>. Acesso em: 30 out. 2019.

⁴² LESTER, Caroline. What Happens When Your Bomb-Defusing Robot Becomes a Weapon. *The Atlantic*, abr. 2018. Disponível em: <https://www.theatlantic.com/technology/archive/2018/04/what-happens-when-your-bomb-defusing-robot-becomes-a-weapon/558758/>. Acesso em: 30 out. 2019.

⁴³ BOSTROM, Nick. Op. cit., p. 45.

⁴⁴ Em 2014, um terço de todas as vendas da Amazon resultaram de seus sistemas de recomendação e personalização, e para a Netflix, três quartos de seus novos pedidos vieram de recomendações (MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. *Big Data: A revolution that will transform how we live, work and think*. New York: Eamon Dolan/Mariner Books, 2014, p. 34).

⁴⁵ Para uma análise acerca dos vieses discriminatórios existentes em tais softwares, veja-se: ANGWIN, Julia; LARSON, Jeff; MATTU, Surya; KIRCHNE, Lauren. Machine Bias. *ProPublica*, mai. 2018. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Acesso em: 26 ago. 2019.

a crimes⁴⁶.

No que tange ao processamento de linguagem natural, traduções feitas por máquinas, embora ainda imperfeitas, são boas o bastante em inúmeras aplicações. Aqui, é interessante notar como os primeiros sistemas utilizaram a abordagem tradicional de gramáticas manualmente programadas que foram desenvolvidas do zero por linguistas especializados para cada idioma específico, enquanto que os sistemas atuais utilizam técnicas estatísticas de aprendizado de máquina que automaticamente constroem modelos a partir de padrões de uso observados. Assim, a máquina infere os parâmetros para esses modelos analisando *corpora* bilíngues, dispensando-se a necessidade dos linguistas: os próprios programadores responsáveis pela sua construção não precisam nem mesmo falar a língua com a qual estão trabalhando⁴⁷.

Verifica-se, desse modo, que o campo da inteligência artificial abarca uma ampla variedade de técnicas, que podem ser implementadas de diversas formas – através, por exemplo, do acoplamento de um mecanismo computacional a sensores e atuadores físicos, isto é, um robô –, e aplicadas a setores cada vez mais diversificados.

Seu conceito se aplica a sistemas que apresentam um comportamento inteligente, analisando o seu ambiente e tomando medidas – com um determinado nível de autonomia – para atingir objetivos específicos. Esses sistemas baseados em IA podem ser puramente confinados ao *software*, atuando no mundo virtual, ou podem ser integrados em dispositivos físicos⁴⁸.

Similarmente, e afastando-se de concepções que possuem como medida de sucesso de determinada ação a sua fidelidade ao desempenho humano, Poole e Mackworth definem a inteligência artificial como o campo que estuda a síntese e análise de agentes computacionais que agem de maneira inteligente. O enfoque na noção de agência, em oposição à de simulação de processos cognitivos, possui a vantagem do reconhecimento de que processos adequados de inferência constituem apenas um dos mecanismos para se alcançar a racionalidade, não se

⁴⁶ PERRY, Walter L; MCINNIS, Brian; PRICE, Carter C.; SMITH, Susan C.; HOLLYWOOD, John S. Predictive Policing: The role of crime forecasting in law enforcement operations. *RAND*, 2013. Disponível em: https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf. Acesso em: 26 ago. 2019.

⁴⁷ BOSTROM, Nick. Op. cit., p. 43.

⁴⁸ Definição adotada em UNIÃO EUROPEIA. *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Artificial Intelligence for Europe* (COM/2018/237). Bruxelas: Comissão Europeia, 2018. Disponível em: <<https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>>. Acesso em: 29 out. 2019.

limitando, assim, ao paradigma lógico tradicional - que exigia a formulação de conceitos informais nos termos formais necessários à notação lógica.⁴⁹

Um agente computacional é aquele cujas decisões sobre suas ações podem ser explicadas em termos de computação – significa dizer que a decisão pode ser dividida em operações primitivas que podem ser implementadas em um dispositivo físico. Já um agente inteligente pode ser classificado como tal quando, dadas suas limitações perceptivas e computacionais, faz o que é apropriado para suas circunstâncias e objetivos, adaptando-se a novos ambientes e aprendendo com a experiência⁵⁰ – características que serão melhor desdobradas no próximo tópico.

O objetivo central da IA, portanto, é compreender os princípios que tornam possível o comportamento inteligente em sistemas naturais ou artificiais, através da análise de agentes, naturais e artificiais, que assim se comportam, e da pesquisa e desenvolvimento de sistemas computacionais capazes de executar tais tarefas comumente vistas como exigindo inteligência⁵¹.

Além disso, inteligência artificial é um conceito que comporta duas acepções distintas – inteligência artificial forte (*strong AI*) e inteligência artificial fraca (*weak AI*). Em linhas gerais, a primeira diz respeito à crença de que seria possível, em algum momento futuro, replicar a consciência humana em qualquer sistema físico que tivesse um programa correto (isto é, com *inputs* e *outputs* corretos). O segundo entendimento, por vez, parte do pressuposto que programas computacionais são ferramentas úteis para o estudo da mente humana e o desenvolvimento de tecnologias, todavia, um programa de inteligência artificial é, no máximo, a simulação de um processo cognitivo – e não um processo cognitivo por si só⁵².

Esses termos são muitas vezes mal utilizados para descrever a distinção entre sistemas que exibem comportamento inteligente geral (*general purpose AI*) e aqueles que são limitados a um domínio restrito de aplicações (*narrow AI*), não possuindo qualquer relação com a questão da consciência⁵³. Por “inteligência geral” não se quer dizer, necessariamente, inteligência humana, mas sim tecnologias capazes de solucionar diferentes tipos de problemas sem serem programadas com conhecimento específico de um único domínio⁵⁴.

⁴⁹ RUSSELL, Stuart J.; NORVIG, Peter. Op. cit., p. 4.

⁵⁰ POOLE, David; MACKWORTH, Alan. Artificial Intelligence: Foundations of Computational Agents. Cambridge (UK): Cambridge University Press, 2010, p. 4.

⁵¹ Id., p. 4.

⁵² SEARLE, John. *Mente, Cérebro e Ciência*. Tradução de Artur Morão. Lisboa: Biblioteca de Filosofia Contemporânea, 2000, p. 36.

⁵³ KAPLAN, Jerry. Op. cit., p. 68.

⁵⁴ LOUKIDES, Mike. Op. cit., p. 2.

Meredith Broussard resume bem a questão ao destacar que o conceito de inteligência artificial geral está contaminado pelo senso comum e pela ficção, podendo ser compreendido como a IA que se gostaria de ter disponível⁵⁵. Já a inteligência artificial estrita se refere aos programas de computador de que atualmente se dispõe, quais sejam, sistemas altamente especializados e de propósito único que podem desempenhar determinada função excepcionalmente bem – como os populares *Deep Blue* e *AlphaGo*.

Finalmente, cumpre ressaltar que, embora existam discussões relativas à criação de uma inteligência artificial forte, e a consequente necessidade de a ela se atribuir um status moral, reconhecendo-lhe direitos e deveres a princípio somente aplicáveis à pessoa natural⁵⁶, uma parcela considerável de especialistas tanto em inteligência artificial como em inteligência humana admitem se tratar, até o momento, apenas de uma possibilidade remota⁵⁷.

Do mesmo modo, são minoritárias as opiniões especializadas demonstrando preocupação com a criação, em um futuro próximo, de uma inteligência artificial geral capaz de se aperfeiçoar continuamente, desembocando em uma superinteligência que poderia subjugar, intencional ou acidentalmente, os seres humanos⁵⁸.

Justamente por isso, apenas a inteligência artificial em suas acepções fraca e restrita estão abarcadas pelo escopo da presente pesquisa. Mais especificamente, tratar-se-á de agentes de inteligência artificial capazes de atuação autônoma, assunto tratado a seguir.

2. Agentes artificiais autônomos

⁵⁵ BROUSSARD, Meredith. *Artificial Intelligence: How computers misunderstand the world*. Cambridge: MIT Press, 2018.

⁵⁶ Veja-se: COECKELBERGH, Mark. Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, v. 12, n. 3, p. 209-221, 2010; TEUBNER, Gunther. Rights of non-humans? Electronic agents and animals as new actors in politics and law. *Journal of Law and Society*, v. 33, n. 4, p. 497-521, 2006.

⁵⁷ Nesse sentido: KAPLAN, Jerry. *Artificial Intelligence: What everyone needs to know*. New York: Oxford University Press, 2016; NICOLELIS, Miguel A. L.; CICUREL, Ronald. *O Cérebro Relativístico: Como ele funciona e por que ele não pode ser simulado por uma máquina de Turing*. São Paulo: Kios Press, 2015; SEARLE, John. *Mente, Cérebro e Ciência*. [Trad.] Artur Morão. Lisboa: Biblioteca de Filosofia Contemporânea, 2000; KELLY, Kevin. The Mith of a Superhuman AI. *Wired*, 2017. Disponível em: <https://www.wired.com/2017/04/the-myth-of-a-superhuman-ai/>. Acesso em: 06 abr. 2020. PINKER, Steven. *O Novo Iluminismo: Em defesa da razão, da ciência e do humanismo*. São Paulo: Companhia das Letras, 2018; PEARL, Judea; MACKENZIE, Dana. *The Book of Why: The new science of cause and effect*. New York: Basic Books, 1. Ed., 2018.

⁵⁸ KURZWEIL, Ray. *The Singularity is Near: When humans transcend biology*. Nova York: Penguin Books, 2005; BOSTROM, Nick. *Superinteligência: Caminhos, perigos e estratégias para um novo mundo*. Rio de Janeiro: DarkSide, 2014; HAWKING, Stephen. *Brief Answers to Big Questions*. Nova York: Bantam Books Publishing, 2018. PIPER, Kelsey. Why Elon Musk Fears Artificial Intelligence. *Vox*, 2018. Disponível em: <https://www.vox.com/future-perfect/2018/11/2/18053418/elon-musk-artificial-intelligence-google-deepmind-openai>. Acesso em: 06 abr. 2020.

Nas interações sociais e econômicas de hoje, é possível notar cada vez mais a presença ubíqua de uma nova categoria de ser: o agente artificial. Responsáveis pela compra e venda de mercadorias, por determinar a elegibilidade para benefícios como a obtenção de crédito ou assistência médica, pelo processamento de informações financeiras e negociação nas bolsas de valores, tais agentes artificiais representam uma continuação lógica da adoção de tecnologias de automação e aprimoramento de capacidades e habilidades humanas.

A crescente frequência de interação com esses agentes artificiais, em ambientes não supervisionados e sem a presença de mediadores humanos, torna fundamental a compreensão do papel por eles desempenhados no complexo de relações sociais, que perpassam desde questões filosóficas acerca de seu estatuto moral, até problemas jurídicos, como a sua qualificação como atores legais e a imputação de responsabilidade pelos eventuais danos que venham a causar.

O termo *agente* abarca uma pluralidade de significados, que variam de acordo com a área em que é aplicado. Na vida cotidiana, o termo possui um significado intuitivo, definindo-se um agente como algo que possui a capacidade de executar ações. Assim, agentes seriam distintos de outras entidades na medida em que realizam ações, ao invés de serem apenas receptáculos de fenômenos e atuações alheias. Negar algo ou alguém agência é negar a capacidade de agir, pois as ações de um agente o distinguem do restante do mundo⁵⁹.

Agentes também podem ser considerados como algo que age em nome de alguém. Para a Economia, agência significa qualquer relacionamento em que uma pessoa contrata outra para executar um serviço em circunstâncias que envolvam delegar alguma discricção sobre a tomada de decisão ao prestador de serviços⁶⁰.

Similarmente, nos países adeptos ao sistema de *common law*, define-se a agência como a relação jurídica que emerge quando uma pessoa (denominada “*principal*”) manifesta consentimento para que outra pessoa, seu agente (“*agente*”), atue em seu lugar e sujeita a seu controle, podendo, inclusive, pleitear direitos e criar obrigações em seu nome⁶¹. Trata-se de um agente em sentido estrito – em oposição à utilização também frequente do termo para denominar, de maneira geral, entes que possuem direitos e deveres próprios, ou seja, pessoas⁶².

⁵⁹ CHOPRA, Samir; LAWRENCE, White. *A Legal Theory for Artificial Autonomous Agents*. Ann Arbor: University of Michigan Press, 2011, p. 11.

⁶⁰ DEMOTT, Deborah A. Disloyal agents. *Alabama Law Review*, v. 58, p. 1049–1067, 2007.

⁶¹ ESTADOS UNIDOS. *Restatement (Third) of Agency*. 1933.

⁶² PAGALLO, Ugo. *The laws of robots: crimes, contracts, and torts*. Heidelberg: Springer, 2013. p. 40.

Distingue-se, portanto, do entendimento filosófico acerca da agência, que se preocupa em determinar ou caracterizar a autonomia de um ator, e não as consequências do seu relacionamento com outros atores⁶³. Nesse sentido, possuir agência significa ser o originador de uma ação, ser movido por motivações, propósitos e desejos, relacionando-se, portanto, com a atribuição de intenções, pelo conceito de ação autodirigida ou “agir por razões”⁶⁴. Nas palavras de Taylor, o que é crucial sobre agentes é o fato de que as coisas importam para eles, o que significa dizer que é possível atribuir propósitos, desejos, aversões a elas, em um forte e original sentido⁶⁵.

Por sua vez, na literatura especializada de ciência da computação, o termo agente representa um conjunto amplo de tecnologias, notadamente aquelas cujos sistemas de processamento de informações são relativamente autônomos. Pode-se aplicá-lo, portanto, a uma variedade de situações em que programas semiautônomos, com diferentes arquiteturas e graus de sofisticação, são utilizados para executar tarefas humanas⁶⁶.

Por exemplo, pode-se encontrar a definição de agente como um software que atua em nome de seu usuário e busca atingir determinados objetivos ou concluir tarefas sem nenhum input ou supervisão diretos deste usuário⁶⁷; ou ainda como sistemas computacionais que habitam um ambiente dinâmico complexo, detectam e agem de forma autônoma nesse ambiente e, ao fazê-lo, realizam um conjunto de objetivos ou tarefas para os quais foram projetados⁶⁸.

Para Russell e Norvig, pode-se definir um agente como “qualquer coisa que possa ser vista como percebendo seu ambiente por meio de sensores e atuando nesse ambiente através de atuadores” (tradução nossa)⁶⁹. Note-se que, segundo a definição por eles proposta, é possível abarcar dentro do conceito uma ampla variedade de entes – humanos, robóticos ou de software. Um agente humano possui olhos, ouvidos e outros órgãos como sensores, e mãos, pernas e o trato vocal como atuadores. Um agente robótico pode possuir câmeras e sensores

⁶³ DEMOTT, Deborah A. Op. cit., p. 1051.

⁶⁴ CHOPRA, Samir; WHITE, Laurence F. Op. cit., p. 11.

⁶⁵ TAYLOR, Charles. *Human Agency and Language (Philosophical Papers v. 1)*. Cambridge University Press: Cambridge, 1985, p. 98.

⁶⁶ CHOPRA, Samir; WHITE, Laurence F. Op. cit., p. 6.

⁶⁷ BORKING, John; B. M. A. van Eck; SIEPEL, P. Intelligent software agents: Turning a privacy threat into a privacy protector. *Information and Privacy Commissioner Ontario (Canada)*, 1999. Disponível em: https://www.researchgate.net/publication/237219149_INTELLIGENT_SOFTWARE_AGENTS_AND_PRIVACY. Acesso em: 30 out. 2019.

⁶⁸ MAES, Pattie. Artificial life meets entertainment: Lifelike autonomous agents. *Communications of the ACM* 38 (11), 1995. Disponível em: <https://dl.acm.org/citation.cfm?id=219808>. Acesso em: 30 out. 2019.

⁶⁹ Na versão original dos autores: “an agent is anything that can be viewed as perceiving its environment through sensors and acting on that environment through actuators” (RUSSELL, Stuart J.; NORVIG, Peter. Op. cit., p. 34).

de alcance infravermelho, e variados motores como atuadores. Finalmente, um agente de software recebe conteúdos de arquivos ou o pressionamento de teclas como inputs sensoriais, e atua no ambiente exibindo algo em uma tela, escrevendo arquivos ou enviando pacotes de rede⁷⁰.

Assim, o ambiente de um agente pode variar desde o mundo real (no caso de um humano ou robô), até ambientes simulados no ciberespaço ou ambientes informacionais consistentes em bases de dados, aplicações, serviços e redes (no caso de um agente de software). Cada um desses ambientes pode, por óbvio, incluir outros agentes, bem como ser modificado à medida em que o agente em questão prossegue com suas tarefas. Já a adaptabilidade de um agente pode ser avaliada pela sua capacidade de funcionar em ambientes especificados imprecisa ou incompletamente, de maneira dinâmica e sensível ao contexto em questão.⁷¹

Precisamente por esta razão, pode-se dizer que um agente possui maior ou menor autonomia a depender de sua capacidade de aprendizado e planejamento. Um agente capaz de aprender é capaz de modificar a si mesmo, seu comportamento e suas respostas diante de suas experiências com o meio ambiente. Desse modo, conforme ele dependa do conhecimento prévio de seu programador – no caso de um agente artificial - e não se baseie em suas próprias percepções, é possível afirmar que lhe falta autonomia⁷².

Agentes capazes de planejamento, por vez, podem construir sequências de ações para realizar tarefas, utilizando-se de uma ampla variedade de técnicas algorítmicas para lidar com a incerteza em seus ambientes. Possuem, portanto, uma natureza direcionada ao alcance de objetivos (*goal-directed*): um resultado final pode ser especificado e, dado o conhecimento do agente sobre as ações necessárias para alcançá-lo, ele poderá realizá-las considerando os recursos existentes e as características ambientais, selecionando dentre as diversas opções disponíveis de forma a maximizar sua performance e obter o melhor resultado possível⁷³.

Tal descrição enquadra-se na definição proposta por Russell e Norvig de um agente racional: aquele que atua de modo a alcançar o melhor resultado ou, havendo incerteza, o melhor resultado esperado. Para os autores, enquanto todos os programas de computador podem ser considerados como agentes no sentido ordinário do termo, vez que por óbvio eles realizam ações, espera-se que um agente de computador faça mais do que isso: que ele opere

⁷⁰ Id. p. 34.

⁷¹ CHOPRA, Samir; WHITE, Laurence F. Op. cit., p. 8-9.

⁷² RUSSELL, Stuart J.; NORVIG, Peter. Op. cit., p. 39.

⁷³ CHOPRA, Samir; WHITE, Laurence F. Op. cit., p. 9.

de forma autônoma, perceba o seu ambiente, persista por um período prolongado, adapte-se a mudanças e busque seus objetivos⁷⁴.

Quando um agente é colocado em determinado ambiente, ele irá gerar uma sequência de ações conforme as percepções que irá receber. Essa sequência de ações faz com que o ambiente passe, também, por uma sequência variada de estados. Assim, se a sequência é desejável, pode-se afirmar que o agente teve um bom desempenho, e essa noção de conveniência é capturada por uma medida de desempenho que avalia qualquer sequência de estados do ambiente.

A racionalidade de um agente dependerá, portanto: da medida de desempenho que define o critério de sucesso; do seu conhecimento prévio sobre o meio ambiente em que se encontra; das ações que ele pretende executar; e da sua sequência de percepção até o momento (que é, em resumo, a história completa de tudo aquilo que o agente já percebeu através de seus sensores). Desse modo, para cada sequência de percepção possível, um agente, para que seja considerado racional, deverá selecionar determinada ação que maximize a sua medida de desempenho, dadas todas as suas percepções até aquele instante, somadas a qualquer conhecimento prévio embutido no agente⁷⁵.

Nesse sentido, Poole e MackWorth ressaltam que dois agentes determinísticos com o mesmo conhecimento prévio, história, habilidades e objetivos deverão, necessariamente, realizar a mesma ação; conseqüentemente, a alteração de qualquer uma dessas variáveis poderá resultar em ações diferentes⁷⁶.

Abrangendo todo o exposto anteriormente, Chopra e White classificam, em síntese, um agente artificial como autônomo na medida em que este possui um grau relativamente elevado de um ou mais dos seguintes requisitos: (i) capacidade de operar sem a intervenção direta de seres humanos ou outros agentes, e de exercer controle não supervisionado sobre suas próprias ações e estados internos; (ii) capacidade social ou capacidade de interagir com outros agentes artificiais ou seres humanos; (iii) capacidade proativa de iniciar um comportamento direcionado a objetivos; (iv) capacidade reativa de perceber um ambiente e responder a mudanças nele; (v) capacidade de se ajustar aos hábitos, métodos de trabalho e preferências de usuários, outros agentes ou seres humanos; (vi) capacidade de se mover em

⁷⁴ RUSSELL, Stuart J.; NORVIG, Peter. Op. cit., p. 4.

⁷⁵ Id., p. 37.

⁷⁶ POOLE, David; MACKWORTH, Alan. Op. cit., p. 4.

um ambiente virtual ou físico; (vii) representatividade, ou o atributo de ser um representante ou intermediário de outro agente ou pessoa⁷⁷.

Seguindo esta linha, e tendo em vista os fins almejados pela presente pesquisa, optou-se por evidenciar os aspectos relativos à capacidade de aprendizado, adaptabilidade e ausência de input ou supervisão direta de um ser humano sobre suas ações. Assim, um agente artificial autônomo será considerado como tal se agir independentemente de instrução humana direta, com base nas informações que ele próprio adquire e analisa, sendo capaz de direcionar suas próprias atividades em um ambiente imprevisível ou em constante modificação.

É importante ressaltar que essa definição de autonomia pode ser aplicada tanto a sistemas que interagem no mundo real, ou seja, em um ambiente físico (como um carro autônomo), quanto àqueles que atuam em um ambiente virtual. É o exemplo dos já mencionados algoritmos de negociação de alta frequência, popularmente denominados “robôs investidores”, que interagem com informações de maneira autodirigida, sem necessariamente produzir ações no mundo físico.

Além disso, pode-se contrastar o termo “autônomo” com outro semelhante: “automático”. Este normalmente se refere à execução de uma sequência pré-programada de ações - um braço robótico utilizado em uma linha industrial de montagem, por exemplo, pode ser considerado automático. “Autônomo”, por vez, sugere a capacidade de reagir independentemente em resposta aos seus inputs sensoriais, fazendo escolhas que, de outra modo, seriam feitas por uma pessoa em um sistema não-autônomo (dirigido por humanos). É o caso de um aspirador de pó robótico que, ainda que de modo simplificado, determinará seu curso de ação com base na presença ou ausência de obstáculos em seu caminho.⁷⁸

Nesse sentido, a autonomização tecnológica enquadra-se naquilo que Klaus Schwab denominou de Quarta Revolução Industrial, ou revolução tecnológica, que tem como marco inicial a virada deste século. Tendo em vista que o termo Revolução Industrial designa o conjunto de alterações drásticas nos métodos de produção, com a constituição de novas infraestruturas e aumento de produtividade em um curto espaço de tempo⁷⁹, o autor considera que, ao longo da história, existiram três grandes marcos na indústria.

As duas primeiras revoluções industriais partiram de sistemas físicos – a invenção da máquina a vapor e o desenvolvimento da eletricidade, respectivamente –, e a terceira chegou aos sistemas cibernéticos com o uso da informática. Finalmente, a quarta revolução industrial

⁷⁷ CHOPRA, Samir; WHITE, Laurence F. Op. cit., p. 10.

⁷⁸ WINFIELD, Alan. *Robotics: A very short introduction*. Oxford: Oxford University Press, 2012, p. 12.

⁷⁹ SCHWAB, Klaus. *The Fourth Industrial Revolution*. Danvers: Currency, 2017.

viabilizou o desenvolvimento de sistemas físico cibernéticos (*cyber-physical systems*), combinando, portanto, várias tecnologias que se fundem e permitem a interação entre os domínios físicos, biológicos e digitais. Assim, com o advento de tecnologias de IA, a automação se revestiria de uma nova qualidade: a capacidade de aprendizado, por parte de softwares, para melhoria de seu próprio desempenho⁸⁰.

Engenheiros tradicionalmente classificam os sistemas ao longo de um espectro de autonomia, a depender do grau com o qual eles tomam decisões sobre suas próprias ações. Em uma extremidade do espectro, se um humano está tomando todas as decisões mais importantes, tem-se um sistema com pouca ou nenhuma autonomia. Por outro lado, se um computador estiver tomando todas as decisões mais importantes, pode-se dizer que se trata de um sistema totalmente autônomo. No meio termo, encontram-se os sistemas parciais ou semiautônomos, nos quais algumas ações importantes são decididas por seres humanos e outras por um computador⁸¹.

Atualmente, pode-se dizer que a grande maioria das máquinas disponíveis ao público em geral são, no máximo, semiautônomas. A mão humana define, orienta e, em última instância, controla o processo, seja diretamente ou devido à capacidade de substituir a máquina e assumir o controle. Consistem, assim, em meras ferramentas, ainda que extraordinariamente sofisticadas, usadas pelos seres humanos⁸².

Por exemplo, no campo de veículos aéreos não tripulados (*unmanned aerial vehicles* ou UAVs), drones como o RQ-1 e MQ-1 *Predators*, da Força Aérea norte-americana, devem ser considerados semiautônomos; enquanto outros, como o *Global Hawk* e o sistema de defesa antimísseis *Phalanx CIWS* (este, pertencente à Marinha norte-americana) operam completamente sozinhos, independentemente de *inputs* de controle de um piloto em tempo real⁸³.

Nesse mesmo sentido, veículos totalmente autônomos ainda não estão disponíveis para consumidores, mas já operam como veículos de pesquisa protótipos ou em ambientes

⁸⁰ VERONESE, Alexandre; SILVEIRA, Alessandra; LEMOS, Amanda. Inteligência Artificial, Mercado Único Digital e a Postulação de um Direito às inferências Justas e Razoáveis: Uma questão jurídica entre a ética e a técnica. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). Inteligência Artificial e Direito: Ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, 2019, p. 233-264.

⁸¹ SURDEN, Harry; WILLIAMS, Mary-Anne. Technological Opacity, Predictability, and Self-Driving Cars. *Cardozo Law Review*, v. 38, p. 121-181, 2016.

⁸² VLADECK, David C. Machines Without Principles: Liability Rules and Artificial Intelligence. *Washington Law Review*, v. 89, n. 01, p. 117-150, 2014.

⁸³ PAGALLO, Ugo. Op. cit., p. 3. Para uma discussão aprofundada sobre o uso de armas autônomas, ver: ANDERSON, Kenneth; WAXMAN, Matthew C. Debating Autonomous Weapon Systems, Their Ethics and Their Regulation Under International Law. *Washington College of Law Research Paper No. 2017-21*, p. 1097-1117, 2017.

comerciais controlados, como os caminhões autônomos utilizados para transportar materiais de mineração no *Outback* australiano, e em construções no estado americano da Flórida⁸⁴.

Por outro lado, veículos parcialmente autônomos exibem uma mistura de controle humano e digital, com algumas atividades importantes dirigidas por computador (por exemplo, frenagem de emergência) e outras por seres humanos (por exemplo, frenagem comum, direção, aceleração)⁸⁵.

Devido a algumas barreiras tecnológicas, como a dificuldade de direção em condições climáticas adversas ou em centros urbanos densos, bem como o preço elevado de alguns dos equipamentos necessários ao seu funcionamento, há uma ampla gama de estimativas acerca de quando veículos totalmente autônomos serão vendidos ao público. A maioria dos especialistas prevê que a primeira venda a consumidores ocorrerá entre os anos de 2020 e 2035. No entanto, mesmo que tais veículos comecem a ser comercializados em 2020, conforme essas previsões, é provável que um tempo maior seja necessário até que haja uma inserção substancial no mercado, com carros totalmente autônomos não atingindo uma proporção elevadas dos veículos nas estradas até o ano de 2030 ou posteriormente⁸⁶.

Por mais inovadoras que sejam, porém, as máquinas existentes hoje possuem uma característica comum fundamental na análise da responsabilidade: em cada caso, é possível identificar, por trás de seu agir, as diretivas humanas responsáveis por determinado funcionamento ou tomada de decisão, rastreando-a diretamente de volta à programação do software, design da máquina, ou dados que lhe foram alimentados ao longo do processo de aprendizagem.

Na medida em que agentes artificiais passem a atuar de maneira verdadeiramente autônoma, interagindo ostensivamente com outros agentes, humanos ou não, em ambientes complexos e não supervisionados, será possível que eles executem determinadas ações não antecipadas por seus criadores, muitas delas com consequências danosas.

3. Os problemas decorrentes da autonomia de agentes artificiais

⁸⁴ GRAY, Darren. No one behind the wheel: The new workforce driving Australia's mines. *The Sydney Morning Herald*, abr. 2019. Disponível em:

<https://www.smh.com.au/business/companies/no-one-behind-the-wheel-the-new-workforce-driving-australia-s-mines-20190411-p51dd2.html>. Acesso em: 30 out. 2019; MARSHALL, Aarian. Starsky Robotics Unleashes Its Truly Driverless Trucks in Florida. *WIRED*, set. 2018. Disponível em:

<https://www.wired.com/story/starsky-robotics-truck-self-driving-florida-test/>. Acesso em: 30 out. 2019.

⁸⁵ SURDEN, Harry; WILLIAMS, Mary-Anne. Op. cit., p. 133.

⁸⁶ Id., p. 137.

Agentes autônomos artificiais são, por definição, capazes de atuar sobre o ambiente em que se encontram, percebendo as mudanças que nele ocorrem e respondendo a elas de acordo com seus objetivos pré-definidos, sem, contudo, necessitar de intervenção ou supervisão direta humana para operar. São, portanto, capazes de se adaptar e de aprender a partir de suas experiências e interação com o ambiente e com outros agentes, artificiais ou não.

São precisamente estas características que, atualmente, constituem o principal atrativo para o seu desenvolvimento e emprego nas mais diversas áreas. Entretanto, são também a justificativa por trás das principais preocupações levantadas a respeito do uso de inteligência artificial. Conforme será visto adiante, os ganhos em autonomia e na consequente abdicação de controle por parte de seres humanos trazem consigo riscos associados ao comportamento imprevisível e, frequentemente, ininteligível desses agentes.

3.1. Imprevisibilidade

O debate atual acerca da responsabilidade por danos causados por uma IA, que gira em torno de aplicações concretas semiautônomas, está relacionado com o comportamento errático de uma máquina, isto é, de maneira diversa daquela que havia sido inicialmente projetada por seus programadores. Nesses casos, existe não apenas a possibilidade de identificação de erro humano, mas também a manutenção da capacidade decisória quanto ao uso da IA por um humano⁸⁷.

Enquadram-se aqui os acidentes fatais ocorridos envolvendo carros com funcionalidades autônomas desenvolvidos pelas empresas Uber e Tesla. No primeiro caso, ocorrido em março de 2018 no estado americano do Arizona, um carro autônomo em fase de teste atropelou uma pedestre que atravessava fora da faixa de pedestres junto à sua bicicleta, sem que houvessem sinais de frenagem ao se aproximar⁸⁸. O relatório da *National Transportation Safety Board* (NTSB) divulgou que os sistemas de radar do veículo detectaram a pedestre 5.6 segundos antes do impacto e, embora o sistema tenha continuado a

⁸⁷ MULHOLLAND, Caitlin. Responsabilidade Civil e Processos Decisórios Autônomos em Sistemas de Inteligência Artificial (IA): Autonomia, imputabilidade e responsabilidade. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 325-348.

⁸⁸ LEVIN, Sam; WONG, Julia Carrie. Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian. *The Guardian*, mar. 2018. Disponível em: <https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe>. Acesso em: 30 out. 2019.

rastreá-la até o acidente, ele nunca identificou com precisão o objeto que atravessa a estrada como um pedestre, ou foi capaz de prever o seu caminho.⁸⁹

Entretanto, os promotores do caso decidiram não prestar queixas por considerar que a Uber não era criminalmente responsável pelo ocorrido, e que o acidente poderia ter sido evitado pela operadora de plantão caso a mesma estivesse atenta no momento de sua ocorrência - carros autônomos em fase de testes possuem um motorista humano de *backup*, que deve estar alerta em caso de eventuais falhas. No caso em tela, a operadora estava assistindo a um programa televisivo⁹⁰.

Similarmente, o relatório da NTSB apontou, para além da falha no software, para o fato de que se a operadora do veículo estivesse atenta, provavelmente teria tido tempo suficiente para detectar e reagir à pedestre para evitar o acidente ou mitigar o impacto. Mas não apenas. No relatório, ressalta-se que, não obstante os gerentes da Uber tivessem a capacidade de monitorar retroativamente o comportamento dos operadores de veículos, eles raramente o faziam, e a supervisão ineficaz da empresa foi exacerbada por sua decisão de remover um segundo operador do veículo durante os testes do sistema de direção automatizado⁹¹.

Já os acidentes envolvendo a empresa Tesla, embora não tenham ocorrido com carros propriamente autônomos, aconteceram quando o recurso semiautônomo do piloto automático estava ativado, em 2016 e novamente em 2019, nos Estados Unidos. Embora o relatório final emitido pela NTSB acerca do primeiro acidente tenha reconhecido a culpa por parte do caminhão envolvido na colisão, também a atribuiu parcialmente à Tesla por projetar um sistema que permitia ao motorista depender excessivamente do sistema de automação por um período prolongado⁹².

Verificam-se, assim, questões atinentes ao funcionamento adequado dos sistemas autônomos utilizados nesses veículos, bem como à necessidade de envolvimento humano no processo (*human in the loop*).

Por outro lado, a absoluta independência da interferência humana para alcançar

⁸⁹ ESTADOS UNIDOS. National Transportation Safety Board: Public Meeting of November 19, 2019, Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian, p. 1-5, 2019. Disponível em: <https://www.nts.gov/news/events/Documents/2019-HWY18MH010-BMG-abstract.pdf>. Acesso em: 30 nov. 2019.

⁹⁰ SHEPARDSON, David. NTSB to determine probable cause of fatal Uber self-driving crash. *Reuters*, out. 2019. Disponível em: <https://www.reuters.com/article/us-uber-crash-safety/ntsb-to-determine-probable-cause-of-fatal-uber-self-driving-crash-idUSKBN1WW2IR>. Acesso em: 30 out. 2019.

⁹¹ ESTADOS UNIDOS. National Transportation... Op. cit., p. 3.

⁹² DAVIES, Alex. Tesla's Latest Autopilot Death Looks Just a Prior Crash. *WIRED*, mai. 2019. Disponível em: <https://www.wired.com/story/teslas-latest-autopilot-death-looks-like-prior-crash/>. Acesso em: 30 out. 2019.

resultados apresenta, como consequência, a impossibilidade de antecipação dos efeitos obtidos. Nesse sentido, na medida em que a influência do programador sobre a máquina diminui, a influência do ambiente operacional aumenta. Essencialmente, o programador transfere parte de seu controle sobre o produto para o ambiente, sobretudo em se tratando de máquinas que continuam aprendendo e se adaptando em seu ambiente operacional final. Como nessa situação elas têm que interagir com um número potencialmente grande de pessoas e situações, via de regra não será possível prever ou controlar a influência do ambiente operacional⁹³.

Assim, uma importante característica que agentes artificiais autônomos apresentam e que impõe novos desafios jurídicos refere-se ao conceito de previsibilidade. Conforme aponta Scherer, em inúmeras instâncias nas quais uma inteligência artificial atua de maneira que pode ser considerada criativa (em sentido estrito), tem-se, na verdade, a consequência da combinação entre os recursos computacionais a ela disponíveis e a sua liberdade de preconceitos cognitivos que afetam os seres humanos⁹⁴.

Estes, incapazes de analisar todas ou mesmo a maioria das informações à sua disposição, sobretudo quando confrontados com restrições temporais, muito frequentemente contentam-se com uma solução satisfatória, ao invés da solução ótima para o problema tratado. Por outro lado, o poder computacional dos computadores modernos possibilita a um programa de IA avaliar um número infinitamente superior de possibilidades em um período reduzido de tempo. Podem, portanto, conceber soluções nem ao menos consideradas por seres humanos, ou então rejeitadas em favor de opções mais intuitivamente atrativas.

Em um universo de possibilidades suficientemente compacto – como no jogo de damas – uma inteligência artificial é capaz até mesmo de gerar uma solução ideal em vez de uma alternativa meramente satisfatória. E, mesmo em ambientes mais complexos – como no jogo Go -, a solução encontrada pode desviar de maneira substancial daquela normalmente produzida pelos processos cognitivos humanos – sendo, pois, imprevisível mesmo para quem a projetou⁹⁵.

É justamente essa capacidade de gerar soluções únicas que torna o uso da inteligência artificial tão atrativo em uma variedade cada vez mais ampla de áreas. Na área médica, dois

⁹³ MATTHIAS, Andreas. The responsibility gap: ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, v. 6, n. 3, p. 175-183, 2004..

⁹⁴ SCHERER, Matthew U. Regulating Artificial Intelligence Systems: Risks, challenges, competencies, and strategies. *Harvard Journal of Law and Technology*, v. 29, n. 2, p. 353-400, 2016. Disponível em: <http://jolt.law.harvard.edu/articles/pdf/v29/29HarvJLTech353.pdf>. Acesso em: 29 out. 2019.

⁹⁵ Id., p. 364-365.

exemplos ilustram bem essa hipótese. O primeiro diz respeito ao programa de aprendizado de máquina *C-Path* (sigla derivativa de *Computational Pathologist*), cujo objetivo é analisar imagens de tecidos cancerígenos e prever a sobrevida do paciente. Patologistas suspeitavam que o estudo de componentes do tecido circundante de células cancerígenas (denominado estroma) poderia, em combinação com o estudo das próprias células tumorais, auxiliar no prognóstico do câncer. Entretanto, a análise do programa revelou que as características do estroma eram, na verdade, um indicador prognóstico mais adequado para casos de câncer de mama do que as próprias características das células cancerígenas - uma conclusão que contrariou tanto o senso comum quanto o pensamento médico predominante⁹⁶.

Exemplar ainda da capacidade inovadora de algoritmos de inteligência artificial é o Projeto Artemis (em referência à deusa grega responsável pela proteção de bebês e grávidas). Trata-se de plataforma de análise de dados desenvolvida em conjunto pela IBM e pelo Instituto de Tecnologia da Universidade de Ontário, capaz de detectar infecções em bebês prematuros 24 horas antes do aparecimento dos respectivos sintomas. Surpreendentemente, o programa foi capaz de identificar correlações que iam de encontro à sabedoria convencional dos médicos, como o fato de que sinais vitais constantes frequentemente são detectados antes de uma infecção grave (e não a esperada deterioração dos mesmos)⁹⁷.

É importante destacar que, ainda que as ações desses sistemas tenham sido inesperadas - certamente para observadores externos e talvez até para os seus programadores -, a sua capacidade de produzir ações imprevisíveis, em si, muito frequentemente resulta do planejamento dos seus desenvolvedores. Afinal, existe cada vez mais um considerável incentivo econômico para criar sistemas de IA capazes de gerar soluções inovadoras⁹⁸. Nesse sentido, Ryan Calo relembra que o comportamento emergente é um objetivo claramente declarado da robótica e da inteligência artificial⁹⁹.

A imprevisibilidade repercute também na definição do que exatamente deve ser considerado como uma falha no funcionamento do código de programação (ou *bug*), sendo tênue a linha divisória entre uma ação resultante do funcionamento normal de um agente artificial autônomo, e aquela decorrente de um defeito.

⁹⁶ MYERS, Andrew. Stanford team trains computer to evaluate breast cancer. *Stanford Medicine News Center*, nov. 2011. Disponível em: <https://med.stanford.edu/news/all-news/2011/11/stanford-team-trains-computer-to-evaluate-breast-cancer.html>. Acesso em: 29 out. 2019.

⁹⁷ BEYMAN, Michael. Big Data's powerful effect on tiny babies. *CNN*, set. 2013. Disponível em: <https://www.cnn.com/2013/09/13/big-datas-powerful-effect-on-tiny-babies.html>. Acesso em: 30 out. 2019.

⁹⁸ SCHERER, Matthew. Op. cit., p. 365.

⁹⁹ CALO, Ryan. Robotics and the Lessons of Cyberlaw. *California Law Review*, v. 103, n. 3, p. 513-563, 2015.

Um software altamente complexo, muito provavelmente terá falhas ou produzirá resultados imprevisíveis. Falhas podem ser difíceis de detectar, e se desenvolverem mediante a combinação de múltiplas modificações e adições. Assim, pode ser extremamente difícil atribuir a responsabilidade por falhas que emergem das camadas de desenvolvimento de software por muitas mãos. E, na medida em que robôs e programas de inteligência artificial aprendam a modificar seu próprio código, as questões atinentes à responsabilidade tornar-se-ão ainda mais difusas¹⁰⁰.

Similarmente, Čerka et. al chamam atenção para o fato de que, uma vez que softwares de aprendizado de máquina têm como característica essencial a autoaprendizagem com base em suas experiências e a capacidade de agir autonomamente, seria bastante improvável conseguir identificar se determinada falha existia desde a sua concepção, ou se decorreu do próprio processo de aprendizado - a menos que se equipare a tomada de decisão autônoma, característica distintiva de um agente artificial autônomo, a um defeito¹⁰¹.

Na verdade, como esses algoritmos são feitos para aprender a partir da interação com ambientes externos dinâmicos, seus códigos não determinam *a priori* as correlações e parâmetros de análise, mas antes um padrão de probabilidades a ser preenchido a partir dessa comunicação com o ambiente¹⁰².

Visando ilustrar a habilidade de um agente artificial de aprender e responder ao seu ambiente, independentemente da vontade do seu desenvolvedor, os autores relatam o ocorrido em 2002 no evento *Living Robots*, na Inglaterra. Nele, robôs confinados em uma arena eram divididos em grupos de “presas” e “caçadores”, cada qual com a sua respectiva função. Utilizando-se de sensores infravermelhos, estes deveriam buscar alimentos (indicados por luz), enquanto que aqueles os caçariam e drenariam sua energia. Buscava-se, desse modo, verificar se os robôs participantes conseguiriam, independentemente, criar novas técnicas de autodefesa e de caça¹⁰³.

Durante o experimento, o robô denominado Gaak e pertencente ao grupo das presas, ao ser acidentalmente deixado sem vigilância por 15 minutos, conseguiu escapar dos confins

¹⁰⁰ BALKIN, Jack M. The Three Laws of Robotics in the Age of Big Data. *Ohio State Law Journal*, v. 78, p.1217-1241, 2017.

¹⁰¹ ČERKA, Paulius; GRIGIENĖ, Jurgita; SIRBIKYTĖ, Gintarė. Liability for damages caused by artificial intelligence. *Computer Law & Security Review*, v. 31, n. 3, p. 376-389, 2015.

¹⁰² GUTIERREZ, Andrei. Op. cit., p. 90.

¹⁰³ ČERKA, Paulius et al. Op. cit., p. 381.

da arena do evento e, ao chegar a ao estacionamento, foi atropelado por um carro (que sofreu danos consideráveis)¹⁰⁴.

Outro exemplo comumente citado da ocorrência de danos decorrentes do funcionamento normal de agentes artificiais autônomos (no sentido de agirem em consonância com a sua finalidade programada) é o *Flash Crash* ocorrido em 2010. Na ocasião, o índice Down Jones sofreu, em um intervalo de cinco minutos, uma queda brusca de mais de mil pontos (ou nove por cento) em relação ao momento de abertura, sem motivo aparente. Em questão de poucos minutos, mais de um trilhão de dólares em valores de investidores (economias, aposentadorias garantidas por fundos de pensão, dentre outros) foram perdidos¹⁰⁵.

Foi necessário cerca de seis meses para que a *U.S. Exchange and Securities Comission* determinasse o que havia ocorrido e, ao contrário do imaginado, não se tratou de um grave erro de funcionamento, mas sim de programas de *high frequency trading* fazendo exatamente aquilo para o qual foram programados. O problema teve sua origem na oferta de venda, por um grande fundo de investimentos, de ações agrupadas em uma aplicação financeira altamente diversificada. Contudo, não havendo de imediato (isto é, em questão de milissegundos) compradores no mercado, e nenhum investidor humano supervisionando a operação, os preços caíram, o que foi detectado pelos *bots* de HFT cujas ações no mercado se influenciaram reciprocamente, desencadeando automaticamente estratégias de controle de perdas (*stop loss*), com ordens de venda a qualquer preço¹⁰⁶.

Dos fatos narrados até agora, extrai-se que: i) agentes artificiais autônomos, conforme a definição adotada neste trabalho, podem agir de maneira imprevisível, seja em relação a usuários, seja em relação aos seus próprio desenvolvedores; ii) tal imprevisibilidade emana da própria capacidade de aprendizado do agente e de sua interação com o meio ambiente, e não necessariamente de erros de design, programação ou treinamento; iii) precisamente por esta razão, não é possível atribuí-los ao ser humano diretamente responsável por uma destas funções.

Quando se observam as mudanças ocorridas no papel do programador de uma máquina, desde os conceitos simples de programação da engenharia de software tradicional até a construção de sistemas de inteligência artificial autônomos, pode-se claramente

¹⁰⁴ WAINWRIGHT, Martin. Robot fails to find a place in the sun. *The Guardian*, 20 jun. 2002. Disponível em: <http://www.theguardian.com/uk/2002/jun/20/engineering.highereducation>. Acesso em: 29 out. 2019.

¹⁰⁵ KAPLAN, Jerry. *Humans Need Not Apply: A guide to wealth and work in the age of artificial intelligence*. New Haven: Yale University Press, 2015, p. 61-63.

¹⁰⁶ Id., p. 61-63.

identificar um processo no qual aquele perde cada vez mais o controle sobre ela e transfere gradualmente controle para a própria máquina.

Inicialmente, o programador era responsável pela escrita do código a ser utilizado - e, portanto, o comportamento operacional da máquina - linha por linha, em uma representação linguística a ser executada diretamente por ela. O programador, no controle do comportamento da máquina em cada detalhe, podia explicar o funcionamento de seu algoritmo, e, do mesmo modo, um observador poderia seguir tal explicação e verificar a sua correção. Erros, portanto, eram sempre erros do programador, e não erros do programa, passíveis de serem identificados, corrigidos, e a ele atribuídos¹⁰⁷.

Todavia, quando as ações de um agente artificial baseiam-se não apenas no seu programa inicial (dados pré-programados), mas dependem em grande parte de sua experiência pós-design, até mesmo os programadores e fabricantes mais diligentes não serão capazes de controlar ou prever tudo que o agente experimentará uma vez que deixar os seus cuidados.

Tecnologias úteis na prática, embora sejam previamente treinadas em ambientes controlados antes de terem sua interação com o público liberada (lembre-se do caso citado anteriormente da presença de um motorista de *backup* em carros autônomos), irão se deparar com situações no mundo real que podem não ter feito parte de seu processo de treinamento (dando continuidade ao exemplo de um veículo autônomo, imagine-se uma sacola plástica voando em sua direção e, não tendo antes se deparado com uma situação similar, o veículo freia bruscamente acreditando se tratar de um obstáculo na pista). Isso significa que tais tecnologias terão que cometer erros durante a sua operação, que poderão ocasionar danos (sendo um “erro” a exploração, pela própria máquina, do espaço de alternativas possíveis, o que permite que ela alcance autonomamente novas soluções e aprenda com o ocorrido¹⁰⁸).

Além disso, é interessante destacar a espécie de paradoxo representado por agentes artificiais autônomos no que diz respeito à sua previsibilidade. Tratando especificamente de veículos autônomos, Surden e Williams apontam que, na grande maioria dos casos, estes serão mais previsíveis do que os motoristas humanos¹⁰⁹, seguindo infalivelmente as instruções que lhes foram dadas e as regras de direção. No entanto, a questão deve ser avaliada não

¹⁰⁷ MATTHIAS, Andreas. Op. cit., p. 181.

¹⁰⁸ Id., p. 182.

¹⁰⁹ Nesse contexto, estima-se, atualmente, que cerca de 1,25 milhão de pessoas morrem anualmente em acidentes de trânsito (o dobro das mortes causadas por guerras, crimes e atos terroristas somadas), e 90% desses acidentes são causados por erros humanos. Espera-se, dessa forma, que a substituição de motoristas humanos por carros autônomos reduzirá mortes e ferimentos na estrada em cerca de 90% (HARARI, Yuval Noah. *21 Lições para o Século 21*. [Trad.] Paulo Geiger. São Paulo: Companhia das Letras, 2018, p. 45-46).

apenas do ponto de vista da confiabilidade tecnológica ou da engenharia de sistemas, mas também da intuição das pessoas comuns que trabalham, vivem e se movem na mesma proximidade física desses veículos. Simplesmente observando um carro autônomo, um leigo não pode dizer com facilidade quais dados ele coletou, em que, dentre muitos sensores, o sistema está “prestando atenção”, nem qual comportamento os algoritmos de controle interno o instruirão a seguir. Seus estados internos não são transparentes e nem intuitivamente compreensíveis para as pessoas, a menos que sejam especificamente projetados de outra maneira (enquadrando-se no conceito dos autores de opacidade tecnológica)¹¹⁰.

Do mesmo modo, outros tipos de agentes artificiais autônomos, como robôs ou drones aéreos, cada vez mais farão a transição de contextos especializados (como fábricas ou laboratórios) para espaços de consumo, levantando questões semelhantes.

Isso porque seres humanos desenvolveram sistemas cognitivos que permitem prever com segurança os movimentos de curto prazo daqueles que os rodeiam (capacidade englobada no conceito de teoria da mente). Mas, em se tratando de avaliar as ações futuras de máquinas, inexistem habilidades intuitivas comparáveis – o que, via de regra, não representou um problema até este ponto, vez que as grandes máquinas automatizadas que se deslocam no ambiente físico - como elevadores, escadas rolantes e equipamentos de fábrica - tendem a ser restritas em sua amplitude de movimento e, portanto, amplamente previsíveis¹¹¹.

Os mecanismos cognitivos de teoria da mente evoluíram em seres humanos por serem criaturas sociais, possibilitando viver cooperativamente em grupos, e de forma a antecipar as ações de curto prazo de outros seres humanos - e não o comportamento de máquinas. Em outras palavras, os humanos não têm as bases instintivas necessárias para entender o que um sistema tecnológico controlado por algoritmos, como um veículo autônomo, fará em seguida.

Por exemplo, imagine um pedestre prestes a atravessar a rua, na faixa de pedestres, e na frente de um carro que se aproxima. Para fazê-lo de maneira segura, ele deverá prever a probabilidade de que o motorista pare, através de uma série de observações instantâneas sobre suas percepções, capacidades e intenções: o motorista vê o pedestre ou encontra-se distraído ao celular? Ele está diminuindo a velocidade, ou então comunicou a sua intenção de parar através de um aceno ou contato visual? Tais observações permitem aos seres humanos reagir a circunstâncias dinâmicas em seu ambiente físico imediato.

Agora, no contexto de um veículo autônomo que diminui visivelmente a velocidade quando se aproxima da faixa, o pedestre pode ser levado a acreditar que aquele está

¹¹⁰ SURDEN, Harry; WILLIAMS, Mary-Anne. Op. cit., p. 152-153.

¹¹¹ Id., p. 151-152.

comunicando sua intenção de parar, similarmente a um motorista humano. No entanto, o veículo pode não ter realmente detectado o pedestre, e estar diminuindo a velocidade por um motivo completamente diverso (ele pode ter uma regra segundo a qual a sua velocidade é automaticamente reduzida à medida que se aproxima de uma faixa de pedestres, ainda que não haja a previsão de parar completamente)¹¹².

Assim, a menos que mecanismos eficazes de comunicação sejam implementados nesses sistemas, de forma a reduzir a sua opacidade tecnológica, poderão haver acidentes decorrentes de sua interação com seres humanos¹¹³, sobretudo inicialmente, quando essas novas tecnologias estiverem em período de transição para o ambiente de consumo.

Dos fatos narrados, percebe-se que a disseminação de agentes autônomos baseados em inteligência artificial na atualidade contemporânea representa novos desafios jurídicos, em especial para a disciplina da responsabilidade civil.

Afirma-se que, para que seja atribuída responsabilidade – sob qualquer fundamento teórico –, é necessário identificar uma pessoa a quem possa ser imputada a obrigação de reparar um dano. A personalidade jurídica constitui, portanto, pressuposto para a responsabilização. Mesmo que o dano tenha sido ocasionado por uma coisa, ou seja, um bem ou ente desprovido de personalidade, a obrigação de indenizar deverá ser deferida a uma pessoa natural ou jurídica, únicas figuras a quem a grande maioria dos ordenamentos legais concede, atualmente, capacidade de atuação jurídica¹¹⁴.

Ainda, responsabilidade e autonomia constituem duas faces de uma mesma moeda, no sentido de que, para que determinada pessoa seja obrigada a reparar um dano injusto, é fundamental que ela possua autonomia de atuação. Significa dizer que ela deverá ter a capacidade de reconhecer a licitude ou ilicitude de sua conduta e, ao mesmo tempo, de identificar e prever a potencialidade danosa desta¹¹⁵.

Uma vez que os agentes artificiais ora tratados detêm autonomia relativamente aos seres humanos que os desenvolveram ou os operaram, podendo desempenhar ações por eles

¹¹² O exemplo remonta a SURDEN, Harry; WILLIAMS, Mary-Anne. Id., p. 124-126.

¹¹³ Em fevereiro de 2016, ao analisar um acidente entre um veículo autônomo da Google e um ônibus, o porta voz da empresa afirmou se tratar de um exemplo da negociação costumeira no trânsito, em que todos tentam prever os movimentos dos demais. O carro havia detectado o ônibus que se aproximava, mas previu que ele cederia a passagem; da mesma forma, pode-se imaginar que o motorista do ônibus presumiu que o carro iria permanecer no mesmo local. Note-se que a incapacidade de prever com segurança o comportamento do veículo autônomo desempenhou um papel crucial no acidente em comento (LEE, Dave. Google self-driving car hits a bus. *BBC News*, fev. 2016. Disponível em: <http://www.bbc.com/news/technology-35692845>. Acesso em: 31 out. 2019).

¹¹⁴ MULHOLLAND, Caitlin. Op. cit., p. 332. Sobre as hipóteses previstas de responsabilidade indireta, quais sejam, responsabilidade por fato de outrem, pelo fato da coisa ou pelo fato do produto, ver artigos 931, 932, 933, 936 e 937 do Código Civil, e artigo 12 do Código de Defesa do Consumidor.

¹¹⁵ Id., p. 332.

não antecipadas, questiona-se: a sua atuação seria suficiente para romper o nexo causal¹¹⁶ entre a conduta do programador, operador, ou usuário, e o dano reparável? Ou, ainda, esses indivíduos poderiam se isentar da responsabilidade ao argumento de que a conduta do sistema era para eles imprevisível?

Tem-se, aqui, aquilo que Andreas Matthias denominou “lacuna de responsabilidade” (*responsibility gap*), resumida nos seguintes termos:

Tradicionalmente, tomamos o operador ou o fabricante da máquina como responsável pelas consequências de sua operação, ou ‘ninguém’ (em casos em que nenhuma falha pessoal possa ser identificada). Agora, pode ser demonstrado que há uma crescente classe de ações de máquinas, onde as formas tradicionais de atribuição de responsabilidade não são compatíveis com nosso senso de justiça e com a estrutura moral da sociedade, porque ninguém tem *controle* suficiente sobre as ações da máquina para poder assumir a responsabilidade por elas (tradução nossa)¹¹⁷.

Similarmente, Mittelstadt et al. ressaltam que o design modular desses sistemas pode significar que nenhuma pessoa ou grupo pode entender completamente a maneira pela qual ele interage ou responde a um fluxo complexo de novas entradas. Da programação linear tradicional aos algoritmos autônomos, o controle comportamental é gradualmente transferido do programador para o algoritmo e seu ambiente operacional. A diferença entre o controle do criador e o comportamento do algoritmo cria, portanto, uma lacuna de responsabilização em que a culpa pode potencialmente ser atribuída a vários agentes morais simultaneamente¹¹⁸.

É interessante notar que, no tocante a essa questão, em relatório publicado pela UNESCO em 2017 sobre Ética Robótica, ressaltou-se que nesses casos parece haver uma responsabilidade “compartilhada” ou “distribuída” entre designers da máquina, engenheiros, programadores, fabricantes e usuários. Nenhum deles pode ser indicado como fonte última de ação de um sistema artificial autônomo. Ocorre que, ao mesmo tempo, tal solução tende a diluir completamente a noção de responsabilidade: se todos têm uma contribuição sobre a responsabilidade total, acaba que ninguém é totalmente responsável (situação conhecida como

¹¹⁶ Trata-se do elemento que liga a conduta do agente ao dano sofrido pela vítima (TEPEDINO, Gustavo; SILVA, Rodrigo. *Inteligência Artificial e Elementos da Responsabilidade Civil*. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 293-325).

¹¹⁷ No original: “Traditionally we hold either the operator/manufacturer of the machine responsible for the consequences of its operation, or ‘nobody’ (in cases, where no personal fault can be identified). Now it can be shown that there is an increasing class of machine actions, where the traditional ways of responsibility ascription are not compatible with our sense of justice and the moral framework of society because nobody has enough control over the machine’s actions to be able to assume the responsibility for them” (ANDREAS, Matthias. Op. cit., p. 7).

¹¹⁸ MITTELSTADDT, Brent; FLORIDI, Luciano; ALLO, Patrick; TADDEO, Mariarosaria; WACHTER, Sandra. The ethics of algorithms: Mapping the debate. *Big Data & Society*, v. 3(2), p. 1-21, 2016.

o “problema das muitas mãos” ou “*the problem of the many hands*”)¹¹⁹.

Deste modo, na ocorrência de um dano injusto causado por um agente artificial autônomo, duas questões se sobressaem: A primeira é relativa a quem será atribuído o dever de reparação por esse dano; a segunda, atinente ao fundamento da responsabilidade civil capaz de justificar tal imposição. A eles, soma-se ainda um elemento agravante: a ininteligibilidade dos outputs desses sistemas.

3.2. Ininteligibilidade

Viu-se que, por trás de grande parte das aplicações da inteligência artificial que atualmente se fazem cada vez mais presentes não apenas em laboratórios de pesquisas, mas também no cotidiano das pessoas, existe um número incomensurável de dados que viabilizam o seu funcionamento.

De fato, o entusiasmo recente no que diz respeito à pesquisa e desenvolvimento de tecnologias baseadas em IA pode ser resumido na ocorrência de quatro fatores interligados e que, portanto, se retroalimentam: a existência de métodos estatísticos e probabilísticos cada vez mais sofisticados; a disponibilidade de um número expansivo de dados; a acessibilidade a um enorme poder computacional a baixo custo; e a crescente adequação de ambientes às novas tecnologias, como a automação residencial e a criação de cidades ‘inteligentes’.¹²⁰

Tais fatores integram aquilo que se denomina indústria 4.0 - *Big Data*, Inteligência Artificial e Internet das Coisas (IoT ou *Internet of Things*), resultante da transição do mundo analógico para o digital.

Estima-se que, no ano 2000, apenas um quarto das informações armazenadas no mundo era digital, estando o restante delas em papel, filme, vinis, fitas cassete magnéticas, dentre outros meios analógicos. Mas, como os dados digitais rapidamente se expandem - dobrando um pouco mais do que a cada três anos, segundo Martin Hilbert - a situação logo se inverteu e, já em 2007, apenas 7% dos dados existentes eram analógicos¹²¹.

¹¹⁹ NAÇÕES UNIDAS. UNESCO’s Report of COMEST on robotics ethics, 2017, p. 42. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000253952>. Acesso em: 31 out. 2019.

¹²⁰ FLORIDI, Luciano; CATH, Carine; MITTELSTADT, Brent; WACHTER, Sandra; TADDEO, Mariarosaria. Artificial Intelligence and the ‘Good Society’: the US, EU, and UK approach. *Science and Engineering Ethics* (Springer), p. 1-24, 2017. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2906249. Acesso em: 29 out. 2019.

¹²¹ HILBERT, Martin; LOPES, Priscila. How to Measure the World’s Technological Capacity to Communicate, Store and Compute Information. *International Journal of Communications*, v. 6, p. 956-979, 2012. Disponível em: <https://ijoc.org/index.php/ijoc/article/viewFile/1562/742&from=32207443>. Acesso em: 30 out. 2019.

Em 2013, a quantidade total de informações armazenadas no mundo era estimada em cerca de 1.200 exabytes¹²² (em oposição a 300 exabytes apenas 6 anos antes), dentre as quais menos de 2% não eram digitais. Em 2014, a quantidade de informações armazenadas crescia quatro vezes mais rápido que a economia mundial, e o poder de processamento dos computadores nove vezes mais rápido¹²³.

Surgiram, portanto, as condições ideais para, a partir da aplicação matemática a grandes quantidades de dados, inferir probabilidades, ideia basilar por trás do chamado *Big Data*. A análise de *Big Data* permite inferir informações preditivas de elevados números de dados, a fim de adquirir mais conhecimento, por exemplo, sobre indivíduos e grupos¹²⁴. Embora seja comumente descrito como parte do ramo do aprendizado de máquina no campo da inteligência artificial, trata-se, na verdade, de conceitos distintos, sendo este, em linhas gerais, a técnica que possibilita a realização daquele.

Se, no passado, o valor de cada dado era atribuído de maneira estática, ou seja, sua utilidade se perdia uma vez que o propósito para o qual ele havia sido coletado era alcançado, com o *Big Data* seu valor latente não é claro no momento da coleta, vindo a ser obtido somente na medida em que os dados são reutilizados repetidamente para finalidades diversas. Além disso, esse valor pode ser aprimorado não apenas com um aumento quantitativo dos dados analisados, mas principalmente quando combinados com bases distintas de dados.

Essa mudança na forma de se valorar os dados é responsável por criar um incentivo econômico elevado na maneira como eles são tratados: em primeiro lugar, sua coleta passa a ser oportunista, sendo realizada sempre que houver a possibilidade para tanto e ainda que nenhum uso concreto seja evidenciado de imediato. Ademais, estes dados passam a ser mantidos pelo maior tempo possível, muito além do uso inicial previsto, visando à sua reutilização e recombinação junto a outras bases de dados¹²⁵.

Assim, podem ser utilizados para propósitos não necessariamente relacionados àqueles iniciais da coleta, e muitas vezes prejudiciais aos titulares desses dados (por exemplo, para a construção de perfis que irão diferenciar o tipo de acesso a bens, serviços e até mesmo direitos básicos a que esses grupos ou indivíduos terão acesso). É justamente nesse contexto

¹²² Embora inexista uma boa maneira de refletir sobre o significado dessa quantidade de dados, imagine que, se todos eles fossem impressos em livros, cobririam toda a superfície dos Estados Unidos em 52 camadas de espessura. Se fossem colocados em CD-ROMs, estes poderiam ser separados em cinco pilhas diferentes que se estenderiam até a lua. (MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. Op. cit., p. 11).

¹²³ Id., p. 12.

¹²⁴ Id., p. 6.

¹²⁵ MAYER-SCHÖNBERGER, Viktor; PADOVA, Yann. Regime Change? Enabling big data through Europe's new data protection regulation. *The Columbia Science and Technology Law Review*, v. 17, p. 315-335, 2017.

que as recentes leis de proteção de dados pessoais se enquadram, como a Lei Geral de Proteção de Dados Pessoais brasileira (Lei 13.709/2018 ou LGPD) e o Regulamento Geral de Proteção de Dados da União Europeia (GDPR ou *General Data Protection Regulation*)¹²⁶.

Além disso, Mayer-Schönberger e Cukier apontam para três grandes mudanças de mentalidade proporcionadas pelo *Big Data*. Em primeiro lugar, a capacidade de analisar grandes quantidades de dados sobre um tópico, em vez de se ter que conformar com conjuntos menores (através, por exemplo, de métodos estatísticos que permitem extrapolar, a partir de uma pequena amostra, conhecimentos úteis sobre a população em geral). Segundo, a disposição para aceitar a confusão do mundo real dos dados, em vez de privilegiar a exatidão (o *Big Data* transforma números em algo mais probabilístico do que preciso e, em muitas situações, isso é bom o bastante). Finalmente, a terceira mudança diz respeito à transição da causalidade por correlações simples: não saber exatamente por que algo acontece, mas apenas que algo de fato acontece¹²⁷.

Com efeito, nas palavras de Pearl e Mackenzie, dados são eminentemente “burros”¹²⁸. Eles podem dizer que pessoas que tomaram determinado medicamento se recuperaram mais rápido do que aquelas que não o tomaram, mas não podem explicar o porquê disto.

Para os autores, vive-se em uma época que presume que o *Big Data* é a solução para todo e qualquer problema vivenciado. Entretanto, tudo que um programa de *machine learning* é capaz de fazer é ajustar uma função aos dados. Nenhuma máquina pode extrair explicações a partir apenas de dados brutos, para tanto, ela deve dominar ao menos três níveis distintos de capacidade cognitiva, ou três níveis de causalidade.

O primeiro deles consiste em detectar regularidades a partir de observações, capacidade compartilhada por muitos animais e também pelas máquinas de aprendizado atuais. É isto que o programa *AlphaGo* faz quando analisa uma base de dados de milhões de partidas do jogo Go para descobrir quais jogadas estão associadas a um maior percentual de

¹²⁶ BRASIL. Lei. n. 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, [2019]. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 26/08/2019 ; UNIÃO EUROPEIA. *Regulamento 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016*. Relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados). Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679&from=pt>. Acesso em: 26/08/2019.

¹²⁷ MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. Op. cit., p. 16.

¹²⁸ PEARL, Judea; MACKENZIE, Dana. *The Book of Why: The new science of cause and effect*. New York: Basic Books, 1. Ed., 2018, p. 6.

vitórias.

Em sua essência, uma correlação quantifica a relação estatística entre dois valores de dados, podendo ser classificada como forte quando, se um dos valores dos dados muda, é altamente provável que o outro também mude; ou fraca, se um valor dos dados muda e pouco ou nada acontece com o outro¹²⁹. Trata-se de previsões baseadas em observações passivas e, embora algumas associações possam ter interpretações causais óbvias, métodos estatísticos não são capazes de derivar, por si só, qual é a causa e qual é o efeito¹³⁰.

O segundo nível, intervenção, implica prever os efeitos de alterações deliberadas no ambiente, escolhendo dentre elas aquela que produzirá o resultado almejado¹³¹. Não é possível responder a perguntas sobre intervenções apenas com dados passivamente coletados, não importa o tamanho da base utilizada ou a profundidade da rede neural que realiza a análise. Para tanto, são necessários experimentos sob condições cuidadosamente controladas.

Finalmente, o terceiro nível apontado pelos autores envolve a capacidade de imaginar como as coisas poderiam ser diferentes, ou seja, de raciocinar contrafactualmente – algo que possibilitou aos seres humanos aprender com a história e com as experiências dos outros, distinguindo-se das demais espécies¹³².

A vantagem em se imaginar contrafactuais, portanto, encontra-se na capacidade de refletir e aprimorar ações passadas e, talvez mais significativamente, assumir a responsabilidade por essas ações:

Responsabilidade e culpa, arrependimento e crédito: esses conceitos são a moeda de uma mente causal. Para entendê-los, devemos ser capazes de comparar o que aconteceu com o que teria acontecido sob alguma hipótese alternativa (tradução nossa)¹³³.

A intuição humana é fundada em uma lógica causal, e não estatística, inclinando-se a enxergar causas mesmo onde não existem: ao se deparar com dois eventos sucessivos, existe um forte viés de interpretá-los em termos causais¹³⁴. Além disso, a análise da maneira pela qual seres humanos dão e recebem explicações revelam a preferência por aquelas

¹²⁹ MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. Op. cit., p. 33.

¹³⁰ PEARL, Judea; MACKENZIE, Dana. Op. cit., p. 29.

¹³¹ Id., p. 27.

¹³² HARARI, Yuval Noah. *Sapiens: Uma breve história da humanidade*. São Paulo: L&PM, 2015.

¹³³ No original: “Responsibility and blame, regret and credit: these concepts are the currency of a causal mind. To make any sense of them, we must be able to compare what did happen with what would have happened under some alternative hypothesis” (PEARL, Judea; MACKENZIE, Dana. Op. cit., p. 260).

¹³⁴ Para Kahneman, a maneira rápida e intuitiva de pensar (sistema 1) é fortemente tendenciosa a enxergar vínculos causais mesmo quando não há nenhum, sendo enviesada a confirmar conhecimentos e crenças pré-existentes (KAHNEMAN, Daniel. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2013).

contrafactuais e contrastantes, ou seja, as pessoas não pedem explicações do porquê dado evento ocorreu, e sim por que ele ocorreu no lugar de uma alternativa diversa. Busca-se uma declaração de como o mundo teria de ser diferente para que um resultado desejável ocorresse¹³⁵.

É no nível das contrafactuais que questões jurídicas atinentes à responsabilidade se encontram. Na seara cível, para que surja o dever de indenizar, é preciso que o dano sofrido pela vítima seja consequência da conduta do agente, o que requer a análise comparativa entre o mundo real e o contrafactual: retirando-se a conduta do agente, verifica-se se o dano ainda assim teria ocorrido. No que pese as diversas construções teóricas que visam explicar o nexo causal, tem-se, em linhas gerais, os contornos essenciais da teoria da causalidade direta e imediata¹³⁶.

O raciocínio contrafactual possui, portanto, uma relação especialmente problemática com dados, uma vez que estes são, por definição, fatos. Assim sendo, eles não podem dizer o que aconteceria em um mundo imaginário, onde é possível negar ou alterar alguns dados observados¹³⁷ (como a conduta de um agente que resultou em dano).

Justamente por não possuírem a capacidade de visualizar realidades alternativas e contrastá-las com a realidade atual existente é que uma IA não poderia passar naquilo que Pearl e Mackenzie denominaram mini Teste de Turing¹³⁸: um teste para verificar se uma máquina estaria apta a responder corretamente a perguntas causais, como um ser humano é capaz de fazê-lo (ou seja, através da compreensão básica dos porquês).

Os autores ressaltam que, nos últimos anos, os notáveis progressos em inteligência artificial ocorreram na área de *deep learning*, através da utilização de redes neurais. Estas, contudo, não lidam com a incerteza de maneira rigorosa ou transparente, e menos ainda

¹³⁵ MITTELSTADT, Brent; WACHTER, Sandra; RUSSELL, Chris. Explaining Explanations in AI. *Proceedings of FAT* '19: Conference on Fairness, Accountability, and Transparency*, 2019. Disponível em: <https://arxiv.org/pdf/1811.01439.pdf>. Acesso em: 31 out. 2019.

¹³⁶ À teoria da causalidade direta e imediata se adicionou, em doutrina, a exigência (por vezes denominada *subteoria*) da necessidade da causa, de modo a se concluir que surge o dever de indenizar quando o dano é efeito necessário de certa causa. A causa deixa de ser relevante, desse modo, não por estar distante do dano, mas sim pela interferência de outra causa capaz de produzi-lo. As maiores dificuldades surgem, então, nas hipóteses em que uma pluralidade de causas contribui para a produção do efeito danoso, cabendo ao magistrado definir a qual delas deve ser imputado o dever de reparar. (TEPEDINO, Gustavo; SILVA, Rodrigo. Op. cit., p. 310).

¹³⁷ Por óbvio, isto não significa que a sua análise seja inútil: em muitos casos, probabilidades, e não certezas, são suficientes para solucionar determinado problema, e podem também ser um primeiro passo essencial para procurar padrões interessantes de associação e levantar questões interpretativas mais precisas (PEARL, Judea; MACKENZIE, Dana. Op. cit., p. 33).

¹³⁸ Id., p. 349. O teste é "mini" por duas razões: primeiro, limita-se ao raciocínio causal, excluindo outros aspectos da inteligência humana, como visão e linguagem natural. Segundo, é permitido à máquina codificar a história que lhe é apresentada em qualquer representação que lhe seja conveniente, retirando o fardo de ter que adquiri-la a partir de sua própria experiência pessoal. Segundo os autores, o entendimento dos três níveis de causalção seria o caminho para se obter uma inteligência artificial forte.

incorporam qualquer representação explícita do ambiente em que operam. Em vez disso, a arquitetura da rede fica livre para evoluir por conta própria e, ao final do treinamento, seu programador não tem ideia de quais cálculos estão sendo executados ou por que eles funcionam. Consequentemente, se a rede falhar, ele não terá uma ideia clara de como corrigi-la¹³⁹.

O aprendizado profundo, para eles, seria a antítese da transparência. Para muitos, a transparência não é de fato necessária, em analogia, por exemplo, ao cérebro humano, sobre cujo funcionamento se possui uma compreensão limitada, porém suficiente. Do mesmo modo, questionam-se, por que não criar e fazer amplo uso desses sistemas sem saber em detalhes como eles funcionam? Pearl e Mackenzie alertam que

Sim, perdoamos nossa escassa compreensão de como o cérebro humano funciona, mas ainda podemos nos comunicar com outros humanos, aprender com eles, instruí-los e motivá-los em nossa própria língua nativa de causas e efeitos. Podemos fazer isso porque nossos cérebros funcionam da mesma maneira. Se todos os nossos robôs forem tão opacos quanto o AlphaGo, não seremos capazes de manter uma conversa significativa com eles, o que seria bastante lamentável (tradução nossa)¹⁴⁰.

A transparência possibilita, assim, uma comunicação eficaz. No caso de sistemas especialistas, verificou-se que regras explicitamente definidas transformam as variáveis de entrada (como informações acerca de transações realizadas em um cartão de crédito) em variáveis de saída (um sinalizador de fraude). A seu turno, algoritmos de *machine learning*, viu-se, são alimentados por ambas as variáveis conjuntamente, de forma que o sistema aprenda as correlações necessárias. Esse processo treina o modelo exibindo não uma lógica explícita, mas sim implícita e normalmente não otimizada para o entendimento humano.

Nesse mesmo sentido, Veale e Edwards ressaltam que esses algoritmos são gerados tendo como prioridade o desempenho preditivo, ao invés da interpretabilidade de suas ações. Aprendizado, nesse contexto, refere-se ao aperfeiçoamento do modelo em uma tarefa específica, conforme a medida de desempenho escolhida por seus desenvolvedores. Esta avaliação não se dá através do interrogatório de sua estrutura interna, mas sim examinando seu comportamento diante da métrica de desenvolvimento adotada¹⁴¹.

¹³⁹ Id., p. 323.

¹⁴⁰ No original: “Yes, we forgive our meager understanding of how human brains work, but we can still communicate with other humans, learn from them, instruct them, and motivate them in our own native language of cause and effect. We can do this because our brains work the same way. If our robots will all be as opaque as AlphaGo, we will not be able to hold a meaningful conversation with them, and that would be quite unfortunate” (Id., p. 362).

¹⁴¹ EDWARDS, Lilian; VEALE, Michael. Op. cit., p. 25.

Em linhas gerais, é possível identificar duas categorias de explicações tecnicamente viáveis sobre o funcionamento de um sistema de *machine learning*. A primeira delas é centrada no próprio modelo (*model-centric explanations* ou MCEs), consistindo em uma explicação global de todo o modelo, sem focar especificamente em determinada decisão ou dados de entrada. As informações fornecidas a partir de uma abordagem MCE são úteis para a verificação da consistência ou regularidade procedimental dos outputs do modelo, e podem incluir:

- informações de configuração: as intenções por trás do processo de modelagem, a família de modelos (rede neural, floresta aleatória, combinação de conjuntos), os parâmetros usados para especificá-lo antes do treinamento;
- metadados de treinamento: estatísticas resumidas e descrições qualitativas dos dados de entrada usados para treinar o modelo, a proveniência de tais dados e os dados ou classificações de saída previstos neste modelo;
- métricas de desempenho: informações sobre a habilidade preditiva do modelo sobre dados não vistos (...);
- lógicas globais estimadas: são formas simplificadas, aproximadas e compreensíveis pelo homem de como as entradas são transformadas em saídas, que por definição não são completas; caso contrário, poderiam ser usadas em vez do modelo complexo para obter os mesmos resultados (...);
- informações do processo: como o modelo foi testado, treinado ou rastreado quanto a propriedades indesejáveis (tradução nossa).¹⁴²

Por outro lado, explicações centradas no sujeito (*subject-centric explanations* ou SCEs) são elaboradas tendo por base um input específico, razão pela qual são também referidas como explicações locais, estando restritas à região em torno de um conjunto de dados. Essa abordagem se faz necessária uma vez que modelos complexos não podem ser explicados de maneira eficaz em sua totalidade (resultando em sua frequente comparação a uma “caixa preta”, na terminologia proposta por Pasquale)¹⁴³.

Introduzindo o conceito de dimensionalidade, frequente na práxis da ciência da computação, Edwards e Veale explicam que dados podem ser pensados geometricamente: com duas variáveis numéricas, é possível exibi-los em um gráfico de dispersão bidimensional. Com três variáveis, em um gráfico tridimensional. Conceitualmente, isto pode ser escalado infinitamente a depender de quantas variáveis existirem nos dados analisados. Na medida em que se aumentam as dimensões (isto é, o número de variáveis), a quantidade de maneiras pelas quais todos os valores potenciais podem ser combinados aumenta exponencialmente.

¹⁴² Id., p. 55-56.

¹⁴³ O termo pode se referir a um dispositivo de gravação, como os sistemas de monitoramento de dados em aviões ou trens; mas também pode significar um sistema cujo funcionamento é misterioso, sendo possível observar suas entradas e saídas, mas sem a certeza de como umas se transformam nas outras PASQUALE, Frank. *The Black Box Society: The secret algorithms that control money and information*. Cambridge: Harvard University Press, 2015.

Assim, modelos multicamadas que misturam combinações arbitrárias de diferentes variáveis, de maneira paralela e interdependente, resultam em um grau de complexidade que inviabiliza a explicação de todos os processos internos do modelo de uma única vez.

Apesar disso, é possível obter explicações locais quando se amplia e se destaca apenas a parte do espaço dentro e ao redor do vetor de variáveis que interessa investigar. Ao analisar apenas uma pequena porção do sistema, ele pode se tornar consideravelmente mais inteligível, e é precisamente isto que as abordagens SCE fazem: partindo-se do modelo a ser investigado, selecionam uma determinada região e a modelam de maneira mais simples e interpretável.

Note-se, todavia, que otimizar um modelo de explicação para a interpretabilidade humana significa necessariamente diluir o desempenho preditivo para capturar apenas as principais lógicas do sistema. Como sistemas com mais variáveis geralmente desempenham melhor do que sistemas mais simples, acaba por existir um *trade off* entre performance e explicabilidade.

Além disso, existe o problema fundamental de que, na grande maioria desses sistemas:

não existe uma teoria que correlacione variáveis de entrada com coisas que os seres humanos entendem como causais (...) diferentemente dos modelos de simulação, os recursos que estão sendo alimentados podem não ter nenhuma interpretação humana conveniente ou clara. O LinkedIn, por exemplo, afirma ter mais de 100.000 variáveis para cada usuário, que são alimentadas no modelo de machine learning. Muitas delas não serão variáveis claras, como "idade", mas maneiras mais abstratas de interagir com a página da web, como quanto tempo o usuário leva para dar um clique (...) Essas variáveis podem muito bem conter sinais preditivos sobre características ou comportamentos individuais, *mas não há maneiras convincentes de exibir claramente essas explicações para uma interpretação humana significativa* (tradução e grifo nossos).¹⁴⁴

Nesse contexto, a interpretabilidade pode ser definida como a medida do grau de um observador humano que tem de compreender as razões por trás de uma predição feita por um modelo de inteligência artificial. Especificamente no caso de algoritmos de aprendizado de máquina, é possível apontar três níveis distintos de interpretabilidade: alta, que inclui algoritmos tradicionais de regressão, árvores de decisão e classificadores baseados em regras (métodos de fácil compreensão); média, no caso de algoritmos um pouco mais avançados, tais

¹⁴⁴ No original: "In some systems there is no theory correlating input variables to things humans understand as causal or even as "things." In ML systems, unlike simulation models, the features that are being fed in might lack any convenient or clear human interpretation in the first place, even if we are creative about it. LinkedIn, for example, claims to have over 100,000 variables held on every user that feed into ML modelling. Many of these will not be clear variables like "age," but more abstract ways you interact with the webpage, such as how long you take to click, the time you spend reading, or even text. you type in a text box but later delete without posting. These variables may well hold predictive signals about individual characteristics or behaviours, but we lack compelling ways to clearly display these explanations for meaningful human interpretation" (Id., p. 59).

como modelos gráficos; e baixa interpretabilidade, compreendendo as técnicas avançadas de *machine learning*, tais como SVM (*Support Vector Machines*), *Ensemble Methods* e Redes Neurais Profundas. Esses métodos podem, no máximo, esboçar ideias de importâncias das variáveis do problema como medida de interpretabilidade¹⁴⁵.

Uma maneira de obter explicabilidade em sistemas de inteligência artificial é fazendo uso de algoritmos que são inerentemente explicáveis, ou seja, aqueles qualificados como de alta interpretabilidade na classificação acima. Entretanto, justamente pelo *tradeoff* apontado previamente, avanços importantes na área têm sido obtidos através de algoritmos com baixa interpretabilidade, sobretudo redes neurais multicamadas.

Notando a necessidade de fornecer explicações também para essas outras abordagens algorítmicas mais complexas, surgiu o conceito de “*Explainable AI*”, campo emergente no aprendizado de máquina que visa abordar como são tomadas as decisões por esses algoritmos¹⁴⁶.

Também no contexto de combate à opacidade de sistemas de IA, especialmente tendo em vista que, cada vez mais, eles são utilizados para gerar resultados que podem influenciar ou até mesmo basear diretamente decisões capazes de impactar de maneira crucial a vida das pessoas a elas sujeitas (resultando na perda de uma oportunidade de emprego, ou na negativa de crédito, por exemplo), é que se enquadra o chamado direito à explicação.

Esse direito, embora não mencionado expressamente nos referidos diplomas, decorreria tanto da Lei Geral de Proteção de Dados brasileira (LGPD) quanto do Regulamento Geral de Proteção de Dados da União Europeia (GDPR). Naquela, além da possibilidade de o titular solicitar a revisão, por um ser humano, de decisões tomadas unicamente com base em tratamento automatizado de seus dados pessoais, “o controlador deverá fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada”¹⁴⁷.

¹⁴⁵ DA SILVA, Nilton Correia. Inteligência Artificial. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 35-52.

¹⁴⁶ SCHMELZER, Ron. Understanding Explainable AI. *Forbes*, jul. 2019. Disponível em: <https://www.forbes.com/sites/cognitiveworld/2019/07/23/understanding-explainable-ai/#73fc6b337c9e>. Acesso em: 30 out. 2019.

¹⁴⁷ Artigo 20, § 1º, Lei 13.709/2018. Sobre o direito à explicação na Lei 13.709/2018, ver: LEITE, Renato. Existe um direito à explicação na Lei Geral de Proteção de Dados Pessoais do Brasil? Artigo Estratégico 24. Rio de Janeiro: Instituto Igarapé, dezembro de 2018. Disponível em: <https://igarape.org.br/wp-content/uploads/2018/12/Existe-um-direito-a-explicacao-na-Lei-Geral-de-Protecao-de-Dados-no-Brasil.pdf>. Acesso em: 31 out. 2019. Veja-se, ainda, o PL n. 4496/2019, de autoria do Senador Styvenson Valentim, que visa alterar a Lei n. 13.709/2018 para definir a expressão “decisão automatizada”. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/138136>. Acesso em: 06 dez. 2019.

Similarmente, no GDPR, exige-se a notificação dos titulares de dados pessoais sempre que estes forem utilizados para fins de tomada de decisão automatizada, sobre informações úteis acerca da lógica subjacente ao processo, e as significâncias e consequências previstas que poderão surgir da decisão¹⁴⁸.

Nesse sentido, em conferência realizada em janeiro de 2017 em Asilomar, na Califórnia, cujo objetivo principal era definir uma série de princípios para o desenvolvimento da IA de forma benéfica, elencou-se dentre eles princípio atinente à transparência em caso de falhas (“*failure transparency*”). Segundo ele, se um sistema de inteligência artificial causar danos, deverá ser possível verificar o porquê¹⁴⁹.

Conforme se verá adiante, as preocupações atinentes ao uso disseminado de inteligência artificial têm aparecido não apenas nas discussões sobre a proteção à privacidade e aos dados pessoais, ou então na forma de um direito à explicação e clamores por transparência algorítmica, mas também em propostas normativas que visam solucionar o problema da lacuna de responsabilidade.

4. Personalidade jurídica (eletrônica): solução à lacuna de responsabilidade?

Resumindo as questões suscitadas até o momento, tem-se que: i) agentes artificiais autônomos, por serem capazes de agir de maneira imprevisível, são também capazes de causar efetivamente danos a uma pessoa; ii) uma vez que a imprevisibilidade emana da própria capacidade de aprendizado dos agentes, é impraticável identificar *a priori* todas as potencialidades danosas da aplicação do sistema; iii) portanto, o dano causado estará diretamente ligado à conduta do agente, mas não necessariamente à conduta culposa do programador, operador ou usuário.

¹⁴⁸ Artigo 13(2), GDPR. Sobre o debate doutrinário acerca da existência e dos contornos do direito à explicação no âmbito europeu, ver: WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, p. 1-47, 2017. Disponível em: <https://ssrn.com/abstract=2903469>. Acesso em: 25 ago. 2019; SELBST, Andrew; POWLES, Julia. Meaningful Information and the Right to Explanation. *International Data Privacy Law*, v. 7, n. 4, p. 1-20, 2017. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3039125. Acesso em: 30 out. 2019; GOODMAN, Bryce; FLAXMAN, Seth. EU Regulations on Algorithmic Decision-Making and a “Right to Explanation”. *AI Magazine*, v. 38, n. 03, p. 1-10, 2016. Disponível em: <https://arxiv.org/abs/1606.08813>. Acesso em: 30 out. 2019; EDWARDS, Lilian; VEALE, Michael. Enslaving the Algorithm: From a ‘Right to an Explanation’ to a ‘Right to Better Decisions’? *IEEE Security & Privacy*, v. 16, n. 3, p. 46-54, 2018. Disponível em: <https://ssrn.com/abstract=3052831>. Acesso em: 26 ago. 2019. EDWARDS, Lilian; VEALE, Michael. Slave to the Algorithm? Why a ‘right to an explanation’ is probably not the remedy you are looking for. *Duke Law & Technology Review*, v. 16, n. 01, p. 18-84, 2016. Disponível em: <https://ssrn.com/abstract=2972855>. Acesso em: 26 ago. 2019.

¹⁴⁹ Disponível em <https://futureoflife.org/ai-principles/>. Acesso em: 29 out. 2019.

A tais fatos acrescenta-se, ainda, um elemento agravante: na medida em que os processos de aprendizagem e decisão independem da atuação e da racionalidade humana, é virtualmente impossível obter explicações acerca dos processos que levaram o agente a determinado resultado (output) ocasionador do dano.

Com isto em mente, cada vez mais a alteração do status desses agentes tem sido tema de iniciativas e diálogos tanto no âmbito internacional quanto em propostas regulatórias domésticas¹⁵⁰, como possível resposta aos problemas previamente tratados de atribuição de responsabilidade em caso de eventuais danos.

No que tange à regulação normativa, Polido classifica três campos preparatórios de intervenção do direito envolvendo atores estatais e não estatais: diálogos multilaterais e regionais (em particular por organizações como as Nações Unidas e a União Europeia); estratégias nacionais em IA (como a adoção de políticas governamentais direcionadas ao estímulo da indústria); e padrões transnacionais de regulação (regulação privada transnacional)¹⁵¹.

No sistema multilateral, a União Internacional das Telecomunicações organizou, em 2017, a Primeira Cúpula Global sobre Inteligência Artificial para o Bem¹⁵², esforço de diálogo a reunir as principais agências especializadas das Nações Unidas, representantes da indústria, organizações não governamentais e academia, buscando avanços em uma agenda multissetorial e exploratória sobre o tema. Tal agenda aponta para dois amplos eixos no plano internacional: a promoção de políticas e regulação da inteligência artificial (*“policy & regulation on AI”*) e a criação de padrões normativos internacionais para a mesma (*“international standards”*)¹⁵³.

Em maio de 2019, a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) aprovou a Recomendação do Conselho de Inteligência Artificial¹⁵⁴, e seus países-membros comprometeram-se com a adoção de princípios que visam ao desenvolvimento de uma IA inovadora e confiável, que respeite os direitos humanos e os valores democráticos. Os Princípios da OCDE sobre Inteligência Artificial são os primeiros do tipo a serem assinados

¹⁵⁰ Sobre o tema, ver: DUTTON, Tim. *An Overview of National AI Strategies*. Disponível em: <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>. Acesso em: 29 out. 2019.

¹⁵¹ POLIDO, Fabrício. Novas Perspectivas para Regulação da Inteligência Artificial: Diálogos entre as políticas domésticas e os processos legais transnacionais. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 179-206.

¹⁵² ITU. *AI for Good Global Summit*, 2018. Relatórios e documentos das conferências estão disponíveis em: [\[https://www.itu.int/web/pp-18/en/background/artificial-intelligence-for-good\]](https://www.itu.int/web/pp-18/en/background/artificial-intelligence-for-good). Acesso em: 25 nov. 2019.

¹⁵³ POLIDO, Fabrício. Op. cit., p. 185-186.

¹⁵⁴ OCDE. *Recommendation of the Council on Artificial Intelligence*, 2019. Disponível em: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>. Acesso em: 25 nov. 2019.

por governos e, para além dos membros da organização, têm como signatários também o Brasil, Argentina, Colômbia, Costa Rica, Peru e Romênia.

Dentre eles, destaca-se o dever de transparência e divulgação responsável (*“responsible disclosure”*) de sistemas de IA de forma a garantir que as pessoas entendam os seus resultados e possam desafiá-los; a necessidade de que esses sistemas funcionem de maneira segura ao longo de seus ciclos de vida, com a avaliação e gerenciamento contínuos de seus potenciais riscos; e a afirmação de que as organizações e indivíduos que desenvolvem, implementam ou operam sistemas de IA devem ser responsabilizados pelo seu bom funcionamento, de acordo com os demais princípios.

No que diz respeito às estratégias nacionais em inteligência artificial, menciona-se, nos Estados Unidos, o *Future of Artificial Intelligence Act*¹⁵⁵ e o *House Resolution 153* (H. RES. 153)¹⁵⁶, ambos de 2017, o *Algorithmic Accountability Act*¹⁵⁷ e a Ordem Executiva Presidencial sobre a manutenção da liderança norte americana em IA de 2019¹⁵⁸. Importante também destacar os relatórios do Comitê sobre Inteligência Artificial do Reino Unido, de 2017¹⁵⁹; e os Princípios para a Governança de Inteligência Artificial na China¹⁶⁰, do mesmo ano.

No Brasil, foi proposto em 2019 o Projeto de Lei do Senado n. 5.051, que “estabelece os princípios para o uso da inteligência artificial no Brasil”. Trata-se de iniciativa pioneira no trato legislativo da matéria no país, que traz consigo princípios fundamentais como: a garantia da proteção da privacidade e dos dados pessoais; a transparência, a confiabilidade e a possibilidade de auditoria dos sistemas; a supervisão humana. Além disso, trata também de questões atinentes ao trabalho e ao desenvolvimento econômico, e apresenta diretrizes de atuação para os Entes Federativos.

¹⁵⁵ ESTADOS UNIDOS. *House Resolution 4625: Future of Artificial Intelligence Act*, 2017. Disponível em: <https://www.congress.gov/bill/115th-congress/house-bill/4625?q=%7B%22search%22%3A%5B%22the%22%5D%7D>. Acesso em: 25 nov. 2019.

¹⁵⁶ ESTADOS UNIDOS. *House Resolution 153: Supporting the development of guidelines for ethical development of artificial intelligence*. Disponível em: <https://www.congress.gov/bill/116th-congress/house-resolution/153/text>. Acesso em: 25 nov. 2019.

¹⁵⁷ ESTADOS UNIDOS. *House Resolution 2231: Algorithmic Accountability Act*. Disponível em: <https://www.congress.gov/bill/116th-congress/house-bill/2231/text>. Acesso em: 25 nov. 2019.

¹⁵⁸ ESTADOS UNIDOS. *Executive Order on Maintaining American Leadership in Artificial Intelligence*, 2019. Disponível em: <https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/>. Acesso em: 25 nov. 2019.

¹⁵⁹ REINO UNIDO. *AI in the UK: Ready? Willing and Able?* 2017. Disponível em: <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>. Acesso em: 25 nov. 2019.

¹⁶⁰ CHINA. *Governance Principles for a New Generation of Artificial Intelligence: Develop responsible artificial intelligence*. 2017.

Paralelamente à construção de estratégias e políticas de IA nos planos doméstico e internacional, observa-se, ainda, o movimento crescente em torno da elaboração de princípios, códigos de conduta, diretrizes e recomendações por parte de atores não estatais – especificamente organizações não governamentais e associações representativas de segmentos da indústria de tecnologia da informação, robótica e internet¹⁶¹.

Cita-se aqui o Instituto de Engenheiros Elétricos e Eletrônicos (IEEE), responsável pela elaboração de Princípios Gerais da Iniciativa Global sobre Ética de Sistemas Autônomos e Inteligentes¹⁶² e pela criação da Comunidade Aberta para Ética em Sistemas Autônomos e Inteligentes (OCEANIS); e também a *Open AI*, empresa de pesquisa sem fins lucrativos voltada para soluções em IA geral segura, tendo elaborado a Carta da Inteligência Artificial Aberta (*“Open AI Charter”*)¹⁶³

Já no âmbito de um sistema regulatório regional, optou-se por abordar, em maior profundidade, as discussões em pauta na União Europeia, uma vez que tratam especificamente o problema da lacuna de responsabilidade, cogitando-se da personificação de determinados sistemas de inteligência artificial capazes de atuação autônoma como possível solução para ele.

Em 16 de fevereiro de 2017, o Parlamento Europeu editou uma resolução com recomendações à Comissão de Direito Civil sobre Robótica (2015/2103-INL)¹⁶⁴. O documento aborda variados temas relativos ao emprego crescente de tecnologias de inteligência artificial, como o seu impacto sobre a educação e o mercado de trabalho, uso de veículos autônomos, propriedade intelectual, pesquisa e inovação. Advoga também pela criação de uma agência europeia para robótica e inteligência artificial, apta a prover a *expertise* técnica, ética e regulatória necessária ao tema.

Especificamente no que tange à responsabilidade civil, o documento destaca que, na medida em que a autonomia de sistemas de inteligência artificial aumenta, não mais será possível considerá-los como meras ferramentas nas mãos de outros atores (como seu fabricante, operador, proprietário ou usuário). Nesse sentido, questiona se a atual normativa da responsabilidade civil é suficiente, ou se será necessária a introdução de novos princípios e

¹⁶¹ POLIDO, Fabrício. Op. cit., p. 196.

¹⁶² IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Disponível em: <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>. Acesso em: 25 nov. 2019.

¹⁶³ Open Community for Ethics in Autonomous Systems – OCEANIS. Informações institucionais e estatuto da Comunidade podem ser encontrados em: <https://ethicsstandards.org/>. Acesso em: 25 nov. 2019.

¹⁶⁴ UNIÃO EUROPEIA. European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103-INL). Estrasburgo: Parlamento Europeu, 2017. Disponível em: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html?redirect. Acesso em: 30 out. 2019.

regras para lidar com esse novo cenário, em que as ações ou omissões de uma IA não podem ser diretamente rastreadas de volta a um ator humano específico.

Uma vez que, sob a normativa europeia vigente, um agente artificial não pode ser responsabilizado *per se* por tais atos ou omissões que resultem em danos para terceiros, solicita-se que a Comissão, quando da avaliação do impacto de seu futuro instrumento legislativo, explore todas as soluções normativas possíveis para a questão. Dentre elas, a criação de um status jurídico específico para tais agentes – o de “pessoas eletrônicas - com direitos e obrigações específicas, incluindo-se o dever de reparar qualquer dano que possam causar; e aplicando-se a agentes que tomem decisões inteligentes e autônomas, ou interajam com terceiros independentemente¹⁶⁵. Isto seria viabilizado através da criação de um sistema de seguro ou de um fundo de garantia, de modo a evitar que as vítimas de danos causados por uma IA não fiquem sem a devida compensação.

Nesse mesmo sentido, Vladeck aponta que

uma solução seria conceituar novamente essas máquinas autônomas e inteligentes como entidades com o *status* de ‘pessoa’, de acordo com a lei. Conferir ‘personalidade’ a essas máquinas resolveria a questão da agência; as máquinas se tornariam atores em seu próprio direito, e, juntamente com o novo estatuto jurídico, viriam novos encargos legais, incluindo o ônus do seguro próprio (tradução nossa)¹⁶⁶.

Dando continuidade aos esforços de consolidar uma posição de vanguarda no cenário global da inteligência artificial e reconhecendo a sua importância para o fortalecimento do Mercado Único Digital, em abril de 2018, a Comissão Europeia, instituição da União Europeia responsável pela realização de propostas normativas e pela implementação de decisões, emitiu uma Comunicação (COM/2018/237) na qual delineia uma abordagem visando impulsionar o investimento no campo da IA e definir orientações éticas para sua pesquisa e desenvolvimento¹⁶⁷.

¹⁶⁵ Note-se que a proposta normativa europeia limita sua abordagem apenas a questões envolvendo a inteligência artificial aplicada a robôs, ou seja, corpórea. Entretanto, deve-se ter em mente que a robótica constitui apenas uma das áreas de estudo do campo da inteligência artificial, que também inclui a possibilidade de criação de agentes artificiais incorpóreos. No presente trabalho, a discussão acerca da atribuição de personalidade será estendida à inteligência artificial considerada em todos os seus aspectos, seja ela corpórea ou não, desde que capaz de atuar de maneira autônoma.

¹⁶⁶ No original em inglês: “One solution would be to reconceptualize these autonomous, intelligent machines as entities with the status of a “person” under the law. Confering “personhood” on these machines would resolve the agency question; the machines become principals in their own right, and along with new legal status would come new legal burdens, including the burden of self-insurance” (VLADECK, David C. Op. cit., p. 150).

¹⁶⁷ UNIÃO EUROPEIA. *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Artificial Intelligence for Europe* (COM/2018/237). Bruxelas: Comissão Europeia, 2018. Disponível em:

Dentre os inúmeros temas abordados pelo documento, apresentam-se, novamente, as preocupações relativas à segurança e à responsabilidade civil no caso de danos, especialmente diante da possibilidade de tomada de decisão autônoma de uma IA. Desse modo, a Comissão propõe-se a investigar se as regras já estabelecidas de responsabilização são adequadas para enfrentar esses novos desafios, ou se eventuais lacunas deverão ser preenchidas – e afirma, ainda, que um elevado nível de segurança e mecanismos eficazes de reparação às vítimas contribuem para o estabelecimento de confiança do usuário e a aceitação social de tais tecnologias.

A proposta europeia de considerar agentes autônomos de inteligência artificial como *pessoas eletrônicas* visa, portanto, à criação de uma nova categoria de entes expressamente dotados, ainda que de maneira limitada, de capacidade jurídica, de forma a possibilitar a sua responsabilização direta em hipótese de dano. Aproxima-se, dessa forma, da personalidade em sentido puramente formal¹⁶⁸, compreendida como a aptidão genérica para figurar em um dos polos de uma relação jurídica, e que permite que entes abstratos como sociedades e associações sejam subjetivados pelo ordenamento jurídico.

Seguindo esta linha, pode-se definir o problema a ser investigado na presente pesquisa nos seguintes termos: (i) poderia um agente autônomo de inteligência artificial ser considerado pessoa, para fins de atribuição de responsabilidade? E, como consequência direta do problema principal, (ii) a quem poderá ser imputada a responsabilidade na hipótese de ocorrência de um dano causado por esse agente, sem interferência humana?

<https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>. Acesso em: 29 out. 2019.

¹⁶⁸ Para uma análise aprofundada da personalidade como valor, veja-se: STANCIOLI, Brunello. *Renúncia ao Exercício de Direitos da Personalidade: Ou como alguém se torna o que quiser*. Belo Horizonte: Del Rey, 2010.

CAPÍTULO II

AGENTES ARTIFICIAIS AUTÔNOMOS COMO PESSOAS JURÍDICAS (ELETRÔNICAS)

1. O debate acerca da personificação de agentes artificiais autônomos

Embora tenha ganhado fôlego recentemente com a Resolução 2015/2103-INL do Parlamento Europeu, a problemática sobre a atribuição de personalidade a sistemas de inteligência artificial dotados de certo grau de autonomia é passível de ser encontrada na literatura especializada desde o início da década de 1990.

À época, com o advento da internet e a crescente utilização de máquinas e sistemas totalmente automatizados em papéis tradicionalmente desempenhados por seres humanos, como caixas eletrônicos e *bots* de compra e venda online, passou-se a indagar se tais sistemas poderiam, na medida em que fossem se aperfeiçoando e ganhando autonomia, continuar sendo considerados como meras ferramentas nas mãos dos seres humanos que lhes empregavam, ou se novas categorias jurídicas seriam necessárias para lidar com a situação nascente.

Nesse sentido, conforme preconizado por Leon Wein em 1992,

o século XXI confrontará a lei com um mundo em que a interação com os computadores é a regra e não a exceção. As últimas décadas deste século estão testemunhando um crescimento sem precedentes na utilização de artefatos autônomos e altamente sofisticados em praticamente todas as atividades de nossa vida cotidiana, e a substituição do homem por máquina está se acelerando à medida que a experiência tecnológica é cada vez mais dedicada a essa tarefa (tradução nossa)¹⁶⁹.

O autor continua ressaltando que os resultados obtidos por computadores são tão propensos a erros quanto qualquer decisão humana e, como consequência, espera-se que o ordenamento jurídico distribua a responsabilidade pelos equívocos ocorridos entre o proprietário da máquina, seu designer, seu programador, e outros possíveis réus. E questiona

¹⁶⁹ WEIN, Leon E. The Responsibility of Intelligent Artifacts: Toward an Automation Jurisprudence. *Harvard J.L. & Tech*, v. 6, n. 103, p. 103-154. No original: “The twenty-first century will confront the law with a world in which interaction with computers is the rule rather than the exception. The last decades of this century are witnessing unprecedented growth in the utilization of highly sophisticated, unattended artifacts in virtually every activity of our daily life, and replacement of man by machine is accelerating as technological expertise is increasingly devoted to this task”.

se seria de fato indispensável a existência de um “mestre” humano para que se atribua responsabilidade a um sistema automatizado¹⁷⁰.

Para Wein, ao se atribuir deveres a um dispositivo inteligente autônomo, seria suficiente que as interações com os mesmos resultem em consequências e resultados legais idiossincráticos, que desviam daqueles decorrentes de transações análogas realizadas sem intermediários automatizados. Desse modo, como esses dispositivos geram responsabilidades de uma ordem ou grau diferente dos humanos que executam tarefa equivalente, eles poderiam ser vistos mais apropriadamente como espécies de pessoas jurídicas “incompletas”¹⁷¹.

O autor, ao analisar a jurisprudência norte americana da época envolvendo autômatos primitivos, como armários que funcionavam a base de moedas, distribuidoras de ingressos e máquinas de venda, destaca que o intermédio de uma dessas máquinas era capaz de aumentar, diminuir ou até mesmo isentar a responsabilidade atribuída em comparação com transações não automatizadas funcionalmente análogas. Conclui, então, que os tribunais concediam à instalação automatizada um grau significativo de autonomia legal, e que

A pedra de toque nesse tipo de caso é se a máquina deve substituir um ator humano e, nesse caso, se um maior ônus da responsabilidade deve ser imposto à rotina automatizada. Onde as máquinas substituem satisfatoriamente os seres humanos, o relacionamento legal entre consumidor e máquina deve refletir o que o consumidor tinha anteriormente com o agente humano. No entanto, há casos em que um maior ônus da responsabilidade é atribuído a um processo automatizado porque o processo carece de um análogo adequado para alguma resposta humana crítica indispensável ao bom funcionamento do sistema (tradução nossa)¹⁷².

Deveria, portanto, existir a possibilidade de se processar tais máquinas diretamente, vez que, em certas ocasiões, a decisão (humana) de programá-las ou projetá-las de uma certa maneira pode não ser negligente, enquanto que a decisão (do computador) de se comportar de determinado modo em determinada situação seria considerada negligente, se a máquina pudesse ser processada¹⁷³.

É interessante notar que, em 1979, em artigo intitulado “*Are there decisions computers should never make?*”¹⁷⁴, James Moor levanta a possibilidade de que algum dia

¹⁷⁰ Id., p. 105.

¹⁷¹ Id., p. 107.

¹⁷² Id., p. 130. No original: “The touchstone in this type of case is whether the machine should be permitted to replace a human actor and, if so, whether a greater burden of liability ought to be imposed on the automated routine. Where machines satisfactorily substitute for humans, the legal relationship between consumer and machine should mirror that which the consumer formerly had with the human agent. Yet, there are instances in which a greater burden of liability is assigned to an automated process because the process lacks an adequate analogue for some critical human response indispensable to the proper functioning of the system”.

¹⁷³ Id., p. 114.

¹⁷⁴ MOOR, James H. Are there decisions computers should never make? *Nature & System*, v. 1, n. 4, p. 217-229.

existiriam computadores que serviriam não apenas como ferramentas de cálculo e consulta, mas tomariam, por si só, decisões relevantes acerca de assuntos envolvendo o bem estar humano. Para o autor, “tomar uma decisão” abrangeria os processos que levaram a determinada escolha, e não apenas à decisão em si. Significaria investigar possíveis cursos de ação, avaliando estratégias diversas e selecionando o melhor caminho a seguir com base na investigação e análise realizadas.

Imagine-se, por exemplo, uma máquina que monitore a condição dos pacientes sob terapia intensiva para, posteriormente, administrar medicamentos. Essa máquina pode não ser capaz de verificar tantas variáveis quanto um médico humano, podendo administrar o medicamento errado em situações nas quais aquele administraria o medicamento correto. Não obstante, a necessidade de atenção ininterrupta e o alto custo do atendimento médico direto poderiam incentivar um hospital a usar a máquina de qualquer maneira, aceitando a possibilidade de algum resultado inadequado.

Entretanto, a máquina em questão não poderia ser responsabilizada diretamente pelos danos decorrentes do equívoco, uma vez que a responsabilidade deve estar ligada a um agente humano, como o programador, o proprietário da máquina ou quem quer que tenha delegado o poder decisório a ela. O poder para tomar decisões não se confundiria com a habilidade para tal, envolvendo estar na situação apropriada e possuir autoridade para a escolha. Assim, seria possível delegar o poder decisório a um computador - o que implicaria também em delegar controle, razão pela qual o autor argumenta que algumas decisões sobre assuntos relevantes não deveriam ficar a cargo de computadores.

Entretanto, mesmo que o tipo de computador tratado por Moor seja causalmente responsável pelas decisões que venha a tomar, ele não poderia ser considerado responsável por elas legalmente, tendo em vista que não são considerados pessoas. Seria necessário o conhecimento e engajamento com as normas relevantes, tratando-se de um processo interativo, e não simplesmente imbuído em destinatários passivos.

No que diz respeito à atribuição de responsabilidade por danos, factualmente, a causalidade engloba a questão de saber se existe ou não uma conexão histórica entre a conduta do acusado e o dano sofrido pelo reclamante. Estabelecer a causa factual de um dano é fator necessário, mas não suficiente para responsabilizar alguém. Para tanto, deve haver ainda uma causa imediata, ou seja, a conexão histórica entre a conduta ilícita do acusado e o

dano sofrido pelo reclamante deve ser forte o suficiente para justificar a atribuição de responsabilidade¹⁷⁵.

Para John Snapper, por outro lado, o posicionamento de Moor seria inadequado pois, tradicionalmente, considera-se responsável aquele que de fato tomou uma decisão – argumentando, assim, que se deveria considerar a possibilidade de imputação de responsabilidade diretamente à máquina no exemplo citado¹⁷⁶.

O autor destaca que o principal argumento colocado por Moor para negar essa responsabilização seria a impossibilidade de levar a juízo e, conseqüentemente, obter uma restituição pelo dano, da própria máquina. Mas ressalta que, muito frequentemente, corporações possuem capacidade para serem partes em processos de natureza cível, inexistindo maiores impedimentos em se considerar um computador como responsável civilmente do que uma empresa¹⁷⁷. Em ambos os casos, não se trata de um agente humano (apesar de que, na segunda hipótese, mesmo não sendo considerada como pessoa no sentido ordinário do termo, considera-se uma empresa como pessoa perante a lei).

Além disso, Snapper evidencia que, para que possa ser responsabilizado, basta que um computador seja capaz de pagar uma reparação ou multa, não havendo dificuldade conceitual em se imaginar alternativas jurídicas para tanto. Dando continuidade ao exemplo tratado, alega que o ordenamento deveria exigir dos hospitais que segurassem seus computadores de terapia intensiva contra ações judiciais, assim como o fazem com a equipe médica humana¹⁷⁸. Nota-se, desse modo, a proposta de contratação de seguros como forma de viabilizar a personificação de tais entes.

Ainda no que diz respeito à atribuição de personalidade a agentes artificiais autônomos, merece também destaque trabalho de Lawrence B. Solum publicado em 1992 intitulado “*Legal personhood for artificial intelligences*”¹⁷⁹. Nele, o autor analisa o tema através de uma série de experimentos mentais de forma a transfigurar a questão puramente teórica em termos pragmáticos – questionando-se, por exemplo, se seria possível conceder direitos fundamentais, como o direito à liberdade de expressão, à uma IA; ou então se a mesma poderia atuar como um administrador fiduciário.

¹⁷⁵ NOLAN, Donal. Causation and the Goals of Tort Law. In: ROBERTSON, Andrew; TANG, Hang Wu (eds.). *The Goals of Private Law* Donal Nolan, “Causation and the Goals of Tort Law”, in *The Goals of Private Law*. Oxford: Hart Publishing, 2009, p. 165-190.

¹⁷⁶ SNAPPER, John W. Responsibility for Computer Based Errors. *Metaphilosophy*, v. 16, n. 4, 1985, p. 289-295.

¹⁷⁷ Id., p. 290.

¹⁷⁸ Idem.

¹⁷⁹ SOLUM, Lawrence B. Legal Personhood for Artificial Intelligences. *North Carolina Law Review*, v. 70, n. 4, 1992, p. 1231-1287.

No que tange ao primeiro exemplo, Solum analisa os principais argumentos contrários à atribuição de certos direitos a princípio somente aplicáveis à pessoa humana também a esses entes. Três objeções são destacadas: a de que somente às pessoas naturais devem ser concedidos tais direitos; a de que uma IA não possui algum componente indispensável à personalidade (como consciência ou intencionalidade); e a de que tais entes, como criações humanas, jamais poderão ser considerados como mais do que propriedade de seres humanos¹⁸⁰.

O tema é interessante uma vez que tangencia questões atinentes à atribuição de personalidade também a outros entes, como algumas espécies animais, questionando o fundamento antropocêntrico por trás de determinadas objeções¹⁸¹. Entretanto, diante do atual estado da arte, discussões sobre o reconhecimento de uma inteligência artificial como similar a uma pessoa humana são meramente especulativas, e não compreendem o escopo deste trabalho.

Para a discussão ora proposta, possui maior relevância o segundo exemplo apresentado, qual seja, a personificação de uma IA que atuasse como administrador fiduciário: o “cenário especula sobre as consequências legais do desenvolvimento de um sistema especialista capaz de fazer as coisas que um administrador humano pode fazer” (tradução nossa)¹⁸², denominado por Solum como o problema da competência.

No exemplo, tal sistema evoluiria em três estágios. No primeiro, o programa apenas auxiliaria um administrador humano na administração de um grande número de fundos, este sim responsável por tomar as decisões finais sobre como investi-los (ainda que baseie tais escolhas nas indicações do programa). Mas o desempenho real das tarefas diárias é amplamente automatizado, realizado pelo programa sem a necessidade de intervenção humana.

No segundo estágio, são desenvolvidos sistemas especialistas que superam os seres humanos como investidores em valores mobiliários negociados em bolsa. Os administradores passam, portanto, a ser instruídos a seguir os conselhos da IA ao tomar decisões de investimento em relação aos fundos. À medida em que as capacidades do sistema aumentam, a necessidade do administrador humano para a tomada de decisões diminui gradativamente.

¹⁸⁰ Id., p. 1258.

¹⁸¹ Sobre o tema, ver: MICHALCZAK, Rafal. *Animal's Race Against the Machines*. In: KURKI, Visa A. J.; PIETRZYKOWSKI, Tomasz (Eds.). *Legal Personhood: Animals, artificial intelligence and the unborn*. Cham (ZG): Springer, 2017, p. 91-101.

¹⁸² Na versão original: “The first scenario speculates about the legal consequences of developing an expert system capable of doing the things a human trustee can do” (SOLUM, Lawrence. Op. cit., p. 1240).

Como consequência, ele passa a ser capaz de supervisionar uma quantidade ainda maior de fundos, envolvendo-se apenas em casos excepcionais em que seja necessário tomar uma decisão não prevista, como a ocorrência de um processo judicial. Com o tempo, porém, o sistema também passa a lidar com esses tipos de tarefa.

Finalmente, o terceiro e último estágio tem início quando o depositário do fundo de investimentos decide encerrar o envolvimento humano em suas atividades (talvez visando economizar com as taxas relacionadas à sua atuação, ou talvez porque administradores humanos insistam em não seguir as instruções do programa, resultando em maus investimentos). Nessa hipótese, Solum questiona o que aconteceria se o depositário tentasse tornar o próprio programa um administrador, rebatendo duas objeções ao tema.

A primeira delas pode ser resumida na suposta incapacidade de uma inteligência artificial para fazer julgamentos e exercer discricção. Por exemplo, o autor ressalta que a legislação prevê a possibilidade de um administrador fiduciário desviar dos termos inicialmente acordados em caso de mudanças imprevistas capazes de prejudicar o próprio propósito do fundo em questão¹⁸³.

Em hipóteses como esta, ele destaca a impossibilidade de se programar o sistema de forma a antecipar todas as possíveis mudanças circunstanciais que possam vir a ocorrer (lembre-se que o exemplo trata de um sistema especialista, e não de aprendizado de máquina), sendo eventualmente necessária a intervenção de uma pessoa natural quando um julgamento discricionário fosse exigido.

Em linhas gerais, o autor considera não haver sentido afirmar que, dentre os milhares de fundos administrados pelo sistema autonomamente, o verdadeiro administrador seria considerado o ser humano que não possui qualquer influência sobre as ações normais de cada fundo, mas apenas pode vir a ser chamado a intervir em situações de anomalia. Isso porque, para Solum, a questão da atribuição de personalidade à IA deveria ser analisada por um viés prático, verificando se o seu tratamento como administradora fiduciária traria consigo alguma vantagem – o que é inegável.

A segunda objeção, a seu turno, diz respeito à responsabilidade, ou seja, como a IA do exemplo poderia ser responsabilizada por qualquer perda ou depreciação no valor do fundo como resultado de uma quebra de confiança, como não exercer as suas atividades com a habilidade e cuidado necessários. Nesse sentido, o autor destaca que

¹⁸³ Id., p. 1249.

Atualmente, a lei possui um mecanismo para atribuir responsabilidade no caso de um sistema especialista com mau funcionamento: o fabricante do sistema pode ser responsabilizado pela responsabilidade do produto. Mas a própria IA poderia ser responsabilizada? Existe uma maneira pela qual uma IA pode ter a capacidade de ser responsabilizada por danos, apesar da falta de bens pessoais. Ela pode adquirir um seguro (...) Se a IA pudesse ser segurada, a um custo razoável, o risco de ser considerada responsável por violar o dever de exercer cuidados razoáveis, então, funcionalmente, ela seria capaz de assumir tanto este dever quanto a responsabilidade correspondente (tradução nossa).¹⁸⁴

Veja-se, portanto, que assim como no trabalho de Snapper, como na Resolução do Parlamento Europeu, a atribuição de personalidade a um sistema artificial autônomo emerge, em Solum, como possível solução ao problema de se atribuir responsabilidade pelos danos causados pelo sistema, opção a ser viabilizada através da contratação de um seguro.

Nota-se, pelo exposto, a atribuição de personalidade a determinados sistemas artificiais autônomos como possível solução para salvaguardar os interesses das pessoas envolvidas nas interações com os mesmos, evitando a sua responsabilização direta pelas decisões e ações por eles tomadas autonomamente, bem como a não restituição pelos danos causados. Entretanto, a personificação aparece também em trabalhos que buscaram lidar com outro problema: o de se contabilizar a formação de contratos eletrônicos realizados por softwares.

Conforme apontado por Tom Allen e Robin Widdison em 1996, no âmbito do comércio online, cada vez mais máquinas seriam não apenas capazes de facilitar a comunicação entre os agentes envolvidos, mas também de dar início a essa comunicação. No nível mais básico, um computador poderia ser programado automaticamente para emitir uma oferta padrão e para reconhecer e registrar a aceitação de parceiros comerciais. Seu papel, portanto, estaria evoluindo rapidamente de uma instrumentalidade passiva para o de participante ativo no processo de negociação.¹⁸⁵

Para usufruir desses benefícios, os comerciantes humanos que empregavam máquinas autônomas no processo de negociação deveriam ter certeza de que as transações por elas geradas seriam válidas e eficazes perante a lei. Caso contrário, seria necessário que eles pessoalmente autorizassem – ou, ao menos, ratificassem – todas as transações individuais

¹⁸⁴ No original: “The law currently has a mechanism for assigning liability in the case of a malfunctioning expert system: the manufacturer of the system may be held responsible for product liability. But could the AI itself be held liable? There is a way in which an AI might have the capacity to be liable in damages despite the lack of personal assets. The AI might purchase insurance (...) If the AI could insure, at a reasonable cost, against the risk that it would be found liable for breaching the duty to exercise reasonable care, then functionally the AI would be able to assume both the duty and the corresponding liability” (Id., p. 1245).

¹⁸⁵ ALLEN, Tom; WIDDISON, Robin. Can Computers Make Contracts. *Harv. JL & Tech.*, v. 9, 1996, p. 25-52.

realizadas por um computador, perdendo-se, assim, as vantagens que justificaram a automatização do processo em primeiro lugar.

Ocorre que, para que um contrato seja legalmente exequível, deve haver, dentre outros elementos necessários, um acordo entre as partes envolvidas (*consensus ad idem*). Quando determinada transação é iniciada, negociada e executada sem que uma ou ambas as partes sequer tenham conhecimento, ao menos previamente, da mesma, como seria possível justificar a existência de um acordo?¹⁸⁶

Nessa situação, poder-se-ia considerar o sistema responsável pela transação como mera “ferramenta contratual”, vinculando as pessoas, naturais ou jurídicas, que empregaram aquele método particular de fazer negócios, e atribuindo a elas a vontade manifestada pelo sistema - nas palavras de Benjamin Allgrove, uma espécie de “responsabilidade vicária contratual”.¹⁸⁷

Alternativamente, seria possível tratar o sistema como uma “ferramenta comunicacional”, isto é, um mecanismo responsável pela transmissão das intenções das pessoas envolvidas na transação. Aqui, também, o elemento consensual estaria baseado em uma ficção sempre que os termos da transação fossem formados autonomamente pelo computador, sem qualquer envolvimento das partes¹⁸⁸.

Uma possível alternativa seria considerar o relacionamento entre o sistema responsável por conduzir a transação e a pessoa em nome de quem ele atua como uma relação jurídica de agência. Conforme visto anteriormente, um agente possui a autoridade e a capacidade de criar relações jurídicas entre aquele que lhe conferiu poderes para tanto (“*principal*”) e terceiros.

Assim, a aplicação de aspectos do direito de agência forneceria um conjunto útil de princípios que, operando conjuntamente com fundamentos contratuais, poderiam auxiliar na determinação se houve de fato uma manifestação de consentimento em transações eletrônicas específicas¹⁸⁹.

Ao tornar mais transparentes e menos problemáticas as maneiras pelas quais sistemas artificiais podem constituir relações jurídicas em nome de seus mandantes, considerá-los

¹⁸⁶ ALLGROVE, Benjamin D. *Legal Personality for Artificial Intellects: Pragmatic solution or science fiction?* 2004, p. 1-101. Disponível em: <https://ssrn.com/abstract=926015>. Acesso em: 12 jan. 2020.

¹⁸⁷ Id., p. 14.

¹⁸⁸ ALLEN, Tom; WIDDISON, Robin. Op. cit., p. 32.

¹⁸⁹ KERR, Ian R. *Providing for autonomous electronic devices in the Uniform Electronic Commerce Act*. Uniform Law Conference of Canada, 1999. Disponível em: <https://www.ulcc.ca/en/1999-winnipeg-mb/359-civil-section-documents/362-providing-for-autonomous-electronic-devices-in-the-electronic-commerce-act-1999>. Acesso em: 28 mar. 2020.

como agentes legais ajudaria a informar um conjunto consistente de normas aplicáveis a essas situações¹⁹⁰.

Ocorre que, embora considerados como agentes no sentido ordinário do termo ou no jargão utilizado na ciência da computação, legalmente falando tais sistemas não o são – para tanto, é necessário se tratar de uma pessoa, natural ou jurídica. Isso não significa dizer que o agente deva ter capacidade jurídica para contratar – por exemplo, um menor de idade pode atuar como um agente -, mas deverá ter personalidade para que possa se sujeitar aos direitos e deveres constituídos a partir de uma relação de agência¹⁹¹.

Assim, para que pudessem ser qualificados legalmente como agentes, seria necessário alterar o status desses sistemas, de forma a vincular diretamente seus usuários às transações por eles acordadas.

Hoje, contratos realizados através da interação de sistemas automatizadas são não apenas reconhecidos mas também juridicamente vinculantes, tendo sido resolvida a controvérsia com a adoção da interpretação dos mesmos como simples ferramentas contratuais. O artigo 12 da Convenção das Nações Unidas sobre a Utilização de Comunicações Eletrônicas em Contratos Internacionais, adotada em novembro de 2005, determina que

Um contrato formado pela interação entre um sistema automatizado de mensagens e uma pessoa natural, ou pela interação entre sistemas automatizados de mensagens, não deverá ser considerado inválido ou inexecutável pelo simples fato de que nenhuma pessoa natural reviu ou interveio em cada uma das ações individuais efetuadas pelo sistema automatizado de mensagens ou o contrato resultante.¹⁹²

E, em nota explicativa emitida pelo Secretariado da UNCITRAL acerca da referida Convenção, estabelece o princípio de que a pessoa, natural ou jurídica, em cujo nome um computador foi programado, deve ser responsável por qualquer mensagem gerada pela máquina:

O artigo 12 é uma disposição habilitante e não deve ser interpretado, de forma equivocada, como a permitir que um sistema de mensagens automatizado ou um computador seja objeto de direitos e obrigações. As comunicações eletrônicas que são geradas automaticamente por sistemas de mensagens ou computadores sem intervenção direta humana devem ser consideradas como “originárias” da entidade jurídica em nome da qual o sistema de mensagens ou computador é operado. As

¹⁹⁰ CHOPRA, Samir; WHITE, Laurence F. Op. cit., p. 22.

¹⁹¹ Allgrove, Benjamin D. Op. cit., p. 16.

¹⁹² NAÇÕES UNIDAS. Convenção sobre a Utilização de Comunicações Eletrônicas em Contratos Internacionais, de 13 novembro de 2005. [Trad.] Bernardo Santos; Pedro Martini; Pedro Soares. Disponível em: http://www.cisg-brasil.net/doc/Traducao_convencao_comunicacoes_eletronicas.pdf. Acesso em: 01 dez. 2019.

questões relevantes à agência que possam surgir nesse contexto devem ser resolvidas de acordo com regras fora da Convenção (tradução nossa).¹⁹³

Garante-se, assim, que uma negociação eletrônica travada por um sistema autônomo seja considerada perfeita, reconhecendo como válida a manifestação de vontade por ela exarada, bem como as obrigações decorrentes, sem, contudo, atribuir a ele personalidade jurídica.

A responsabilidade pelos seus atos é imputada à pessoa em cujo nome o sistema agiu, em conformidade com a regra geral de que o operador de uma ferramenta é responsável pelos resultados obtidos pelo seu uso, uma vez que esta não possui volição independente própria¹⁹⁴.

Atualmente, existem inúmeros sistemas contratuais automatizados em operação – que variam desde a venda de produtos online a consumidores até operações financeiras em alta frequência. Tais sistemas são instruídos a ofertar, negociar e concluir transações, e celebram contratos em nome de pessoas jurídicas reconhecidas. Existem, contudo, exceções: por exemplo, sob a lei japonesa, nem todos esses contratos serão válidos, conforme aponta Fumio Shimpo: se a IA falhar em se identificar como tal e induzir uma pessoa a celebrar um contrato, este poderá ser invalidado, conforme preconiza a codificação civil do Japão.¹⁹⁵

Note-se que, quando os trabalhos supramencionados foram escritos, o mundo ainda estava em meio ao segundo “inverno” do campo da inteligência artificial e, nas duas décadas seguintes, a ideia de se atribuir personalidade a uma IA continuou sendo apenas um experimento mental. Porém, em vista dos desenvolvimentos recentes de suas capacidades e seu crescente uso, tem-se, agora, um momento apropriado para reconsiderar a proposta.

Para tanto, faz-se necessário analisar as razões pelas quais se atribui o status jurídico de pessoa também a entes não humanos, de modo a verificar se, em primeiro lugar, seria possível fazer o mesmo em relação a uma inteligência artificial e, em seguida, se isto seria desejável.

¹⁹³ Seção explicativa n. 213: “Article 12 is an enabling provision and should not be mis-interpreted as allowing for an automated message system or a computer to be made the subject of rights and obligations. Electronic communications that are generated automatically by message systems or computers without direct human intervention should be regarded as ‘originating’ from the legal entity on behalf of which the message system or computer is operated. Questions relevant to agency that might arise in that context are to be settled under rules outside the Convention”.

¹⁹⁴ ČERKA, Paulius; GRIGIENĖ, Jurgita; SIRBIKYTĖ, Gintarė. Liability for damages caused by artificial intelligence. *Computer Law & Security Review*, v. 31, n. 3, 2015, p. 376-389.

¹⁹⁵ SHIMPO, Fumio. The Principal Japanese AI and Robot Strategy Toward Establishing Basic Principles. In: PAGALLO, Ugo; BARFIELD, Woodrow. *Research Handbook on the Law of Artificial Intelligence*. Northampton: Edward Elgard Publishing, 2018.

2. As razões da pessoa jurídica

Em obra dedicada à história do direito comercial, Francesco Galgano destaca que este nasce, como direito distinto daquele responsável pela regulação das relações não comerciais, apenas no século XVII, como *ius mercatorum* ou *lex mercatoria*: não apenas porque regula a atividade dos comerciantes da época, mas principalmente por ser um direito por eles criado, a partir dos estatutos das corporações mercantis, seus costumes e sua jurisprudência.¹⁹⁶

Surge, portanto, sem mediação estatal, como direito imposto em nome de uma classe emergente - e não mais em nome de toda a comunidade – a qualquer um que ingressasse em uma relação com um comerciante. O direito comercial se destaca como direito particular de uma nova classe social, em uma sociedade já habituada a múltiplos direitos particulares, de classes ou de terras, juntamente com direitos que aspiram ser universais: o direito Romano, que os glosadores extraem do *Corpus Iuris* de Justiniano; o direito canônico, que a Igreja impõe a todos em nome da majestade do pontífice¹⁹⁷.

Em relação aos demais, o *ius mercatorum* desempenha posição de concorrência, devendo ser aplicado, em detrimento daqueles, sempre que ao menos uma das partes integrantes da relação jurídica fosse pertencente à classe mercantil. Na ausência desse pressuposto, o relacionamento com os outros direitos existentes passa a ser de coexistência.

Nem sempre, porém, a aplicação da *lex mercatoria* era mais vantajosa ao comerciante, sobretudo quando as normas do direito romano, mais benéficas ao indivíduo, pudessem comprometer a concretização dos interesses da classe comerciante, qual seja, o de máxima valorização do capital comercial. Por exemplo, em se tratando dos bens destinados à atividade mercantil, o direito de propriedade perde a característica de *dominium*, que expressava a mais ampla faculdade de disposição, e imprime a tais bens um vínculo de destinação, do qual o comerciante, embora proprietário, não pode desviar, e o qual os seus credores pessoais devem respeitar¹⁹⁸.

O comerciante não pode, assim, oferecer ativos da empresa como garantia aos credores de negócios não relacionados com a mesma, vinculados juridicamente à garantia dos créditos a ela relativos; não pode, desvalorizando o capital comercial, comprometer o interesse da classe mercantil na distinção, existente até hoje, entre os credores sociais e os credores particulares dos sócios.

¹⁹⁶ GALGANO, Francesco. *Lex Mercatoria*. Bologna: Il Mulino, 2001.

¹⁹⁷ Id., p. 38.

¹⁹⁸ Id., p. 45.

A nova classe que se consolidava buscava satisfazer exigências de valorização da riqueza, concebida, segundo o nascente espírito capitalista, como ferramenta para a criação de mais riqueza. Para tanto, era necessário regular não apenas as relações externas à classe mercantil, de maneira vantajosa para os comerciantes; mas também regular, a seu modo, as relações existentes no interior da própria classe, entre comerciantes.

Diante da inadequação da figura da *societas* romana relativamente aos interesses da classe mercantil, é criado um novo tipo de sociedade - a *societas mercatorum*, designada nos estatutos dos comerciantes como companhia, e correspondente à sociedade em nome coletivo.

Sua característica distintiva era a responsabilidade ilimitada e solidária de todos os sócios pelas obrigações da sociedade: a obrigação assumida por um único sócio, ainda que sem o conhecimento dos demais, caso o fosse feita em nome da sociedade, ou vinculada ao seu objeto, era uma obrigação pela qual todos os sócios poderiam ser chamados a responder – modelo de administração societária que posteriormente veio a ser denominada disjuntiva.

Galgano destaca que o interesse protegido pela regra de responsabilidade ilimitada e solidária dos sócios era o de ampliar os negócios: o crédito dos comerciantes na sociedade aumentava, e a conclusão dos negócios era facilitada, se a terceiros fosse oferecida como garantia uma pluralidade de patrimônios, ao invés de todo o patrimônio de qualquer um dos sócios. Além disso, a responsabilidade ilimitada e solidária dos sócios englobava também aqueles ocultos – sócios desconhecidos por terceiros e cujo patrimônio, de outro modo, restaria inalcançável.¹⁹⁹

Diferenciava-se, portanto, do princípio da administração conjuntiva, que vigorava no direito romano e disciplinava as sociedades civis, segundo o qual era requerido o acordo unânime dos sócios em todas as operações da sociedade para que estas fossem válidas. As obrigações assumidas por um único sócio eram apenas por ele suportadas perante terceiros, salvo existência de um mandato expresso conferido pelos demais sócios. Essa forma de administração protegia, em última instância, a autonomia individual e o direito de propriedade, e não o interesse capitalista na circulação de riqueza:

Era uma regra que satisfazia a necessidade de preservação da riqueza: o direito romano protegia em grau máximo (...) a autonomia individual e, por meio dela, o direito à propriedade: princípio segundo o qual o sócio não podia ser obrigado em relação a terceiros se não por sua própria vontade (resultante da sua participação direta na conclusão do negócio ou pelo mandato conferido pelos outros sócios),

¹⁹⁹ Id., p. 47.

porque somente ele, enquanto proprietário, poderia alienar seus próprios bens (tradução nossa).²⁰⁰

A regra da responsabilidade limitada dos sócios de uma sociedade, assim como outras características distintivas das pessoas jurídicas que hoje se conhece, veio a surgir apenas no final da Idade Média, em um cenário político dominado por Estados monárquicos, com uma grande base territorial, nacional ou regional. Com o desaparecimento das soberanias comunais, a burguesia acaba por perder qualquer força de direção política, empreendendo, posteriormente, uma recuperação lenta e contrastada dentro das instituições monárquicas, como terceiro Estado no parlamento dos reinos.

Nesse contexto de consolidação dos Estados-nação, a classe mercantil deixa de ser a criadora de suas próprias normas, e o direito comercial passa por uma dupla transformação, não mais sendo um direito de classe, mas sim do Estado (direito público); e não mais sendo direito universal, mas nacional. Suas fontes são as leis emanadas pelo Estado soberano, operantes dentro das fronteiras nacionais, e cambiantes de Estado a Estado. A jurisdição comercial passa das antigas magistraturas mercantis, internas às corporações, aos tribunais estatais²⁰¹.

Não obstante, conforme destaca Galgano, a perda de poder político da classe mercantil não resultou na diminuição do ritmo de acumulação de capital comercial. Pelo contrário, esta passou a se intensificar como resultado da centralização de poderes nas mãos da monarquia, que atinge a sua forma extremada no Estado absoluto.

A política absolutista encontra na expansão do tráfego comercial um de seus principais pilares, tirando maior utilidade do mesmo a partir da coleta de tributos de comerciantes que ingressam e deixam o reino, e encontrando acumulado nas mãos de banqueiros e comerciantes as imensas somas de dinheiro necessárias para a consecução de suas grandes empreitadas²⁰².

De fato, desde o seu surgimento, as interações da classe mercantil com as demais classes detentores de riqueza eram, muitas das vezes, de cooperação, inspiradas por necessidades mútuas: por um lado, as classes dominantes precisavam investir sua renda de maneira lucrativa; e por outro, a classe mercantil tinha necessidade de obter o capital

²⁰⁰ Id., p. 46. Na versão original, em italiano: “Era una regola che soddisfaceva esigenze di conservazione della ricchezza: il diritto romano proteggeva in massimo grado (...) l'autonomia individuale e, per il tramite di questa, il diritto di proprietà: il principio era quello secondo il quale il socio non poteva essere obbligato nei confronti dei terzi se non dalla propria volontà (risultante dalla sua diretta partecipazione alla conclusione dell'affare o dal mandato conferito dal consocio), perché egli solo, in quanto proprietario, poteva disporre dei propri beni”.

²⁰¹ Id., p. 74.

²⁰² SAVARY, Jacques. *Le paifait négociant*. Paris: Librairie Droz, 2011.

necessário para o tráfego comercial²⁰³. E, na medida em que a política mercantil estatal passa a exigir o máximo desenvolvimento do tráfego comercial, seu sucesso passa a representar também a fortuna da classe mercantil.

A convergência de interesses que a política econômica mercantilista determina resulta, durante todo o período do absolutismo monárquico, em múltiplas formas de aliança e colaboração – sendo emblemática de tal período A Companhia das Índias Ocidentais. Trata-se da primeira grande empresa da Era Moderna, arquétipo do que se tornará, com as codificações do século XIX, a forma jurídica típica da sociedade capitalista: a sociedade anônima²⁰⁴.

Nela, aparecerem pela primeira vez as características próprias deste tipo societário, a saber, a limitação da responsabilidade dos sócios e a divisão do capital social em ações. Embora a primeira dessas características fosse já comum à sociedade em comandita simples, ela passa a abarcar todos os sócios, e não apenas uma parcela destes – em outras palavras, a sociedade anônima passa a conceder o benefício da responsabilidade limitada também à classe mercantil.

Isso porque, nas sociedades em comandita simples, apenas uma categoria de sócios, os comanditários, excluídos da direção da sociedade empresária, possuíam responsabilidade limitada relativamente às obrigações por ela contraídas; aos demais, sócios comanditados, cabia necessariamente assumir ilimitadamente tal responsabilidade.

Ocorre que, à classe mercantil, qualquer que fosse a forma jurídica assumida – seja a de comerciante individual ou de sócio comanditário – não lhe era concedido o benefício da limitação de responsabilidade. Este era garantido apenas às demais classes detentoras de bens móveis e, em especial, às classes fundiárias, às quais era possibilitado investir parte de seu patrimônio nas atividades comerciais, sem correr nenhum outro risco além da perda de capital investido²⁰⁵.

A limitação da responsabilidade dos comerciantes através da sociedade anônima passou a ser, nos séculos XVII e XVIII, excepcionalmente concedida às companhias coloniais para explorar os recursos econômicos das recém descobertas colônias. Precisamente por seu caráter excepcional, a sociedade anônima era estabelecida por ato de concessão soberana, que por sua vez representava a fonte do privilégio da responsabilidade limitada de todos os sócios integrantes, bem como o de divisão do capital social em ações.

Desse modo, Galgano destaca que

²⁰³ GALGANO, Francesco. Op cit., p. 44.

²⁰⁴ Id., p. 68.

²⁰⁵ Id., p. 78.

A sociedade anônima nasce, no século XVII, com o caráter de uma espécie de “pacto” celebrado de tempos em tempos por grupos empresariais com o soberano: eles se declararam dispostos a empreender em iniciativas econômicas colossais, como a colonização de terras no exterior, a partir das quais o soberano espera obter o aumento do poder político e econômico do Estado, e nas quais a nobreza vê uma nova e lucrativa fonte de renda, mas que para a classe mercantil importa grandes riscos, desde que seja concedido pelo soberano um “privilégio”, a imunidade ao antigo princípio da responsabilidade ilimitada (tradução nossa)²⁰⁶.

A justificativa para tal incentivo concedido era, à época, as enormes proporções dos empreendimentos realizados que, conseqüentemente, traziam consigo também enormes riscos. O benefício da responsabilidade limitada incidia sobre a relação entre as proporções da iniciativa econômica empreendida e a extensão do risco: significava que as vastas proporções do empreendimento se tornaram um incentivo, e não mais um obstáculo, para os comerciantes envolvidos, pois as eventuais perdas da sociedade, por maiores que fossem, não afetariam o seu patrimônio senão dentro dos limites do capital investido.

A outra característica distintiva da nova figura corporativa, consistente na divisão do capital social em ações, era um incentivo não apenas para a classe mercantil, mas principalmente para as demais classes detentoras de riqueza. A qualidade de sócio passou a ser incorporada em um título livremente transferível - a ação -, cujo valor em dinheiro poderia ser facilmente obtido. Com isso, somava-se à vantagem de participação nos lucros sem assunção de riscos ilimitados, a capacidade de dispor da própria participação na empresa do mesmo modo como se dispõe de qualquer outro bem móvel²⁰⁷.

A ação torna-se, assim, o instrumento que permite à classe mercantil mobilizar, em torno de suas iniciativas, os recursos financeiros de outras classes sociais, como a nobreza e o clero. Esta forma de financiamento tinha o benefício de, diferentemente do que ocorria com empréstimos de banqueiros, o capital aportado pelos acionistas constituir *capital de risco*. Ou seja, se o empreendimento falhasse e o seu capital fosse perdido, nada seria devido aos acionistas que, enquanto tal, deveriam suportar os riscos da empresa.

A partir do trabalho de Francesco Galgano, é possível verificar que o recurso à personificação surge, na Modernidade, como um importante instrumento para a articulação

²⁰⁶ No original: “La società per azioni nasce, nel Seicento, con il carattere di una sorta di “patto” di volta in volta concluso da gruppi imprenditoriali con il sovrano: gli uni si dichiarano disposti ad intraprendere colossali iniziative economiche, quali la colonizzazione delle terre d’oltremare, dalle quali il sovrano si attende l’accrescimento della potenza politica ed economica dello Stato, ed alle quali la nobiltà guarda come ad una nuova e proficua fonte di rendite, ma che per la classe mercantile importano ingenti rischi, a condizione che venga concesso dal sovrano un “privilegio”, l’immunità dall’antico principio della responsabilità illimitata” (Id., p. 79).

²⁰⁷ Id., p. 80.

patrimonial, fazendo com que a separação de um patrimônio em virtude de uma determinada finalidade possa ser melhor compreendida a partir da criação de um novo sujeito. A viabilidade de uma estrutura patrimonial autônoma do ente abstrato, distinta da de seus membros, revela-se, assim, dependente do reconhecimento de uma nova subjetividade jurídica²⁰⁸.

Mas não apenas. A personificação possibilita, ainda, a constituição de um sistema de imputação direta dos atos praticados pelos órgãos da pessoa jurídica, através da mediação orgânica, desenvolvida em contraposição à representação. Enquanto nesta representante e representado são vistos como sujeitos distintos, o que possibilita considerar o ente coletivo indiferente aos atos praticados por seus agentes, através da mediação orgânica, o órgão passa a ser concebido como parte do próprio ente coletivo.

Possibilita-se, desse modo, um processo de imputação mais amplo: não se atribui à pessoa jurídica apenas os efeitos correlatos ao ato praticado, mas o ordenamento a reconhece como verdadeira autora do mesmo. Significa dizer que, ainda que o ato tenha sido materialmente realizado por uma pessoa natural, juridicamente é como se a própria pessoa jurídica o tivesse feito²⁰⁹.

O reconhecimento de uma nova subjetividade permite, ademais, simplificar situações jurídicas complexas com a constituição de um centro unitário de imputação, desempenhando a personalidade jurídica uma função heurística. Nesse sentido, ao simplificar este complexo de relações jurídicas, a expressão “pessoa jurídica” facilitaria a descrição do próprio fenômeno subjacente²¹⁰.

Isso porque não há nada que impeça, de antemão, que todos os integrantes de uma sociedade ou associação, por exemplo, relacionem-se conjuntamente com terceiros. Entretanto, a interposição de uma nova subjetividade permite que esse mesmo processo passe a ser descrito de forma unitária, sem a necessidade de menção, a cada ato praticado, de todos os sócios que se mostram retratados pelo termo pessoa jurídica²¹¹.

Ao contrário da relação entre dois indivíduos, a relação jurídica entre um indivíduo e uma corporação, enquanto pessoa jurídica, sempre poderá ser reduzida a outras inúmeras relações separadamente, com cada indivíduo por trás delas, ainda que as mesmas sejam combinadas em uma única relação jurídica por motivos de conveniência. Desse modo, uma

²⁰⁸ NEGRI, Sérgio Ávila. As Razões da Pessoa Jurídica e a Expropriação da Subjetividade. *Civilistica*, a. 5, n. 2, 2016, p. 1-18.

²⁰⁹ Id., p. 9.

²¹⁰ Id., p. 8.

²¹¹ Idem, p. 8.

definição corrente de personalidade jurídica refere-se à capacidade para integrar um dos polos da relação jurídica. Isso significa que um ente coletivo deve ter a capacidade para estabelecer, manter e efetivar relações jurídicas com outras pessoas, mantendo-se dentro dos limites legais²¹².

Nesse mesmo sentido, Bryson et al. afirmam que a criação de um discurso fictício segundo o qual determinados entes coletivos são pessoas constitui um atalho útil para lhes atribuir direitos e obrigações legais a princípio possuídos apenas por pessoas naturais dentro de um ordenamento jurídico – como o direito de vincular outras subjetividades através de um contrato, ou o dever de satisfazer as obrigações dele decorrentes. Sem esse amplo conjunto de direitos e deveres atualmente característicos da personalidade jurídica, os autores destacam que as empresas não poderiam ter evoluído em seu papel de verdadeiros motores do desenvolvimento econômico²¹³.

Para os autores, três observações se fazem necessárias quando da análise do tema de uma possível personalidade jurídica eletrônica. Primeiro, destacam, deve-se ter em mente que a personalidade jurídica constitui um artifício pois, ao ser atribuída a determinado ente, significa dizer que o sistema jurídico que o fez endereça suas regras a ele, tanto para conceder-lhe direitos quanto para sujeitá-lo a obrigações, transformando-o em um verdadeiro ator legal²¹⁴.

Nesse sentido, Otávio Rodrigues Jr. destaca que nas ordens jurídicas positivistas do século XIX, a personalidade emerge como um atributo reconhecido pelas ordens estatais aos seres humanos, mas nem sempre a todos e nem sempre apenas a eles:

Pessoa é aquele (ou aquilo) a quem o Estado reconhece a aptidão genérica para adquirir e exercer, por si ou por outrem, direitos e deveres na ordem jurídica, a partir de um fato gerador quase sempre associado ao nascimento com vida. Por essa razão, no primeiro grupo de ordens jurídicas, é possível mencionar seres humanos que não são pessoas (os escravos) ou que possuem restrições a sua capacidade de direito (as mulheres, em muitos ordenamentos dos séculos XIX e XX), além de um conjunto de bens afetados a um fim (as fundações) aos quais se defere a personalidade.²¹⁵

Para além dos exemplos supra mencionados por Otávio Rodrigues Jr., é interessante notar a existência de situações em que ordenamentos jurídicos reconheceram certos direitos a

²¹² Solaiman, S. M. Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy. *Artificial Intelligence Law*, n. 25, 2017, p. 155–179.

²¹³ BRYSON, Joanna J.; DIAMANTIS, Mihailis E.; GRANT, Thomas D. Of, for, and by the people: the legal lacuna of synthetic persons. *Artificial Intelligence Law*, n. 25, 2017, p. 237-291.

²¹⁴ Id., p. 277.

²¹⁵ RODRIGUES JR, Otavio Luiz. Pessoa, personalidade, conceito filosófico e conceito jurídico de pessoa: Espécies de pessoas no direito em geral. *Revista de direito do consumidor*, v. 27, n. 118, 2018, p. 281-291.

entes não humanos, como edifícios, objetos, e divindades. Por exemplo, no caso *Bumper Development Corporation Ltd v. Commissioner of Police of Metropolis*²¹⁶, do Reino Unido, o Tribunal Recursal britânico decidiu que um templo indiano, cuja personalidade era reconhecida no país de origem, poderia reivindicar direitos e ajuizar ações perante a lei inglesa. Similarmente, no caso norte-americano *Autocephalous Greek Orthodox Church of Cyprus v. Goldberg*²¹⁷, o Tribunal de Apelações do 7º Circuito dos EUA sustentou que um conjunto de mosaicos deveriam ser devolvidos a determinada igreja, considerada sua proprietária legal.

No caso *Pramatha Nath Mullick v. Pradyumnakumar Mullick*²¹⁸, também conhecido como o caso do ídolo hindu, à época julgado pelo mais alto tribunal indiano (*Judicial Committee of the Privy Council*), reconheceu-se que um ídolo da religião Hindu era, de acordo com os costumes e a autoridade estabelecida, uma “entidade jurídica”, capaz de processar e ser processada. Seus interesses, portanto, seriam atendidos pela pessoa que possuísse a divindade sob sua responsabilidade, atuando como seu representante legal, em analogia a situações em que se discutem os interesses de uma criança.

Mais recentemente, em 2011, um grupo de cidadãos equatorianos ajuizou uma ação legal bem-sucedida em nome do meio ambiente contra o governo provincial de Loja para interromper a expansão de uma estrada, a qual eles alegavam estar danificando uma importante bacia hidrográfica²¹⁹. Já em 2015, a Justiça argentina expediu um *Habeas Corpus* em nome de uma orangotango, reconhecendo-a como “pessoa não-humana” e, conseqüentemente, detentora de alguns direitos básicos, como o direito à liberdade²²⁰.

Verifica-se, desse modo, que a personalidade jurídica não está necessariamente correlacionada a uma noção metafísica ou ética de personalidade. Embora seja desejável que

²¹⁶ REINO UNIDO. Court of Appeals. *Bumper Development Corporation Ltd v. Commissioner of Police of Metropolis*, 1991. Disponível em: <https://lawin.org/bumper-development-corp-ltd-v-commissioner-of-police-of-the-metropolis-and-others/>. Acesso em: 01 abr. 2020.

²¹⁷ VAN PELT, Meredith. *Autocephalous Greek Orthodox Church of Cyprus v. Goldberg and Feldmen Fine Arts, Inc.: A case for the use of civil remedies in effecting the return of stolen art*. *Dickinson Journal of International Law*, v. 8, n. 3, 1990, p. 441-464. Disponível em: <https://elibrary.law.psu.edu/cgi/viewcontent.cgi?article=1126&context=psilr>. Acesso em: 01 abr. 2020.

²¹⁸ INDIA. Bombay High Court. *Pramatha Nath Mullick v. Pradyumnakumar Mullick*, 1925. Disponível em: <https://indiankanoon.org/doc/290902/>. Acesso em: 01 abr. 2020.

²¹⁹ GREENE, Natalia. *The First Successful Case of the Rights of Nature Implementation in Ecuador*. *The Rights of Nature*, 2011. Disponível em: <http://therightsofnature.org/first-ron-case-ecuador/>. Acesso em: 01 abr. 2020.

²²⁰ The Telegraph. *Sandra the Orangutan granted limited human rights*, 2014. Disponível em: <https://www.telegraph.co.uk/news/worldnews/southamerica/argentina/11307205/Sandra-the-orangutan-granted-limited-human-rights.html>. Acesso em: 01 abr. 2020. Ver também: SCHNEIDER, Kevin. *Judge Recognizes two chimpanzees as legal persons, grants them writ of habeas corpus*. *Non Human Rights*, 2015. Disponível em: <https://www.nonhumanrights.org/blog/judge-recognizes-two-chimpanzees-as-legal-persons-grants-them-writ-of-habeas-corpus/>. Acesso em: 01 abr. 2020.

um ordenamento jurídico tenha em mente tais conceitos, em diferentes momentos históricos conferiu-se personalidade a entes que estavam aquém ou além do conjunto de pessoas eticamente consideradas

A personalidade jurídica resulta da decisão de um ordenamento legal em reconhecer que determinado ente a possui. Para Bryson et al., ela poderia ser pensada como uma espécie de status fictício, que pode ser atribuído pelo ordenamento jurídico sempre que o fazer seja adequado para o alcance de seus objetivos.²²¹ Similarmente, Jacob Turner afirma ser a personalidade jurídica uma ficção, algo criado pelos seres humanos por meio de seus sistemas legais e, enquanto tal, cabendo a eles decidir a quem essa ficção deve se estender, e precisar o seu conteúdo²²².

A segunda observação a respeito do tema feita por Bryson et al. é de que a personalidade constitui um agregado de direitos e obrigações sendo, portanto, divisível. Isto quer dizer que as pessoas não precisam possuir idênticos direitos e obrigações, ainda que dentro de um mesmo sistema legal. Este poderá tratar certo ator como uma pessoa jurídica no que diz respeito a certos direitos e obrigações, mas não em relação a outros²²³.

Note-se que aquilo que os atores tratam como divisibilidade da personalidade equivale ao conceito adotado no ordenamento pátrio de capacidade de fato ou de exercício, variável entre diferentes pessoas e até mesmo em diversos estágios da vida de uma mesma pessoa. Diversamente, a personalidade constitui um conceito binário, no sentido de que ou se é, ou não se é pessoa, não sendo admitidas gradações nesse sentido.

Finalmente, os autores observam que a personalidade jurídica de um ator, mesmo que implique em uma vasta gama de direitos e obrigações, não revela, por si só, o seu envolvimento efetivo com o sistema jurídico. Ainda que ele possa ser o beneficiário de normas que lhe confirmam direitos ou o destinatário de obrigações impostas por outras pessoas, tal fato isoladamente não diz quais oportunidades o sistema lhe oferece para tirar proveito dessas normas, ou para que outros atores o responsabilizem por eventuais violações. Em suma, os direitos e obrigações que uma pessoa jurídica possui em virtude da lei podem não corresponder àqueles que ela detém por uma questão de fato²²⁴.

Dizer que a personalidade jurídica constitui uma “ficção” não significa que ela não produza efeitos reais. Pelo contrário, o propósito de se conferir personalidade a um certo ente

²²¹ BRYSON et al. Op cit., p. 277.

²²² TURNER, Jacob. *Robot Rules: Regulating Artificial Intelligence*. Nova York: Palgrave Macmillan, 2019, p. 175.

²²³ BRYSON et al. Op. cit., p. 277.

²²⁴ Idem.

é permitir que ele influencie de certas maneiras o ordenamento jurídico, e seja também por ele afetado.

Bryson et al. destacam que o discurso e a prática de reconhecer a pessoa jurídica se enquadram no tipo de estrutura que filósofos denominam ficcionalismo. Um domínio de discurso é ficcionalista se procura representar algo que não seja a verdade literal. Participantes de um discurso ficcionalista engajam-se (intencionalmente ou não) em uma espécie de pretexto ao assumir uma postura segundo a qual o conteúdo do referido discurso, embora literalmente falso, refere-se a entidades reais e descreve propriedades reais das mesmas.²²⁵

Quando se argumenta que esses domínios do discurso são fictícios, consequentemente assume-se o ônus de informar o motivo pelo qual ele é adotado – o que geralmente envolve explicar por que determinado discurso é útil. No caso das pessoas jurídicas, viu-se que a interposição de uma nova subjetividade em certas relações jurídicas apresenta uma função heurística, de articulação patrimonial, imputação de responsabilidade, organização, dentre outras.

Conforme dito anteriormente, cabe a todo ordenamento decidir a quais entidades atribuir personalidade, devendo fazer essa escolha tendo em mente seus objetivos basilares. A principal questão a ser enfrentada por um sistema jurídico quando da concessão de personalidade jurídica a determinada entidade é se esta decisão avança ou dificulta o alcance desses objetivos.

O processo que leva à criação de uma nova pessoa jurídica é descrito por Gunther Teubner como a observação externa de uma autodescrição: o sistema jurídico analisa, a partir de suas próprias conceituações, como o sistema social organizado observa a si próprio como uma coletividade, ou então como é observado como tal por terceiros em seu ambiente.²²⁶

Baseando-se em Luhmann, um ator coletivo não seria constituído pelo mero agrupamento de indivíduos, mas uma série de comunicações. Desse modo, sob a dupla condição de que uma cadeia de comunicações transmita informações sobre si mesma - isto é, crie uma autodescrição -, e que eventos comunicativos sejam atribuídos a essa autodescrição como suas ações, a realidade social de um ator coletivo emerge com suas próprias estruturas de tomada de decisão e efeito vinculante para o sistema social.²²⁷

²²⁵ Id., p. 278.

²²⁶ TEUBER, Gunther. Enterprise Corporatism: New industrial policy and the "essence" of the legal person. *The American Journal of Comparative Law*, v. 36, n. 1, 1988, p. 130-155.

²²⁷ TEUBNER, Gunther. Rights of non-humans? Electronic agents and animals as new actors in politics and law. *Journal of Law and Society*, v. 33, n. 4, 2006, p. 497-521.

O sistema jurídico, destaca Teubner, não está de maneira alguma vinculado pela autodescrição, e nem por observações externas da mesma (psicológicas, sociológicas ou econômicas, por exemplo). Por conseguinte, nada impede que ele tome qualquer objeto – divindades, santos, templos, terrenos, objetos de arte – como pontos de atribuição e lhes confira capacidade legal. Entretanto, é surpreendente o fato de que, não obstante o grau elevado de liberdade e escolha que possui, existam tamanhas semelhanças estruturais entre a “coletividade” e a “pessoa coletiva”:

Não é apenas que os mecanismos sociais e legais de ação atribuídos a construções produzidas dentro do sistema (indivíduo/coletividade no mundo social, pessoa natural/pessoa jurídica no mundo jurídico) sejam, em princípio, estruturados de maneira paralela. É mais impressionante que a lei hoje praticamente não faça uso de sua liberdade positivista e promova exclusivamente “coletividades” ao status de pessoas jurídicas (tradução nossa).²²⁸

Nesse contexto, o autor propõe que, mesmo sem ser obrigada normativamente a isso, a lei vincula regularmente a capacidade jurídica a um conjunto de pré-requisitos de uma determinada realidade social, que podem ser descritos por um “conceito unitário”. São eles: (1) um sistema de ação (formalmente) organizado; (2) uma autodescrição de uma identidade coletiva; e (3) uma ligação cíclica de identidade e ação através de mecanismos de atribuição.²²⁹

A razão para essa estreita correspondência, embora possa surpreender do ponto de vista das liberdades positivistas, não diz respeito a uma lei natural ou a uma lógica jurídica subjacente, mas sim a uma questão de política legal. Atribuir personalidade jurídica às formações sociais apenas faria sentido a partir do momento em que elas possuem uma ordem interna altamente desenvolvida. A capacidade social para ação, ou seja, a capacidade de atribuir efeitos externos ao sistema social como tal, significa que as ações de certos indivíduos são atribuídas a todos os participantes, conferindo-lhes direitos e vinculando-os socialmente.

Para tanto, Teubner destaca a necessidade de uma ordem que possua muitos pré-requisitos sociais, como o desenvolvimento de estruturas de liderança, a legitimação de regras de representação e processos distributivos com efeitos externos ou internos, e uma certa mitigação de atribuição pessoal. Esses requisitos podem ser resumidos na fórmula “capacidade social para ação coletiva”.

Via de regra, aponta o autor, faz sentido em termos de política legislativa conceder personalidade a sistemas sociais que já possuem capacidade social para ação coletiva. Ele cita

²²⁸ TEUBER, Gunther. *Enterprise...* Op cit., p. 141.

²²⁹ Id., p. 142.

como exemplares de tal fenômeno os partidos políticos e sindicatos que, uma vez que têm sua capacidade social para ação coletiva desenvolvida, passam a expor o ordenamento jurídico à pressão de completar o processo social de personificação através da concessão da personalidade jurídica²³⁰.

Allen e Widdison explicam que, no geral, a maioria das pessoas jurídicas é vista pelas demais como atores dentro de algum contexto extralegal. Uma vez que um grupo de pessoas naturais atinge certo nível organizacional, há a tendência de considerá-lo como uma unidade social e comunicativa. As pessoas de fora da organização a consideram como uma entidade independente de seus membros com a qual fazem negócios; seus membros, por vez, também a consideram como uma entidade diferenciada de si próprios. E assim, como consequência, o sistema legal também passa, eventualmente, a considerar a organização como verdadeira autora a qual comunicações legalmente significativas são atribuídas.²³¹

Por outro lado, conforme demonstrado pelos exemplos da máfia italiana ou das associações sem capacidade legal na Alemanha (“*Nichtrechtsfähiger Verein*” similar à figura da sociedade não personificada do ordenamento pátrio), podem haver importantes razões que militam contra a atribuição de personalidade jurídica a certos sistemas sociais, a despeito dos mesmos serem efetivamente capacitados para ação social.

A segunda razão por trás da estreita correspondência entre a estrutura social e jurídica diz respeito à função desempenhada pela personificação (ou, conforme previamente referido, às razões da pessoa jurídica). Para Teubner, o aspecto funcional é compreendido de maneira bastante inadequada se apenas as vantagens da responsabilidade limitada forem levadas em consideração, e as desvantagens parcialmente compensadas ao se “perfurar o véu corporativo”. Igualmente relevante seriam, por exemplo, as vantagens na coordenação de recursos, a economia proporcionada nos custos de transação, as vantagens posicionais para a organização de contratos com o meio, os ganhos em eficiência decorrentes do apoio do ordenamento ao sistema social, e sua adquirida perpetuidade.²³²

Contudo, a principal “propriedade emergente” do processo de personificação residiria na construção de um sistema autopoietico²³³ de segunda ordem:

²³⁰ Id., p. 143.

²³¹ ALLEN, Tom; WIDDISON, Robin. Op. cit. p. 38.

²³² TEUBNER, Gunther. Rights of... Op. cit., p. 6.

²³³ Segundo a teoria da autopoiese aplicada ao Direito de Teubner, os ordenamentos jurídicos desenvolvem novos significados legais a partir de seu estoque já existente de significados, assim como sistemas matemáticos desenvolvem novos conceitos a partir dos conceitos matemáticos já existentes. ALLEN, Tom; WIDDISON, Robin. Op. cit. p.37.

Pelo vínculo cíclico de identidade e ação aperfeiçoado na pessoa jurídica, a organização adquire uma autonomia até então não alcançada em relação ao seu ambiente, tanto o ambiente externo do mercado e da política quanto o ambiente interno dos membros e outros envolvidos na organização. A personificação legalmente apoiada é um passo decisivo para o fechamento operacional completo, o que ao mesmo tempo significa um novo tipo de abertura ambiental (tradução nossa).²³⁴

De acordo com essa teoria, um ordenamento jurídico trata entes não naturais como pessoas coletivas porque reconhece a sua capacidade para agir de alguma maneira extralegal, e que esta ação deve também conter um significado dentro do universo jurídico. Nesse sentido, Allen e Widdison exemplificam que, quando um Estado declara guerra contra outro, por óbvio tal declaração possui um significado fora do sistema jurídico. Entretanto, pode-se também levantar a questão da legalidade de referida declaração e, para tanto, ela deve ser atribuída a uma pessoa coletiva. Uma vez que, no sentido extralegal, a maioria das pessoas atribuiria a declaração de guerra ao próprio Estado, faz sentido para o ordenamento jurídico atribuir-lhe a declaração, devendo, necessariamente, tratá-lo como um sujeito de direito²³⁵.

A atribuição de personalidade jurídica possibilita a entidades não humanas, desse modo, entrar em negociações políticas de pleno direito e intrincadas transações econômicas. Tais entidades obtêm controle considerável sobre o ambiente, vinculando simultaneamente o seu processo interno de tomada de decisão. Os atores coletivos, especialmente quando formalmente reconhecidos pelo ordenamento, passam a formar suas próprias estratégias, preferências e interesses, não mais redutíveis aos de seus membros, sócios ou proprietários. A lei estabiliza as expectativas sociais sobre esses atores ao construí-los como pessoas, concedendo-lhes direitos e impondo-lhes deveres e responsabilidades.²³⁶

Aplicando a teoria de Gunther Teubner à análise da personificação de agentes de inteligência artificial, Allen e Widdison afirmam haver uma semelhança essencial entre atores coletivos e tais agentes. Apesar daqueles consistirem em sistemas sociais, enquanto que estes se apresentam como sistemas informacionais, a semelhança estaria no fato de que

(...) com as organizações, a busca é pela separação entre a identidade e as ações da organização e as de seus membros. Em outras palavras, é a capacidade social para ação coletiva que é material. Com os computadores, a busca também é pela separação da identidade e das ações do computador de seus controladores humanos. Obviamente, não despreveríamos isso como a capacidade social para ação coletiva.

²³⁴ TEUBNER, Gunther. *Enterprise...* Op. cit., p. 144. No original: "By the cyclical linkage of identity and action perfected in the legal person, the organization acquires a hitherto unachieved autonomy vis-à-vis its environment, both the external environment of market and politics and the internal environment of members and others involved in the organization. The legally supported personification is a decisive step towards complete operational closure, which at the same time means a new type of environmental openness".

²³⁵ ALLEN, Tom; WIDDISON, Robin. Op. cit. p. 38.

²³⁶ TEUBNER, Gunther. *Rights of...* Op. cit., p. 7-8.

No entanto, descreveríamos isso como a capacidade social para ação autônoma. Portanto, podemos traduzir o argumento de Teubner afirmando que “faz sentido em termos de política legal conceder capacidade jurídica a sistemas informacionais que já possuem capacidade social de ação autônoma” (tradução nossa)²³⁷

Essa capacidade social de ação autônoma surgiria, segundo os autores, quando aqueles que interagem com um agente artificial o consideram como a fonte de comunicações relevantes – e não os seus controladores humanos.

Nesse contexto, haveria uma diferença fundamental entre uma máquina de venda automática e um bot responsável por realizar transações autonomamente. No primeiro caso, apenas é possível afirmar superficialmente que a máquina vende algum produto, uma vez que é certo que o preço, meios de pagamento e entrega, quantidade e qualidade do mesmo são determinados por outras pessoas.

Por outro lado, em se tratando de computadores autônomos, Allen e Widdison reiteram que, a exemplo de computadores capazes de jogar xadrez (como o famoso Deep Blue), existe a inclinação de se afirmar que eles o fazem por si só; assim como bots responsáveis por transações online. Nesse sentido prático, extralegal, são atribuídas ações aos próprios computadores, havendo algum sentido em questionar sobre a sua possível personificação.²³⁸

Ao analisar se um agente artificial possuiria capacidade social de ação autônoma, deve-se levar em conta o tipo de ação social em questão; em particular, se ele é capaz de manifestar os comportamentos associados àquela ação (por exemplo, retornando ao caso da validade de contratos realizados por esses agentes artificiais, os comportamentos tradicionalmente associados à realização de um acordo). A demonstração desses tipos de comportamentos, por conseguinte, pressionaria o sistema jurídico a considerar o agente artificial como um sujeito de direito, e não mais mero objeto facilitador de transações.

3. Agentes autônomos de IA como pessoas jurídicas

²³⁷ ALLEN, Tom; WIDDISON, Robin. Op. cit. p. 39. No original: “(...) with organizations, the search is for the separation of the organization's identity and actions from those of its membership. In other words, it is the social capacity for collective action that is material. With computers, the search is also for the separation of the computer's identity and actions from its human controllers. Obviously, we would not describe this as the social capacity for collective action. However, we would describe it as the social capacity for autonomous action. Hence, we can translate Teubner's point by stating that "it makes legal policy sense to grant legal capacity to information systems that already have social capacity for autonomous action.”

²³⁸ Idem.

Não há um princípio único que determine quando um ordenamento jurídico deve reconhecer certa entidade como pessoa, e nem quando deve negar a personalidade jurídica. Analisando o tema através dessa perspectiva, inexistem limitações ao tratamento de agentes autônomos de inteligência artificial como pessoas jurídicas.

Quando um Estado concede personalidade a um ente, a nova pessoa reconhecida passa a ser considerada como um “nacional” daquele Estado. De maneira geral, a liberdade de definir os critérios para a aquisição e a perda de nacionalidade é um dos poderes básicos de todos os Estados soberanos. Portanto, é também uma questão de soberania nacional decidir quanto às entidades as quais concede personalidade. Ainda que leis de nacionalidade não sejam invocadas, um Estado é livre para fazer qualquer coisa que não seja vedada pela lei internacional pública. E, atualmente, tem-se que não há regras internacionais que proíbam a personalidade legal de uma inteligência artificial.²³⁹

A extensão da liberdade que os países possuem para reconhecer pessoas jurídicas é demonstrada pela amplitude das entidades que recebem esse status em todo o mundo, conforme previamente mencionado (de corporações a templos e ídolos religiosos). Conforme explica Thomas Burri, “assim como o legislador nacional poderia determinar que grandes símios ou certos rios são pessoas dentro da ordem jurídica doméstica, também poderia fazê-lo, por exemplo, para páginas da web” (tradução nossa)²⁴⁰.

Turner destaca que, tão logo um país conceda personalidade a uma inteligência artificial, é possível que a decisão tenha um efeito dominó em outras nações. Isso porque grande parte dos ordenamentos adota uma doutrina de mútuo reconhecimento em suas disposições sobre conflitos entre leis, as quais concederão o mesmo status para pessoas jurídicas reconhecidas em outros países, mesmo quando a entidade não é considerada como pessoa sob a lei local²⁴¹.

Foi precisamente o que ocorreu no caso anteriormente narrado *Bumper Development Corp. v. Commissioner of Police for the Metropolis*, em que o Tribunal de Apelação reconheceu que um templo indiano poderia ser tratado como uma pessoa jurídica pelo ordenamento jurídico britânico, uma vez que possuía esse status na Índia.

²³⁹ TURNER, Jacob. Op. cit., p. 179.

²⁴⁰ BURRI, Thomas. Free Movement of Algorithms: Artificially Intelligent Persons Conquer the European Union’s Internal Market. In: PAGALLO, Ugo; BARFIEL, Woodrow (orgs.) *Research Handbook on the Law of Artificial Intelligence*. Cheltenham: Edward Elgar, 2018. No original: “(...) just as national legislature could determine that great apes or certain rivers are persons within the domestic legal order, it could also state so, for instance, for webpages”.

²⁴¹ TURNER, Jacob. Op. cit., p. 180.

Por exemplo, nesse mesmo sentido, as disposições da União Europeia em matéria de liberdade de estabelecimento exigem que todos os seus Estados-membros reconheçam as pessoas jurídicas constituídas em conformidade legal e reconhecidas por ao menos um outro Estado-membro²⁴². Para Turner, caso uma inteligência artificial seja personificada em algum dos Estados da União, o reconhecimento por parte de apenas um dos demais membros pode levar todo o bloco a igualmente fazê-lo. Se a União Europeia conceder tal reconhecimento, é plausível que outras grandes economias mundiais sigam o exemplo, de modo a parecerem o mais atraentes possível para os designers e empreendedores de IA que desejem se aproveitar do novo status legal.²⁴³

Viu-se que as razões que podem levar determinado ordenamento jurídico a conceder o status de pessoa a um ente podem se basear em argumentos de moralidade, conveniência, ou refletir a realidade social.

Na primeira hipótese, o ordenamento jurídico reconhece que tal ente possui interesses distintos daqueles de outras pessoas, que devem ser protegidos por si só, e não apenas na medida em que coincidam com os direitos e interesses de terceiros. É nesse sentido que, atualmente, reconhece-se a personalidade de todos os seres humanos, ou que se pleiteia que o mesmo seja feito ao menos a algumas espécies animais que possuem caracteres considerados moralmente relevantes (como a consciência ou faculdades mentais superiores, como possuir um senso de si mesmo).

No atual estágio da tecnologia, preocupa-se com a proteção daqueles que interagem com agentes artificiais autônomos, e não com o próprio agente individualmente considerado. Estes não apresentam, até o momento, quaisquer desejos ou interesses merecedores de proteção por parte do ordenamento legal.

Em outros casos, verificou-se que a personalidade emerge como o reconhecimento do ordenamento jurídico de determinada realidade social, na medida em que muitos entes coletivos já possuem os atributos determinantes para serem considerados como pessoas em algum sentido extralegal. Assim, o seu reconhecimento enquanto pessoas jurídicas apenas completaria o processo de personificação que teve início em um contexto social.

²⁴² Artigo 54 do Tratado de Funcionamento da União Europeia (Treaty on the Functioning of the European Union ou TFEU): "Companies or firms formed in accordance with the law of a Member State and having their registered office, central administration or principal place of business within the Union shall, for the purposes of this Chapter, be treated in the same way as natural persons who are nationals of Member States.

"Companies or firms" means companies or firms constituted under civil or commercial law, including cooperative societies, and other legal persons governed by public or private law, save for those which are non-profit-making."

²⁴³ TURNER, Jacob. Op. cit., p. 181.

Finalmente, existem ainda casos em que a atribuição de personalidade a uma entidade é pautada pura e simplesmente por conveniência. Aqui, Allen e Widdison relembram que a lei inglesa do Almirantado, por exemplo, trata um navio como uma pessoa jurídica,²⁴⁴ algo que permite que aqueles que têm interesse nos negócios do navio o sujeitem a uma forma de prisão. Tradicionalmente, navios não são considerados como tendo um direito moral à personalidade; e nem como tendo personalidades extralegais. Não obstante, conferir uma forma de subjetividade a eles cumpre um certo objetivo do ordenamento jurídico, de maneira conveniente²⁴⁵.

Conclui-se, desse modo, que ao menos do ponto de vista normativo, não existem óbices à atribuição de personalidade a entes não humanos, notadamente agentes autônomos de inteligência artificial. Entretanto, é preciso questionar, à luz de outros fatores, se tal personificação seria, apesar de possível, também desejável.

Bryson et al. relembram que um ordenamento jurídico serve, em última instância, para proteger os interesses de pessoas, isto é, de pessoas atualmente reconhecidas como tal. Na ausência de alguma necessidade moral convincente, os prováveis custos e benefícios de qualquer mudança no sistema devem ser considerados – e a inclusão da IA na categoria de sujeitos de direito seria, inegavelmente, uma mudança²⁴⁶.

Assim, fatores como a alocação eficiente de riscos, a necessidade de se introduzir a personalidade para explicar todos os fenômenos relevantes, e a existência de explicações alternativas melhores, deverão ser levadas em consideração nas deliberações sobre a atribuição de personalidade a agentes artificiais. E ainda, essa análise deverá avaliar os custos de transação e os benefícios econômicos da introdução de agentes artificiais como atores legais plenos em uma esfera que não está acostumada a um reconhecimento explícito de seu papel²⁴⁷.

Por fim, deverão ser analisados os desafios remanescentes caso se opte pela personificação de agentes de IA autônomos como solução para o problema da lacuna de responsabilidade.

3.1. Argumentos contrários à personificação

²⁴⁴ CHOPRA, Samir; WHITE, Laurence. Artificial Agents-personhood in Law and Philosophy. In: *Proceedings of the 16th European Conference on Artificial Intelligence*. IOS Press, 2004, p. 635-639.

²⁴⁵ ALLEN, Tom; WIDDISON, Robin. Op. cit. p. 42

²⁴⁶ BRYSON et al. Op. cit., p. 274.

²⁴⁷ CHOPRA, Samir; WHITE, Laurence. A Legal Theory... Op. cit., p. 190.

No caso *Darhmouth College v. Woodward*²⁴⁸, em que a Suprema Corte dos Estados Unidos analisou a possibilidade de um estado da federação alterar a natureza privada de uma instituição de ensino, o juiz Thurgood Marshall bem descreve a natureza da pessoa jurídica como:

(...) um ser artificial, invisível e intangível, que existe apenas na contemplação do Direito. Sendo mera criatura da Lei, possui apenas as propriedades que são conferidas expressa ou incidentalmente à sua existência. Tais propriedades são aquelas melhor calculadas para atingir o objetivo para o qual [a pessoa jurídica] foi criada. Entre as mais importantes estão a imortalidade e (...) a individualidade; propriedades pelas quais uma sucessão perpétua de muitos indivíduos passa a ser considerada ao mesmo tempo e a atuar como uma única pessoa. Eles permitem que uma corporação gerencie seus próprios negócios e detenha propriedades sem as complexidades desconcertantes, e a infundável necessidade de estratégias para transmiti-los de mãos em mãos. É principalmente para o propósito de vestir corpos de homens, em sucessão, com essas qualidades e capacidades, que as corporações foram inventadas e estão em uso (tradução nossa).²⁴⁹

Pelo relato de Marshall, fica evidente que a estabilização do processo de coordenação de ações funcionalmente integradas, representado pela noção de organização, constitui uma das principais características que justificam a atribuição de personalidade a entes não humanos. Isso porque a existência de uma estrutura para a coordenação das ações é uma característica inerente ao próprio fenômeno associativo.²⁵⁰

Além disso, Visa A. J. Kurki ressalta que as coletividades humanas constituem um caso especial pois os seus interesses, embora explicáveis em termos individuais, não são redutíveis aos interesses dos membros do grupo. Empregando a ontologia social de Raimo Tuomela, isto é, sua teoria sobre a intencionalidade de grupos, o autor explica: se Mary adquire algum serviço de um grupo de amigos, o pagamento acordado não é devido a nenhum dos membros do grupo individualmente considerado, mas ao grupo em si. Isso porque o recipiente final da soma devida frequentemente não é definido – o grupo pode, por exemplo, decidir usá-la para investir no próprio projeto, dividi-la entre os seus membros, ou doá-la à

²⁴⁸ ESTADOS UNIDOS. Supreme Court. *Trustees of Dartmouth College v. Woodward*, 1819. Disponível em: www.oyez.org/cases/1789-1850/17us518. Acesso em 01 abr. 2020.

²⁴⁹ MARSHALL apud TURNER, Jacob. Op. cit., p. 175. No original: “A corporation is an artificial being, invisible, intangible, and existing only in contemplation of law. Being the mere creature of law, it possesses only those properties which the charter of its creation confers upon it, either expressly or as incidental to its very existence. These are such as are supposed best calculated to effect the object for which it was created. Among the most important are immortality, and, if the expression may be allowed, individuality; properties by which a perpetual succession of many persons are considered at the same, and may act as a single individual. They enable a corporation to manage its own affairs and to hold property without the perplexing intricacies, the hazardous and endless necessity of perpetual conveyances for the purpose of transmitting it from hand to hand. It is chiefly for the purpose of clothing bodies of men, in succession, with these qualities and capacities that corporations were invented and are in use.”

²⁵⁰ NEGRI, Sérgio. Op. cit., p. 10.

caridade. Contudo, deve-se notar que o projeto em si é “de propriedade” dos membros do grupo, que com ele se comprometeram pois o consideram significativo, e provavelmente têm uma opinião sobre como os produtos do projeto grupal devem ser usados.²⁵¹

A intencionalidade coletiva é um tipo particular de intencionalidade em que o “sujeito gramatical” de um estado mental é plural: em vez de um indivíduo e outro acreditam em X, ambos acreditam em X (nós). O exemplo fornecido por Kurki para ilustrar a separação entre intenções coletivas e individuais é o de viajar: pode-se viajar para uma conferência sabendo que outra pessoa da mesma faculdade também estará indo, o que por si só não constitui uma intencionalidade ou intenção coletiva. No entanto, se as duas pessoas decidirem participar da conferência juntas, elas estarão coletivamente comprometidas.²⁵²

Tuomela define os dois modos de intencionalidade como o modo Eu (“eu participarei da conferência sabendo que a outra pessoa P também o fará”) e o modo Nós (“nós – Eu e P – participaremos da conferência”), sendo a segunda modalidade irreduzível à primeira. As ações intencionais do modo Nós envolvem, naturalmente, as intenções do modo Eu necessárias para a participação de uma das partes na intenção conjunta; por exemplo, “como nós participaremos da conferência, eu reservarei as passagens de trem”. A principal razão para executar esta ação é, no entanto, a intenção conjunta.²⁵³

Também deve ser observado que, em grupos maiores, nem todos os membros precisam pensar no modo Nós, mas ao menos uma proporção significativa deles precisa para que o grupo seja funcional. É por essa razão que o funcionamento de uma corporação não pode ser explicado apenas fazendo referência às crenças, desejos e intenções independentes dos seus membros: a intencionalidade do modo Nós simplesmente não existe, exceto no contexto do grupo.²⁵⁴

Embora o processo de criação e distribuição de uma inteligência artificial envolvam uma multiplicidade de agentes, como os programadores, designers e operadores, uma vez

²⁵¹ KURKI, Visa A. J. *A Theory of Legal Personhood*. Oxford: Oxford University Press, 2019. No caso de uma pessoa jurídica formada por apenas um indivíduo (“one-person corporation”), como uma sociedade unipessoal ou microempreendedor individual, o autor emprega o termo “group agent” e não a considera propriamente como uma pessoa jurídica. Na verdade, esse tipo de arranjo forneceria apenas uma nova plataforma legal para o seu proprietário, que se diferencia da sua própria plataforma. Assim, uma pessoa que funda uma empresa individual estaria no controle de duas plataformas legais, com as seguintes características: a primeira delas relativa ao nome (“Mary” e “Mary Inc.”); a segunda, relativa ao fato de que as posições legais em cada plataforma são integradas (Mary pode acabar perdendo sua casa por causa de um contrato que ela assinou); e a terceira, relativa ao fato de as plataformas serem também separadas (Mary normalmente não pode ser pessoalmente responsável pelas dívidas da Mary Inc.).

²⁵² Id., p. 159.

²⁵³ TUOMELA, Raimo. *Social Ontology: Collective Intentionality and Group Agents*. Oxford: Oxford University Press, 2013.

²⁵⁴ KURKI, Visa A. J. Op. cit., p. 159.

desenvolvida, espera-se que a IA atue por si só, autonomamente, sem a necessidade de supervisão ou inputs diretos dos membros dessa cadeia. A noção de intencionalidade, individual ou coletiva, não se faz presente no agente artificial, mas apenas relativamente aos envolvidos no seu processo de desenvolvimento.

Em se tratando de uma inteligência artificial, não se encontra presente a sucessão constante de indivíduos que compõem a pessoa jurídica, o intuito coletivo direcionado a uma determinada finalidade, e a necessidade de organização de múltiplas atividades em um único sujeito, centro unitário de imputação. Em outras palavras, não abrangem uma pluralidade de agentes cujas complexas relações entre si demandam a sua simplificação em um único ente personificado.

No caso de atribuição de personalidade a uma IA, o centro unitário de imputação seria o próprio agente artificial, o que também traz à tona problemas relativos à sua identificação. Isso porque a personificação de uma IA pressupõe a possibilidade de identifica-la com razoável certeza. E, uma vez que esse tipo de tecnologia pode sofrer alterações e se adaptar, é possível questionar até que ponto estaria a se tratar do mesmo programa.

Não obstante futuros desenvolvimentos, atualmente não se pode confundir, necessariamente, um sistema de IA com o hardware físico no qual ele funciona (o que pode ser excepcionado em caso de robôs, cujo componente corpóreo se faz tão relevante quanto). Portanto, deve-se considerar algo inerente e identificável na natureza da IA não física como seu método de identificação, e a questão torna-se particularmente aguda quando a entidade em questão é uma unidade de um todo maior, como parte de um *swarm* (“enxame”)²⁵⁵ ou rede de sistemas de IA²⁵⁶.

Nesse mesmo sentido, Benjamin Allgrove destaca que será preciso, inicialmente, limitar, ao menos em algum grau, o quanto do código subjacente de uma IA pode ser alterado antes que ela deixe de ser a mesma inteligência artificial registrada. Isso potencialmente limitaria a evolução do software e exigiria monitoramento constante para garantir que a realidade corresponda àquela informada.²⁵⁷

É certo que o problema da identificação não é exclusivo da IA, surgindo também com sociedades e associações, cujos membros e controle podem variar com frequência. No

²⁵⁵ A inteligência de enxame (“*swarm intelligence*”) consiste em um ramo da inteligência artificial que tenta fazer com que computadores e robôs imitem o comportamento altamente eficiente de insetos de colônia, como formigas e abelhas. BRUNO, Lee. How Swarm Intelligence Helps AI Make Better Decisions. *All Turtles*, 2018. Disponível em: <https://www.all-turtles.com/2018/11/08/how-swarm-intelligence-helps-ai-make-better-decisions/>. Acesso em: 31 mar. 2020.

²⁵⁶ TURNER, Jacob. Op. cit., p. 198.

²⁵⁷ ALLGROVE, Benjamin. Op. cit., p. 81.

entanto, o registro torna essas entidades identificáveis e, para que uma IA seja tratada como uma pessoa, seria necessário desenvolver um sistema de registro semelhante.²⁵⁸

Para que esses entes artificiais possam gozar dos benefícios que acompanham a personalidade jurídica, portanto, pode ser exigido que eles sejam registrados e marcados com um “selo” de identificação eletrônica indelével e imutável correspondente a esse registro, de modo que a identificação seja sempre possível²⁵⁹ (tarefa não necessariamente fácil, repita-se, em virtude das constantes alterações promovidas no código e adquiridas com a interação ambiental de sistemas de auto aprendizado).

Na resolução de 16 de fevereiro de 2017 do Parlamento Europeu, com recomendações à Comissão de Direito Civil sobre Robótica (2015/2103-INL), o tema foi tratado dentro dos princípios gerais relativos ao desenvolvimento de robótica e inteligência artificial para uso civil:

[O Parlamento Europeu] considera que deveria ser introduzido um sistema abrangente de registro de robôs avançados no mercado interno da União, sempre que tal seja pertinente e necessário para certas categorias específicas de robôs, e insta a Comissão a definir critérios para a classificação dos robôs que deveriam ser registrados; neste contexto, insta a Comissão a ponderar se seria desejável que o sistema de registro e o registro fossem geridos por uma Agência da UE para a Robótica e a Inteligência Artificial.²⁶⁰

E, relacionando a questão à possível criação de um fundo compensatório para o caso de danos ocasionados por uma IA, complementa que deve ser garantido que

a ligação entre um robô e seu fundo seja visível por um número de registro individual que apareça em um registro específico da União, o que permitirá que qualquer pessoa que interaja com o robô seja informada sobre a natureza do fundo, os limites de responsabilidade em caso de danos à propriedade, nomes e funções dos colaboradores e todos os outros detalhes relevantes.²⁶¹

Uma opção apontada por Turner é que sistemas de inteligência artificial sejam registrados em mais de um local. Caso ele receba suas atualizações de uma fonte centralizada, mas também seja personalizado pelo usuário, sua regulação deveria ser individual e coletiva. O autor destaca que princípios similares de deveres sobrepostos se aplicam aos seres humanos: se o motorista de uma empresa colidir o veículo e causar danos, ele poderá ser responsável enquanto pessoa individual, mas também em sua capacidade de funcionário –

²⁵⁸ ALLEN, Tom; WIDDISON, Robin. Op. cit., p. 42.

²⁵⁹ TURNER, Jacob. Op. cit., p. 199.

²⁶⁰ UNIÃO EUROPEIA. Resolution 2015/2103-INL... Op. cit., art. 2.

²⁶¹ Id., art. 59(e).

circunstância na qual é provável que a vítima obtenha ressarcimento daquele que possui melhores condições de garanti-lo (geralmente o empregador).²⁶²

Embora nem toda IA precise ser registrada, isso pode se tornar um pré-requisito essencial para que ela se valha de certas prerrogativas econômicas e jurídicas, como a adoção de seguros, acesso ao sistema bancário ou até mesmo à Internet.²⁶³ Assim, para os agentes autônomos de inteligência artificial que desempenhem atividades em que seja provável a ocorrência de danos e consequente responsabilidade legal, seu registro e licenciamento podem ser obrigatórios.

A principal dificuldade em se introduzir um sistema de registro são os custos associados ao mesmo, que seriam significativos e precisariam ser ponderados com os benefícios de fazê-lo. Nesse sentido, Ian Kerr, tratando da hipótese de personificação como resposta aos problemas envolvendo a validade de transações realizadas por sistemas artificiais, questiona por que alguém iria se dar ao trabalho de criar e implementar tal sistema de registro quando existem mecanismos muito menos dispendiosos e menos exigentes para alcançar o mesmo resultado.²⁶⁴

Como ainda não se sabe com que frequência as ações de agentes autônomos de IA produzirão consequências não desejadas, não está claro se a complicação e despesa adicionais necessárias para desenvolver um sistema de registro são justificadas. Além disso, diferentemente de um registro corporativo, um sistema que atribui personalidade a dispositivos eletrônicos não resolveria necessariamente todos os problemas de identificação, uma vez que agentes artificiais polimórficos estão em constante evolução.²⁶⁵

O autor levanta ainda questionamentos acerca da criação de um sistema de registro internacional como, por exemplo, quem seria responsável por governá-lo, e em que local e de que maneira as disputas seriam resolvidas. A mesma questão é também levantada por Allgrove, que destaca que, embora regimes de personificação sejam nacionais, um sistema de registro de IA quase certamente teria que ser internacional, criando dilemas jurisdicionais e quanto ao compartilhamento de encargos.²⁶⁶

²⁶² TURNER, Jacob. Op. cit., p. 199.

²⁶³ Id., p. 200.

²⁶⁴ KERR, Ian. Providing for Autonomous Electronic Devices in the Electronic Commerce Act. In: ULCC Uniform Law Conference of Canada, 1999. Disponível em: <https://www.ulcc.ca/en/1999-winnipeg-mb/359-civil-section-documents/362-providing-for-autonomous-electronic-devices-in-the-electronic-commerce-act-1999>. Acesso em: 28 mar. 2020.

²⁶⁵ Id., p. 6.

²⁶⁶ ALLGROVE, Benjamin. Op. cit., p. 81.

Além da criação de um centro autônomo de imputação de responsabilidade, conferir personalidade jurídica a sistemas de inteligência artificial não produz outros benefícios que justifiquem as despesas com o registro. Por outro lado, conferir personalidade a empresas, viu-se, serve não apenas para identificar a entidade para pessoas de fora; mas também traz as vantagens legais e econômicas da responsabilidade limitada e a continuação de sua identidade no mundo jurídico. Nenhuma dessas vantagens resultaria da atribuição de personalidade jurídica à IA.²⁶⁷

Retornando à fala do Juiz Marshall em *Darhmouth College v. Woodward*, tem-se que a pessoa jurídica, enquanto “mera criatura da Lei, possui apenas as propriedades que são conferidas expressa ou indiretamente à sua existência”.²⁶⁸ Em outras palavras, cabe ao ordenamento jurídico modular a capacidade de direito dos entes aos quais atribui personalidade, definindo os direitos e deveres que eles poderão deter. É possível questionar, portanto, quais seriam os direitos e as correspondentes obrigações que um agente artificial autônomo, caso fosse personificado, poderia ter.

As questões envolvendo a modulação da capacidade de direito de uma pessoa jurídica (eletrônica) são multifacetadas, sendo, por isso, interessante analisar os três contextos distintos utilizados por Kurki para tratar da questão. Note-se, porém, que segundo a teoria proposta pelo autor, a personalidade não deveria ser tratada como um assunto em preto e branco (ou, conforme previamente colocado, como um conceito binário), mas sim em diferentes tons de cinza, sendo possível, portanto, a sua gradação (tratar-se-ia, nas palavras do autor, uma “*cluster property*”). Em linhas gerais, Kurki destaca que uma pessoa, natural ou jurídica, não seria simplesmente detentora de direitos ou portadora de deveres; pelo contrário, a personalidade consistiria em divisíveis, mas não obstante interconectados, incidentes da personalidade, passíveis de serem classificados em incidentes ativos e passivos.²⁶⁹

Partindo desse pressuposto, o autor avalia a personificação de IAs nos seguintes contextos: o primeiro deles é um contexto valorativo (“*ultimate-value context*”), no qual se questiona se uma inteligência artificial possuiria valor intrínseco sendo, assim, merecedora de algumas das proteções jurídicas que as pessoas naturais gozam. Kurki destaca que a ficção científica está repleta de exemplos de cenários em que algumas características de um robô (normalmente humanoide) suscitam questões sobre o seu tratamento moralmente correto. Garantias fundamentais são especialmente importantes aqui, como a proteção à vida, à

²⁶⁷ ALLEN, Tom; WIDDISON, Robin. Op. cit., p. 43.

²⁶⁸ MARSHALL apud Turner. Op. cit., p. 175.

²⁶⁹ KURKI, Visa A. J. Op. cit., p. 93.

liberdade e à integridade física, a capacidade para figurar como vítima no direito criminal e de sofrer danos de natureza não patrimonial.²⁷⁰

O segundo contexto existente é o da responsabilidade, cujo foco é, como o próprio nome revela, preocupações acerca da responsabilidade legal de uma inteligência artificial. Aqui, são discutidas questões como: carros ou robôs de segurança autônomos podem ser responsabilizados criminal ou civilmente por suas ações? Nesse contexto, o incidente mais relevante é o da “personalidade onerosa” (“*onerous legal personhood*”), consistente na responsabilidade civil, criminal e demais tipos de responsabilidade, como a administrativa. Principalmente no contexto de responsabilização civil, seria também necessária que a IA pudesse ser proprietária de bens.²⁷¹

Finalmente, o contexto comercial está relacionado com uma inteligência artificial atuando como um ator comercial, realizando compras, vendas, e assim por diante. Os três incidentes mais relevantes da personalidade seriam a posse de “direitos especiais” (“*special rights*”), o direito de propriedade e as competências legais (“*legal competences*”).²⁷² Os primeiro surgem de transações específicas entre indivíduos ou de alguma relação existente entre eles, e os direitos e deveres correspondentes limitam-se às partes envolvidas.²⁷³ Eles são diferentes, portanto, do direito de propriedade, passível de ser oponível *erga omnes*. Finalmente, as competências legais se traduzem na capacidade de administrar os outros incidentes da personalidade sem a necessidade de um representante legal.

Dentro do contexto da responsabilidade, a criação de um novo ator legal seria justificada pelo resultante centro autônomo de imputação, distinguindo-se, por conseguinte, a responsabilidade da IA da de seus programadores, usuários ou proprietários. Igualmente justificável e necessário seria o direito de propriedade, os direitos especiais e, a depender do grau de autonomia e sofisticação da IA, as competências legais.

Outros direitos, contudo, apresentam maior resistência de serem aplicados a entidades não humanas, levantando questionamentos sobre a possibilidade de ‘desvalorização’ da personalidade de pessoas naturais (sobretudo direitos enquadrados no contexto valorativo anteriormente apresentado). Nesse contexto, é interessante avaliar como determinados direitos, como o direito à liberdade religiosa, o direito à liberdade de expressão, ou o direito à

²⁷⁰ Id., p. 176.

²⁷¹ Idem.

²⁷² Idem.

²⁷³ Aqui, o autor faz referência a HART, H. L. A. Are There Any Natural Rights?, *The Philosophical Review*, v. 64, n. 2, 1955, p. 175-191.

compensação por danos morais, são estendidos a pessoas jurídicas, e as implicações sociais que essa extensão trazem consigo.

Inicialmente, cumpre ressaltar que negar ou conceder personalidade jurídica a entidades específicas revela um posicionamento acerca da avaliação social da entidade em questão. Para o jurista Stefano Rodotà, o Direito sempre contribuiu para a criação de antropologias, no sentido de que ele é responsável pela construção de figuras sociais. Cada grande operação jurídica delineou um modelo de pessoa, que nunca foi o mero registro de uma natureza “humana”, mas um jogo de plenos e vazios, de seleção daquilo que podia encontrar acolhimento no espaço do Direito e aquilo que deveria ficar de fora.²⁷⁴

Se compreendida como um jogo de soma zero, em que as decisões sobre a personalidade influenciam também outros interesses além dos da entidade em questão, a personificação de entidades não humanas pode representar prejuízos à personalidade de pessoas naturais.²⁷⁵

É nesse sentido que se afirma que a atribuição de personalidade a entes não humanos pode desvalorizar o significado social da personalidade de seres humanos, caso “a igualdade de direitos constitucionais somada a desigualdade de poderes reconhecidos e de fato levar, inexoravelmente, à supremacia de pessoas artificiais sobre pessoas reais” (tradução nossa).²⁷⁶

Bryson et al. explicam que, às vezes, estender a classe de entes personificados pode custar os interesses daqueles que já pertencem à categoria. No passado, a criação de novas pessoas jurídicas levou, em determinados momentos, a assimetrias e corrupções, através da existência de entidades que são responsáveis, mas não possuíam fundos para arcar com essa responsabilidade, ou entidades que possuíam fundos, mas não eram legalmente responsabilizadas. Segundo os autores, em última análise, isso significa enfraquecer as proteções legais de pessoas humanas em relação a pessoas artificiais.²⁷⁷

Isso se relaciona com o conceito proposto por Neil Richards e William Smart de “Falácia Androide” (“*Android Fallacy*”): “a ideia de que robôs são ‘iguais às pessoas’ e que há uma diferença significativa entre robôs humanoides e não humanoides”.²⁷⁸ Embora essa

²⁷⁴ RODOTÀ, Stefano. A antropologia do homo dignus. Trad. Maria Celina Bodin de Moraes. *Civilistica*, v. 6, n. 2, 2017, p. 1-17. Disponível em: <http://civilistica.com/aantropologia-do-homo-dignus/>. Acesso em: 29 mar. 2020.

²⁷⁵ CHOPRA, Samir; WHITE, Lawrence. A Legal Theory ... Op. cit., p. 185.

²⁷⁶ MAYER, Carl J. Personalizing the Impersonal: Corporations and the Bill of Rights. *Hastings Law Journal*, v. 41, n. 3, 1990, p. 577-667. No original: “ (...) equality of constitutional rights plus an inequality of legislated and de facto powers leads inexorably to the supremacy of artificial over real persons”.

²⁷⁷ BRYSON et al. Op. cit., p. 275.

²⁷⁸ RICHARDS, Neil M.; SMART, William D. How should the law think about robots? In: CALO, Ryan; FROMKIN, A. Michael; KERR, Ian (orgs.) *Robot Law*. Cheltenham: Edward Elgar Publishing, 2016.

definição tenha sido inicialmente cunhada para situações envolvendo robôs, isto é, sistemas de inteligência artificial acoplados a um substrato físico, ela é também passível de ser estendida a agentes autônomos de IA como um todo, pois reflete a natural propensão humana de antropomorfizar determinados entes cujo comportamento se assemelhe ao seu.²⁷⁹

Para os autores, essa projeção de atributos humanos é perigosa quando o objetivo é elaborar legislação envolvendo inteligência artificial. Isso porque tais agentes são, e por muitos anos permanecerão sendo, apenas ferramentas, ainda que com elevado grau de sofisticação:

À medida que a autonomia do sistema aumenta, fica cada vez mais difícil definir a conexão entre os comandos de entrada (inputs) e o comportamento do robô (outputs), mas ele existe e é determinístico. O mesmo conjunto de entradas gerará o mesmo conjunto de saídas sempre. O problema, no entanto, é que o robô nunca verá exatamente a mesma entrada duas vezes. Duas imagens da câmera nunca são iguais, devido a mudanças sutis de iluminação e erros de medição na própria câmera. Os humanos podem não ser capazes de ver as diferenças, mas o software do robô vê. O problema é que esse comportamento diferente em situações aparentemente semelhantes pode ser interpretado como ‘livre-arbítrio’ ou agência por parte do robô (tradução nossa).²⁸⁰

Similarmente, em abril de 2018, um grupo de especialistas em IA de 14 países europeus, incluindo cientistas da computação, juristas e CEOs de empresas de tecnologia, alertou que a concessão de personalidade jurídica aos robôs, conforme contemplado na Resolução do Parlamento Europeu de fevereiro de 2017, seria inapropriada sob uma perspectiva ética e legal. Eles afirmaram que, de um ponto de vista técnico, a proposta de criação do status de pessoa eletrônica traz consigo inúmeros vieses pautados em uma supervalorização das reais capacidades de até mesmo os robôs mais avançados, e uma compreensão superficial da imprevisibilidade e capacidade de autoaprendizagem²⁸¹.

É interessante notar os paralelos existentes entre alegações que caem na Falácia Androide, ou seja, alegações de que uma IA seria igual a uma pessoa humana, e discursos que

²⁷⁹ É precisamente por isso que pessoas reagem de maneira diferente a tecnologias que têm forma antropomórfica, que possuam olhos visíveis como um humano, ou falem com uma voz humana, em vez de texto legível. DARLING, Kate. *Extending Legal Rights to Social Robots*. We Robot Conference, Miami, 2012, p. 1-18. Disponível em: http://robots.law.miami.edu/wp-content/uploads/2012/04/Darling_Extending-Legal-Rights-to-Social-Robots-v2.pdf. Acesso em: 28 mar. 2020.

²⁸⁰ RICHARDS, Neil M.; SMART, William D. Op. cit., p. 18. No original: “As the autonomy of the system increases, it becomes harder and harder to form the connection between the inputs (your commands) and the outputs (the robot’s behavior), but it exists and is deterministic. The same set of inputs will generate the same set of outputs every time. The problem, however, is that the robot will never see exactly the same input twice. No two camera images are ever the same, because of subtle lighting changes and measurement errors in the camera itself. Humans might not be able to see the differences, but the robot’s software does. The problem is that this different behavior in apparently similar situations can be interpreted as “free will” or agency on the part of the robot.”

²⁸¹ Disponível em: <http://www.robotics-openletter.eu/>

buscam comparar pessoas jurídicas a pessoas naturais para fins de expansão dos direitos concedidos às primeiras. O fato de que muitas vezes tais discursos resultam no enfraquecimento da proteção concedida às pessoas humanas é capaz de revelar a possibilidade de que o mesmo ocorra no caso de personificação de uma inteligência artificial, sobretudo diante de tendências antropomórficas.

Carl J. Mayer, analisando o contexto de crescente concessão de direitos constitucionais (previstos na *Bill of Rights*) a pessoas jurídicas no sistema legal norte-americano, destaca que o pleito de empresas por certos direitos “intangíveis” emerge como um mecanismo de escape de regulações governamentais, muitas vezes às custas de liberdades e direitos de pessoas humanas.²⁸²

A Suprema Corte dos Estados Unidos concedeu pela primeira vez salvaguardas relativas à liberdade de expressão a pessoas jurídicas em 1978, no caso *First National Bank of Boston v. Bellotti*²⁸³. O posicionamento se manteve posteriormente no julgamento do caso *Pacific Gas and Electric Co. v. Public Utilities Commission* de 1986, no qual se estabeleceu, com raras exceções, o direito de uma empresa, enquanto editora, de se recusar a publicar mensagens com as quais não concorda²⁸⁴.

Ao discordar da conclusão majoritária da Corte de que entes coletivos desfrutavam de liberdade de expressão, o Juiz William Rehnquist opinou que atribuir a corporações um intelecto ou mente para fins de liberdade de expressão seria confundir uma metáfora com a realidade. Para ele, a insistência em tratar identicamente, para fins constitucionais, entidades comprovadamente diferentes consistiria em um pecado jurisprudencial tão grande quanto tratar diferentemente entidades idênticas.²⁸⁵

A importância de direitos relativos à liberdade de expressão se reflete no exercício muitas vezes agressivo dos mesmos pelas empresas, geralmente com consequências adversas para os indivíduos.²⁸⁶ Em 1987, os americanos testemunharam um acirrado debate público envolvendo esses direitos quando um projeto de lei foi introduzido no Congresso para proibir todas as formas de propaganda de cigarros, e as empresas de tabaco reagiram à altura.

Mayer exemplifica com um caso envolvendo a empresa Phillip Morris, que enviou um kit de imprensa a um grupo seletivo de editores de jornais e televisão com uma cópia do

²⁸² MAYER, Carl J. Op. cit., p. 605.

²⁸³ ESTADOS UNIDOS. Supreme Court. *First National Bank of Boston v. Bellotti*, 1978. Disponível em: <https://www.oyez.org/cases/1977/76-1172>. Acesso em 01 abr. 2020.

²⁸⁴ ESTADOS UNIDOS. Supreme Court. *Pacific Gas & Electric Company v. Public Utilities Commission of California*, 1986. Disponível em: <https://www.oyez.org/cases/1985/84-1044>. Acesso em 01 abr. 2020.

²⁸⁵ REHNQUIST apud MAYER, Carl J. Op. cit., p. 644.

²⁸⁶ MAYER, Carl J. Op. cit., p. 614.

Pravda, o jornal oficial do Partido Comunista da União Soviética, acompanhado dos dizeres: “Um jornal mundialmente famoso sem propaganda de cigarro”.²⁸⁷

O fato de uma empresa de tabaco se envolver em táticas tão agressivas para apoiar seu direito à liberdade de expressão é um indicativo da importância desses direitos na economia política moderna. Mas, como destaca Mayer, isso representa dois problemas para os indivíduos: primeiro, como o exemplo da Philip Morris ilustra, a afirmação desses direitos pode envolver certas medidas de intimidação direcionadas a indivíduos que preferem restringir a sua aplicação a empresas. Segundo, a afirmação da liberdade de expressão de corporações pode, por sua vez, privar o indivíduo de um certo tipo de liberdade - a liberdade de ser protegido dos males associados ao fumo e da influência exercida pela sua publicidade, no caso.²⁸⁸

Situações como essa se repetem no ordenamento jurídico norte-americano, em que apenas o privilégio contra a autoincriminação continua sendo a única salvaguarda da *Bill of Rights* ainda indisponível para as pessoas jurídicas.²⁸⁹

Em *Dow Chemical Co. v US*, a Suprema Corte considerou que uma empresa tinha uma expectativa razoável, legítima e objetiva de privacidade, em conformidade à Quarta Emenda Constitucional.²⁹⁰ E em *US v. Martin Linen Supply Co.*, a Corte decidiu que uma empresa tinha direito à proteção pela regra de dupla incriminação (*double jeopardy rule*) para protegê-la da “vergonha, dos custos e do desgaste, os quais a obrigaria a viver em um contínuo estado de ansiedade e insegurança”.²⁹¹

Mayer ressalta que é plenamente plausível que um grupo de pessoas possa se unir para atuar e possuir propriedades em conjunto. É bem menos plausível, contudo, que o mesmo grupo possa falar com uma só voz ou ter um interesse singular de privacidade. Segundo o autor, quanto mais intangível o direito, mais notável é a diferença entre uma pessoa jurídica e uma pessoa natural, e mais difícil fica tratar aquelas em analogia a estas.²⁹²

Enquanto muitos ordenamentos jurídicos permitem a objeção de consciência a indivíduos (como médicos que se recusam, por motivos religiosos, a realizar determinados procedimentos, especialmente de controle de natalidade), um tribunal norte-americano

²⁸⁷ Idem.

²⁸⁸ Idem.

²⁸⁹ Id., 624.

²⁹⁰ ESTADOS UNIDOS. Supreme Court. *Dow Chemical Company v. United States*, 1986. Disponível em: <https://www.oyez.org/cases/1985/84-1259>. Acesso em 01 abr. 2020.

²⁹¹ ESTADOS UNIDOS. Supreme Court. *United States v. Martin Linen Supply Company*, 1977. Disponível em: <https://www.oyez.org/cases/1976/76-120>. Acesso em 01 abr. 2020.

²⁹² MAYER, Carl J. Op. cit., p. 644.

determinou que a empresa farmacêutica Ralph's poderia ser vista como objetora de consciência ao armazenamento de métodos de controle de natalidade com base nas crenças religiosas e morais das pessoas naturais que a compõem (a família Storman).

O caso é emblemático uma vez que, para objetar conscientemente, uma pessoa deve necessariamente possuir crenças morais ou religiosas. Ainda que uma pessoa jurídica possa contratar, atuar em juízo ou ser proprietária, é altamente contra intuitivo atribuir a ela crenças, uma vez que elas não possuem um processo de raciocínio próprio distinto de seus integrantes. Não obstante, o tribunal empregou um modelo organizacional da pessoa jurídica para atribuir à empresa farmacêutica o direito em disputa:

A família Storman alega que a Ralph's não pode vender [a pílula do dia seguinte] Plano B "com base em motivos religiosos e morais" e que "as crenças religiosas de Kevin Storman impedem que ele venda uma droga que intencionalmente interrompe uma vida humana inocente". Os Stormans argumentam que a Ralph's é uma extensão das crenças dos membros da família Storman, e que as crenças da família Storman são as crenças da Ralph's (tradução nossa).²⁹³

As supostas crenças da empresa Ralph's derivam diretamente das crenças dos seus proprietários e funcionários humanos. As crenças religiosas e morais dos constituintes humanos são transferidas, desse modo, para a pessoa jurídica, e assim, a ficção da personalidade jurídica adquire a função de permitir que esforços humanos coletivos desfrutem dos mesmos benefícios e proteções da personalidade natural.²⁹⁴

Ainda mais preocupante é a decisão da Suprema Corte norte-americana de 30 de junho de 2014, no caso *Burwell v. Hobby Lobby*²⁹⁵, na qual se reconheceu, pela primeira vez na história do referido tribunal, a reivindicação de crença religiosa por parte de uma pessoa jurídica com fins lucrativos. No caso em tela, a empresa Hobby Lobby argumentou que a cobertura obrigatória de seguros determinada pela legislação do país, por incluir a oferta de "drogas abortivas", violava o seu direito à liberdade religiosa.

A Hobby Lobby Stores, Inc., uma cadeia nacional de artes e artesanato com mais de 500 lojas e mais de 13.000 funcionários pertence e é operada pela família Green, que

²⁹³ No original "Stormans alleges that Ralph's cannot sell Plan B "based on religious and moral grounds," and that Kevin "Stormans' [s] religious beliefs prevent him from selling a drug that intentionally terminates innocent human life." Stormans argues that Ralph's is an extension of the beliefs of members of the Stormans family, and that the beliefs of the Stormans family are the beliefs of Ralph's". ESTADOS UNIDOS. Court of Appeals for the Ninth Circuit. Stormans, Inc. v. Selecky, 2009. Disponível em: <https://cite.case.law/f3d/586/1109/>. Acesso em: 01 abr. 2020.

²⁹⁴ SOLAIMAN, S. M. Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy. *Artificial Intelligence and Law*, v. 25, n. 2, 2017, p. 155-179.

²⁹⁵ ESTADOS UNIDOS. Supreme Court. *Burwell v. Hobby Lobby*, 573, 2014. Disponível em: <https://www.oyez.org/cases/2013/13-354>. Acesso em: 01 abr. 2020.

organizou o negócio em torno dos princípios da fé cristã, tendo expressado explicitamente o desejo de administrar a empresa de acordo com os preceitos bíblicos, dentre os quais a crença de que o uso de contracepção é imoral.²⁹⁶

De acordo com o *Patient Protection and Affordable Care Act*²⁹⁷, popularmente conhecido como Obamacare, os planos de assistência médica ofertados pelo empregador devem fornecer aos empregados certos tipos de cuidados preventivos, como métodos contraceptivos aprovados pelo governo. Embora existam certas exceções a essa norma, dispensada a empregadores religiosos e instituições religiosas sem fins lucrativos, inexistiam isenções disponíveis para instituições com fins lucrativos, como a Hobby Lobby Stores, Inc.

Assim, em dezembro de 2012, a família Green, enquanto representantes da Hobby Lobby, ajuizaram uma ação em face do Departamento de Saúde e Serviços Humanos, contestando a obrigatoriedade de fornecimento de métodos contraceptivos. Tal exigência violaria a cláusula de liberdade religiosa da Primeira Emenda Constitucional, bem como o *Religious Freedom Restoration Act* (RFRA) de 1993.

A Suprema Corte considerou que, ao criar o RFRA, o Congresso pretendia que o mesmo fosse interpretado como sendo aplicável às corporações, uma vez que elas são compostas por indivíduos que as utilizam para alcançar determinados fins. Como o requisito de contracepção forçaria empresas religiosas a financiar aquilo que consideram como aborto, contrariando seus princípios ou enfrentando multas elevadas, o governo estaria criando um encargo significativo a essas empresas, o que não constitui o método menos restritivo de satisfazer interesses governamentais.

O método menos restritivo, afirmou a Corte, seria a extensão, por parte do Departamento de Saúde e Serviços Humanos, das mesmas isenções aplicadas a organizações religiosas não lucrativas também a corporações, pessoas jurídicas com fins lucrativos como a Hobby Lobby. Em seu voto dissidente, a Juíza Ruth Ginsburg ressaltou que, conforme precedentes, crenças religiosas não servem para afastar direitos de terceiros, como a isenção almejada faria com as mulheres que buscassem contracepção neste caso. Desse modo, verifica-se que, sob a escusa de exercício de uma suposta crença religiosa, permitiu-se que a pessoa jurídica em questão deixasse de ofertar a seus empregados métodos contraceptivos, como parte do seu plano de saúde.

Em sua análise da extensão de direitos constitucionais às pessoas jurídicas no ordenamento norte-americano, Bryson et al. corretamente concluem, portanto, que com

²⁹⁶ Idem.

²⁹⁷ ESTADOS UNIDOS. *Patient Protection and Affordable Care Act* (Public Law 111-148), 2010.

demasiada frequência trata-se de um processo que resulta na diminuição de direitos e poderes de indivíduos reais. Os direitos previstos na Quarta Emenda Constitucional, que proíbem buscas e apreensões desarrazoadas, diminuem o direito do indivíduo de viver em um mundo não poluído ou de desfrutar de privacidade. O exercício corporativo dos direitos da Primeira Emenda Constitucional, que dispõe sobre a liberdade de expressão, de imprensa, religiosa e associativa frustra o direito do indivíduo de participar igualmente nas eleições democráticas, de pagar taxas razoáveis de utilidade e de viver em um ambiente livre de toxinas.²⁹⁸

Para Negri, constata-se no discurso dos direitos da personalidade da pessoa jurídica uma inversão, pois, ao invés de se analisar a pertinência do conteúdo do direito em questão, por meio de uma valoração contextualizada, que leve em conta as particularidades de cada tipo de pessoa jurídica, procura-se conferir um tratamento generalizante ao problema.²⁹⁹

No ordenamento jurídico brasileiro, a despeito do silêncio constitucional sobre o tema, doutrina, jurisprudência e legislação infraconstitucional convergem para a assertiva de que pessoas jurídicas são titulares de direitos fundamentais.³⁰⁰ O artigo 52 do Código Civil brasileiro estende à pessoa jurídica, no que for possível, a proteção dos direitos da personalidade. E grande parte da doutrina, ao interpretar o dispositivo, chega à conclusão de que o Código Civil teria esvaziado a discussão sobre a possibilidade de a pessoa jurídica titularizar direitos da personalidade. Com raras exceções, a única ressalva comumente apontada diz respeito à impossibilidade de a pessoa jurídica reclamar a proteção de direitos que se mostrem incompatíveis com a sua natureza.

O autor destaca que o debate em tela concentra-se na simples análise da subjetividade jurídica. Ou seja, a atribuição de personalidade jurídica, ao promover o desenvolvimento de um novo sujeito, já justificaria por si só o reconhecimento dos direitos fundamentais, culminando no seguinte silogismo: se a pessoa jurídica pode ser considerada um sujeito para o Direito, assim como ocorre com a pessoa natural, então aquela deveria receber a mesma tutela desta.³⁰¹

Nesse contexto, insere-se a súmula 227 do Superior Tribunal de Justiça, que reconhece a possibilidade de a pessoa jurídica sofrer dano moral³⁰², baseando-se no entendimento de que

²⁹⁸ Os exemplos dados por Bryson et al. remetem a casos específicos julgados pela Suprema Corte dos Estados Unidos. Por todos, veja-se: BRYSON et al. Op. cit., p. 658.

²⁹⁹ NEGRI, Sérgio. Op. cit., p. 10.

³⁰⁰ Id., p. 11.

³⁰¹ Id., p. 12.

³⁰² BRASIL. Superior Tribunal de Justiça, Súmula 227: “A pessoa jurídica pode sofrer dano moral”.

a honra, inicialmente vista como um atributo do ser humano, envolveria um aspecto subjetivo e outro objetivo. Este, relacionado à reputação, estender-se-ia também à pessoa jurídica.

Conforme bem destacado por Negri, verifica-se em alguns julgados que a afirmação da titularidade do direito à honra por entes jurídicos coletivos ocorre em colisão com outros direitos fundamentais, como o direito à liberdade de expressão:

Nessas situações, invoca-se, por vezes, a honra profissional ou a imagem da pessoa jurídica para justificar o dano moral sofrido por sociedades empresárias com a divulgação de matérias jornalísticas. Sem se atentar para a diferença entre as razões que determinam a proteção do ser humano e as que determinam a proteção da pessoa jurídica, as normas aplicáveis aos direitos da personalidade são simplesmente transpostas para a proteção da atividade empresarial e para a tutela de interesses de natureza exclusivamente econômica.³⁰³

Nesses julgados, a inadequada elucidação dos interesses presentes no caso concreto representa um obstáculo para a justificação da decisão final que atribui determinado peso a um direito em detrimento de outro. Para o autor, a divulgação de uma matéria denunciando péssimas condições de trabalho ou a violação de direitos humanos por determinada organização empresarial, embora possa gerar prejuízos aos seus controladores no curto prazo, pode também beneficiar os trabalhadores, acionistas minoritários e outras partes interessadas ou afetadas pela atividade desenvolvida.³⁰⁴

A tutela do nome, da honra ou da identidade de pessoas jurídicas não revela aspectos essenciais da personalidade do ente, sendo invocados para garantir, na prática, o exercício da atividade que lhe é correlata. Na tutela do nome da pessoa jurídica não se busca garantir a correta representação da pessoa nas relações sociais, mas apenas a proteção do livre desenvolvimento da função do ente em questão. Existe, portanto, uma diferença fundamental, no que tange aos interesses tutelados, entre a função desempenhada por esses direitos em relação à pessoa jurídica daquela desempenhada relativamente à pessoa natural.³⁰⁵

Perder de vista as razões específicas que informam as pessoas jurídicas, equiparando-as indistintamente às pessoas naturais, pode conduzir ao processo denominado por Stefano Rodotà de “expropriação da subjetividade”³⁰⁶. Tal processo consistiria na usurpação, no plano concreto, de direitos inerentes ao ser humano, sob o pretexto de proteção do sujeito abstrato³⁰⁷.

³⁰³ NEGRI, Sérgio. Op. cit., p. 13.

³⁰⁴ Idem.

³⁰⁵ PERLINGIERI, Carolina. *Enti e diritti della persona*. Napoli: Edizioni Scientifiche Italiane, 2009.

³⁰⁶ RODOTÀ, Stefano. *La vita e le regole. Tra diritto e non diritto*. Milano: Fetrinelli Editore, 2007.

³⁰⁷ NEGRI, Sérgio. Op. cit., p. 2.

Retomando a ideia de que o Direito constrói suas figuras sociais (antropologia), o conceito de expropriação da subjetividade emerge em um contexto de prevalência da ideia abstrata de sujeito de direito, indiferente à realidade das condições materiais dos indivíduos.³⁰⁸

O autor reconhece a importância desempenhada pela criação do sujeito de direito, a qual descreve como um dos grandes êxitos da modernidade. A abstração do sujeito foi indispensável para que a sociedade saísse dos status, e se abrisse o caminho para o reconhecimento da igualdade.³⁰⁹ Ou seja, para que se abandonassem as restrições daqueles que tinham relevância dentro do ordenamento jurídico, passando a reconhecer como pessoas todos os seres humanos.

Ocorre que, embora o sujeito abstrato inicialmente compreendesse a todos, seu prolongar no tempo ameaçava não definir ninguém por completo:

O que deve ser rejeitado é o uso político que pouco a pouco esterilizou a força histórica e teórica dessa invenção, reduzindo o sujeito a um esqueleto que isolou o indivíduo, separou-o de qualquer contexto, abstraindo das condições materiais. Por isso era essencial empreender um caminho diferente. Daí a necessidade de retomar o fio partido da igualdade, subtraindo-a não dos benefícios de uma forma que continua a ser instrumento contra a institucionalização da discriminação, mas da indiferença à realidade do ser, criando assim novas hierarquias e novos abandonos com base na força política e na arrogância do mercado.³¹⁰

O reducionismo da pessoa (natural) ao sujeito acaba por determinar não apenas uma exclusão subjetiva daqueles que não apresentam todas as características consideradas necessárias para realizar as operações econômicas, mas também impõe uma exclusão objetiva de tudo o que pertence a relações irreduzíveis a dados econômicos puros. Assim, destaca o autor, ocorre o processo de expropriação da subjetividade, e uma negação da global autonomia da existência. Condena-se à marginalidade todas as decisões de conteúdo não patrimonial, construindo-se, em essência, um paradigma fechado de normalidade jurídica.³¹¹

O processo de expropriação da subjetividade engloba, portanto, todas as situações em que, sob o pretexto de proteção de um sujeito abstrato, acaba-se por limitar a proteção conferida ao sujeito concreto. Exemplares de tal fenômeno são as decisões judiciais previamente narradas, representativas do discurso da titularidade dos direitos fundamentais

³⁰⁸ RODOTÀ, Stefano. La vita ... Op. cit., p. 19.

³⁰⁹ RODOTÀ, Stefano. A antropologia ... Op. cit., p. 4.

³¹⁰ Id., p. 5.

³¹¹ RODOTÀ, Stefano. La vita... Op. cit., p. 27. Na versão original: "Si realizzava così una espropriazione della soggettività e una negazione della complessiva autonomia dell'esistere. Si condannavo alla marginalità tutte le decisioni a contenuto non patrimoniale. Si costruiva, in sostanza, un paradigma chiuso di normalità giuridica."

por sociedades e associações, na literatura do Direito Constitucional, e na análise dos direitos da personalidade da pessoa jurídica, no âmbito do Direito Privado.

Ao contrário do que se observa com a pessoa natural, o mecanismo de imputação de direitos e deveres à pessoa jurídica não se faz completo. Esse conceito de imputação incompleta tem por objetivo lembrar que no interior da pessoa jurídica a menção aos interesses, crenças e desejos, não pode ter como referência o ente abstrato, mas apenas o ser humano, que representa o verdadeiro responsável pelas ações e omissões imputadas, metaforicamente, ao ente abstrato.³¹²

Aproxima-se da noção de imputação incompleta descrita por Negri o apontamento feito por S. M. Solaiman de que, quando a lei determina que uma pessoa jurídica faça ou se abstenha de fazer certas coisas, tal comando se aplica, essencialmente, aos indivíduos por trás dela; se violados, a pessoa jurídica poderá ser penalizada como forma de punir direta ou indiretamente as pessoas naturais que a administram e/ou a possuem. O ente jurídico coletivo seria, desse modo, utilizado pelo ordenamento como uma mira, com o objetivo final de acertar os indivíduos por trás dele. Portanto, as pessoas naturais não estariam imunes às obrigações decorrentes de falhas no cumprimento de obrigações corporativas ou no cumprimento da lei.³¹³

Da mesma forma, quando da atribuição de direitos ao ente, os mesmos devem ser transpostos, em última instância, aos seres humanos que o compõem:

(...) na realidade, tanto indivíduos quanto empresas podem ser responsabilizados por irregularidades corporativas. Isso significa que, apesar da separação da personalidade determinada por lei, os indivíduos por trás de uma corporação não estão imunes à responsabilidade que pode resultar da violação de um dever corporativo cometido por indivíduos utilizando-se do véu da incorporação. Por outro lado, também seus gerentes, controladores, e acionistas podem ser responsabilizados por dívidas corporativas em extensões variáveis, dependendo do tipo da empresa. Quaisquer multas pecuniárias pagas pelas empresas acabam sendo impostas aos acionistas, que são os reclamantes residuais dos ativos corporativos. Da mesma forma, é provável que quaisquer lucros de uma corporação beneficiem seus membros ao longo do tempo, na forma de dividendos, ganhos de capital ou pagamentos residuais (tradução nossa).³¹⁴

³¹² NEGRI, Sérgio. Op. cit., p. 14.

³¹³ SOLAIMAN, S. M. Op. cit., p. 174.

³¹⁴ SOLAIMAN, S. M. Op. cit., p. 173. Na verso original: “(...) in reality, both individuals and corporations can be held liable for corporate wrongdoings. This means that despite the separation of personality by law, individuals behind a corporation are not immune from the liability that may arise from the breach of a corporate duty committed by individuals wearing the veil of incorporation. On the other hand, as well as managers/controllers, shareholders may be held liable for corporate debts to varying extents, depending on the type of their company. Any pecuniary fines paid by corporations are eventually imposed on stockholders, who are the residual claimants of corporate assets. Correspondingly, any profits of a corporation are likely to benefit its corporator in the course of time, in the form of dividends, capital gains or residual payments.”

Tal assertiva revela outra preocupação comumente apontada quando da análise da personificação de sistemas de inteligência artificial: a de que seres humanos façam uso dos mesmos para se esquivar da responsabilidade.³¹⁵ É possível pressupor que, caso de fato uma IA seja constituída como um sujeito autônomo de direito, atores humanos procurarão explorar tal fato para fins egoístas.

Embora a persecução desses fins nem sempre seja censurável por si só, um ordenamento jurídico equilibrado, fundado no princípio da dignidade da pessoa humana e comprometido com a solidariedade social, deve considerar o impacto das mudanças nas regras no sistema como um todo, principalmente no que diz respeito aos direitos das pessoas já reconhecidas por ele enquanto tal.

Nesse contexto, Bryson et al. revelam preocupações acerca do abuso da personalidade jurídica, ou seja, indivíduos que se utilizem do ente artificial para se protegerem das consequências de sua conduta. Para eles, o reconhecimento da personalidade jurídica eletrônica de uma IA pode apresentar a atores inescrupulosos oportunidades de gerenciamento de sua responsabilidade.³¹⁶

No que tange à atribuição subjetiva de responsabilidade, Kurki aponta que os desenvolvedores certamente podem ser responsabilizados caso, por exemplo, tenham criado intencionalmente uma IA para cometer um ilícito. Esses mesmos desenvolvedores, porém, poderiam de fato se beneficiar da responsabilidade imputada às suas criações na hipótese de que a elas fosse atribuída uma personalidade distinta.³¹⁷

Similarmente, ao analisar a responsabilidade civil dos administradores de sociedades empresárias por decisões tomadas com base em sistemas de inteligência artificial, Ana Frazão destaca que a autonomia desses sistemas não deveria afastar, por si só, a responsabilidade usual dos administradores e das respectivas sociedades que se utilizam dos mesmos.³¹⁸

Para a autora, qualquer que seja o regime adotado (no caso, de responsabilidade civil, mas é possível estender o raciocínio também à personalidade), deve-se evitar que este sirva como mero subterfúgio para a irresponsabilidade ou impunidade, “o que certamente ocorreria

³¹⁵ BRYSON et al. Op. cit., p. 285.

³¹⁶ Id., p. 286.

³¹⁷ KURKI, Visa A. J. Op. cit, p. 181.

³¹⁸ FRAZÃO, Ana. A Responsabilidade Civil dos Administradores de Sociedades Empresárias por Decisões Tomadas com Base em Sistemas de Inteligência Artificial In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 481-521.

se os agentes econômicos que se utilizassem de sistemas de inteligência artificial não mais respondessem pelos danos daí decorrentes”.³¹⁹

Na verdade, conclui a autora que, em um cenário como este, poderia haver até mesmo um incentivo para a automação total, uma vez que isso poderia ser utilizado como verdadeira excludente de responsabilidade dos atores humanos por trás da máquina.³²⁰

A atribuição de personalidade a agentes de inteligência artificial poderia, assim, levar a um cenário de “irresponsabilidade organizada”, expressão cunhada por Günther Teubner inicialmente para alertar para o risco de que redes contratuais possam funcionar como verdadeiros grupos empresariais sem a necessária e correspondente responsabilidade conjunta perante terceiros³²¹; mas que também pode ser utilizada para todas as outras formas de arranjos e colaborações empresariais.

Desse modo, alerta-se para o risco de que as formas jurídicas obscureçam as reais alocações de poder realizadas entre os agentes econômicos³²², com a consequência de que a personalidade jurídica acabe servindo como um escudo de responsabilidade (“*liability shield*”).³²³

Teubner justifica as ações de combate à irresponsabilidade organizada - como a aplicação da responsabilidade civil solidária e objetiva aos integrantes de uma rede contratual nos casos em que se fazem ausentes explicações causais claras para a responsabilidade individual de um acidente - como uma resposta apropriada ao uso indevido da forma jurídica.³²⁴

Em se tratando da personalidade jurídica, já se dispõe de mecanismo para contrabalancear a sua utilização indevida: a desconsideração. Nos casos em que uma pessoa jurídica é usada como um escudo ou ‘véu’ para ações consideradas pelo ordenamento como ilícitas ou indevidas, e seus integrantes procuram explorar a forma jurídica da personalidade para se blindarem, pode-se ignorar a distinção entre a personalidade da pessoa jurídica e a personalidade das pessoas naturais que a compõem, ‘perfurando-se o véu corporativo’. Em tais situações, a própria lei reconhece que a ficção da personalidade jurídica distinta possui utilidade apenas até determinado ponto.

³¹⁹ Id., p. 505.

³²⁰ TEUBNER, Gunther. *Networks as connected contracts*. Oxford: Hart Publishing, 2011.

³²¹ Id., p. 68.

³²² FRAZÃO, Ana. Op. cit., p. 483.

³²³ Expressão utilizada por TURNER, Jacob. Op. cit., p. 191.

³²⁴ Id., p. 70.

Assim, no caso de se conceder personalidade a agentes de inteligência artificial, inexistindo alguma maneira de se contornar o escudo da personalidade, certamente ocorreriam instâncias em que esses agentes seriam projetados e utilizados para realizar atividades que apresentam alto risco (jurídico) para aqueles que o desenvolvem e/ou utilizam (pessoas humanas e entes coletivos). Embora isso possa beneficiar, em um primeiro momento, os seres humanos por trás da IA, isso seria feito às custas de interesses humanos em geral.³²⁵

Ainda, a desconsideração da personalidade no caso de agentes autônomos de inteligência artificial, mesmo que eventualmente venha a ser acolhida legalmente, traz consigo alguns desafios remanescentes. Em primeiro lugar, retornar-se-ia ao problema inicial que a personificação buscava solucionar, ou seja, o de quem deveria responder pelo dano: o próprio usuário da IA, seu programador, o responsável pelo design da máquina, ou ainda quem a comercializou.

Além disso, deverão ser definidos os critérios necessários para promover a desconsideração da personalidade de um agente artificial. Por exemplo, ainda que houvesse um seguro ou patrimônio associado ao ente, mas o mesmo não fosse capaz de cobrir o dano causado em sua extensão, questiona-se se poderia haver a desconsideração e, novamente, quem arcaria com a complementação dos valores devidos.

Outros obstáculos também se fazem presentes, já na análise do instituto da desconsideração quando aplicada a empresas: o fato de que a mesma apenas pode ser decretada em juízo, com a demonstração, pela parte afetada, da existência dos requisitos legais necessários, e muitas vezes apenas após o esgotamento de outras tentativas de obtenção do ressarcimento.

Em suma, é possível concluir que a atribuição de personalidade a agentes autônomos de inteligência artificial, embora seja possível, pode não ser o mecanismo jurídico mais adequado para lidar com os desafios advindos da sua criação e ampla utilização no mercado, em especial a lacuna da responsabilidade. Viu-se que a criação de um sistema de identificação e registro desses agentes – pressuposto indispensável ao seu reconhecimento como pessoas – acarretaria custos e dilemas quanto à sua implementação e abrangência.

Ausentes todas as razões que justificam a criação de uma personalidade jurídica distinta daquela de seus integrantes senão a criação de um centro autônomo de imputação com autonomia patrimonial, é provável que os custos associados ao sistema de identificação e registro de IAs supere os benefícios da sua implementação.

³²⁵ BRYSON et al. Op. cit., p. 287.

Viu-se também que a personificação de agentes autônomos de inteligência artificial pode contribuir no fortalecimento da tendência de os antropomorfizar, equiparando-os indistintamente a pessoas humanas. Tratar-se-ia de um processo similar ao que já ocorre quando da atribuição de direitos fundamentais ou de personalidade a entes coletivos artificiais – e que muitas vezes resulta no enfraquecimento da proteção concedida aos seres humanos que com eles se relacionam (conceito de expropriação da subjetividade de Rodotà).

Mas não apenas: deve-se apontar os riscos envolvidos na utilização indevida do instituto jurídico da personalidade como subterfúgio para empresas e indivíduos que criam, desenvolvem e fazem uso de tais agentes artificiais. A personalidade jurídica distinta dos mesmos pode viabilizar a irresponsabilidade organizada das pessoas envolvidas nesse processo.

Há de se ter em vista, finalmente, a quem favorece a exteriorização dos custos representados pela criação de uma nova subjetividade com autonomia patrimonial, e em desfavor de quem os riscos representados pela atividade desenvolvida são limitados. Sabe-se que a autonomia patrimonial de entes coletivos justifica-se em virtude da necessidade de exteriorizar uma parcela dos custos derivados da atividade desenvolvida, diminuindo-se, assim, o chamado “risco empresa”.

Em outras palavras, atribui-se personalidade ao ente abstrato para que, mediante a afirmação de uma estrutura patrimonial autônoma, haja limitação dos custos da atividade em favor dos sócios e em desfavor da coletividade, possibilitando, assim, o desenvolvimento da própria atividade econômica.

Entretanto, o controle de riscos em favor dos sócios e em desfavor de terceiros não é admitido quando confrontarem os interesses de pessoas naturais em flagrante posição de vulnerabilidade na relação estabelecida com o ente personificado. É o que ocorre, por exemplo, quando do inadimplemento de verbas de natureza laboral: diante da insuficiência do patrimônio da pessoa jurídica, os seus sócios são chamados a responder pelo adimplemento com seu patrimônio particular.

Em se tratando de sistemas de inteligência artificial, não se pode olvidar que sua criação e disponibilização no mercado são feitas, tradicionalmente, por pessoas jurídicas. Na verdade, é interessante notar que apenas sete instituições privadas - Google, Facebook, IBM,

Amazon, Microsoft, Apple e Baidu - possuem capacidades no campo da IA que ultrapassam largamente a de todas as demais instituições da área.³²⁶

Significa dizer, portanto, que os custos associados à atividade exercida (pesquisa e desenvolvimento de tais tecnologias), em sua absoluta maioria, já foram devidamente redistribuídos à comunidade por meio da imputação direta ao ente jurídico abstrato.

A personificação de sistemas artificiais autônomos representaria, portanto, nova exteriorização de custos, agora em favor da pessoa jurídica e em prejuízo das pessoas naturais que, possuidoras de estados subjetivos internos, são aptas a possuir desejos e crenças, formular objetivos e também vê-los frustrados, constituindo o polo mais vulnerável de qualquer relação envolvendo uma IA. Dificultar-se-ia ainda mais, portanto, a reparação integral dos danos por elas suportados, privilegiando única e exclusivamente os interesses de seus desenvolvedores.

De fato, nos casos em que a cadeia causal entre uma pessoa reconhecida pelo ordenamento jurídico e certo resultado danoso é rompida, a interposição de uma nova pessoa (no caso, eletrônica) fornece ao ordenamento uma entidade passível de responsabilização. A personificação de agentes de inteligência artificial, portanto, permitiria que a responsabilidade fosse alcançada com danos mínimos aos conceitos de causalidade e agência, mantendo assim a coerência do sistema como um todo³²⁷.

Entretanto, serão apresentadas no capítulo subsequente alternativas jurídicas viáveis já existentes para lidar com o principal problema que a personificação de agentes autônomos de IA busca solucionar, qual seja, o da responsabilidade pelos danos que eles eventualmente venham a causar. Estas alternativas, somadas ao todo o exposto ao longo deste capítulo, leva à conclusão que, ao menos em um futuro próximo, a personificação desses agentes, ainda que possível, não é desejável.

³²⁶ Dados referentes ao ano de 2017. CALO, Ryan. *Artificial Intelligence Policy: A primer and roadmap*, 2017. Disponível em: <https://ssrn.com/abstract=3015350>. Acesso em: 30 mar. 2020.

³²⁷ TURNER, Jacob. Op. cit., p. 186.

CAPÍTULO III

A RESPONSABILIDADE CIVIL POR DANOS CAUSADOS POR UM AGENTE ARTIFICIAL AUTÔNOMO

Do ponto de vista formal, a responsabilidade civil extracontratual nasce do descumprimento de um dever genérico e universal de não causar danos a outras pessoas. Enquanto na responsabilidade civil contratual tem-se a inobservância de qualquer dever emanado de uma relação obrigacional, na responsabilidade extracontratual prescinde-se de uma relação preexistente entre lesante e lesado, sendo suficiente o descumprimento de um dever que emerge do tráfego social.³²⁸

A responsabilidade civil extracontratual (doravante apenas ‘responsabilidade civil’) tem por objetivo colocar a pessoa prejudicada na posição em que estaria se a violação do dever geral de não causar danos a outros não tivesse ocorrido. Assim, é possível que o dever de reparação surja também em favor de indivíduos prejudicados por suas interações com um agente artificial.

A potencialidade dos danos causados por uma inteligência artificial não é trivial. Como exemplos particularmente alarmantes tem-se os sistemas de controle de baterias de mísseis, pilotos automáticos, sistemas de controle de trens e softwares de controle de medicina nuclear, que podem causar mortes e ferimentos graves se operarem incorretamente. Na esfera econômica, quebras de sistemas de negociação e aconselhamento equivocados de sistemas especialistas podem causar perdas significativas para seus usuários.³²⁹

No que diz respeito à imputação em casos de danos causados por agentes autônomos de inteligência artificial, duas peculiaridades fazem com que a aplicação das teorias tradicionais sobre responsabilidade civil seja dificultada. A primeira delas consiste, conforme apresentado ao longo do primeiro capítulo do presente trabalho, na imprevisibilidade de suas ações.

A segunda peculiaridade, abordada no capítulo anterior, consiste na existência de agência causal sem uma correspondente agência legal³³⁰ - pelo menos até o presente

³²⁸ ROSENVALD, Nelson; FARIAS, Cristiano de Chaves; BRAGA NETTO, Felipe Peixoto. *Curso de Direito Civil: Responsabilidade civil*. 4^a ed. Salvador: Editora Juspodivm, 2017.

³²⁹ Por todos, ver CHOPRA, Samir; WHITE, Lawrence. *A Legal Theory ...* Op. cit., p. 119.

³³⁰ ASARO, Peter M. *The Liability Problem for Autonomous Artificial Agents*. AAAI Spring Symposium Series, 2016. Disponível em: <https://www.aaai.org/ocs/index.php/SSS/SSS16/paper/view/12699>. Acesso em: 09 abr. 2020.

momento. Significa dizer, em linhas gerais, que tais agentes são capazes de produzir ações danosas sem que, correspondentemente, sejam responsabilizados pelas mesmas.

Similarmente, dois conceitos distintos de ‘causalidade’ possuem relevância relativamente à responsabilidade de agentes artificiais. Para que haja a responsabilização de um sujeito no âmbito cível, o ato ou a omissão injusta do réu deverá ter causado um dano ao demandante, no sentido estrito de que, a não ser pelo ato ou omissão, tal prejuízo não teria ocorrido. Mas para que a responsabilidade seja constatada, deve haver não apenas essa causalidade de fato, mas também causalidade normativa: a ação ou omissão também deve ser considerada legalmente relevante para fins de alocação de responsabilidade.³³¹

Nesse sentido, é possível identificar determinadas ações que, embora sejam causa (factual) adequada a um resultado, não culminam na imputação normativa desse resultado ao agente que o causou, pela ausência de intencionalidade direta ou indireta quanto a ele (ausência de dolo e culpa, respectivamente). Se, por exemplo, um indivíduo, ao ser empurrado contra um objeto, causa a sua destruição, não é possível afirmar, normativamente, que foi ele quem causou o resultado, mas sim aquele que o empurrou. A análise dessa intencionalidade do agente, demonstrada através de dolo ou culpa, é o que caracteriza a responsabilidade civil subjetiva.

Da mesma forma, é possível identificar situações em que, ainda que ausente a intencionalidade direta ou indireta quanto ao resultado, imputa-se responsabilidade ao agente (causalidade normativa). A análise apenas do dano causado e do nexo de causalidade entre a ação ou omissão do agente e o resultado danoso, sem levar em consideração elementos volitivos do mesmo, é o traço distintivo da responsabilidade civil objetiva.

A análise da responsabilidade civil por danos causados por uma inteligência artificial perpassa, ainda, duas questões relevantes. A primeira delas diz respeito às hipóteses em que o dano causado pela IA seja resultado direto de um erro humano, ou seja, através de um comportamento negligente, por falha de um dever de cuidado, de um defeito de fabricação ou manutenção, do fornecimento inadequado ou insuficiente de orientações sobre o seu uso seguro e apropriado.

Viu-se que, até o presente momento, por mais sofisticados e inovadores que sejam, a grande maioria das tecnologias de inteligência artificial possui uma característica comum fundamental para a análise da responsabilidade: é possível identificar, por trás de seu agir, as diretivas humanas responsáveis por determinado funcionamento ou tomada de decisão,

³³¹ CHOPRA, Samir; WHITE, Lawrence. *A Legal Theory ...* Op. cit., p. 122.

rastreando-a diretamente de volta à programação do software, design da máquina, ou dados que lhe foram alimentados ao longo do processo de aprendizagem.³³²

Onde o envolvimento humano na tomada de decisão é evidente, não há necessidade de se reexaminar modelos existentes de responsabilidade. Qualquer ser humano (ou entidade personificada) que tem um papel no desenvolvimento da inteligência artificial e ajuda a mapear sua tomada de decisão é potencialmente responsável por atos ilícitos - negligentes ou intencionais - cometidos ou envolvendo a IA.

Nesse casos, a aplicação de modalidades já existentes de responsabilidade civil pode ser feita a uma inteligência artificial autônoma da mesma forma como o é para outra máquina ou ferramenta utilizada por um ser humano, talvez com a exceção de se considerar um padrão mais elevado de cuidado quando da análise de um dano.³³³

Na medida em que sistemas artificiais passem a ser capazes de atuar de maneira verdadeiramente autônoma, independentemente de um input humano direto e baseando-se em informações que o próprio sistema adquire e analisa, será possível que eles executem determinadas ações não antecipadas por seus criadores, muitas delas com consequências danosas.

Assim, a segunda questão adquire relevância se, e somente se, uma IA totalmente autônoma causar danos de maneiras impassíveis de serem rastreadas ou atribuídas a um ser humano.

Nessas hipóteses, questiona-se qual deveria ser o regime de responsabilidade adotado, especialmente quando se tem em mente que o sistema de responsabilidade civil não está disposto, na maior parte das vezes, a deixar que uma pessoa que sofreu um dano sem ter contribuído para o mesmo tenha que suportar sozinha o fardo que lhe foi imposto. A pergunta, então, passa a ser quem deverá arcar com os custos associados a um incidente envolvendo um agente artificial autônomo. Em outras palavras, “quem paga?”³³⁴.

1. Danos diretamente atribuíveis a seres humanos

Em geral, os usuários e operadores de uma inteligência artificial têm o dever de empregar um certo grau de cuidado para impedir que ela cause danos a outros indivíduos ou suas propriedades. O padrão de conduta exigido é aquele que um homem médio, ou seja, um

³³² VLADECK, David. Op. cit., p. 120.

³³³ Id., p. 127.

³³⁴ Id., p. 128.

indivíduo razoável ou diligente, teria nas mesmas circunstâncias, e proporcional ao perigo envolvido na atividade.

Esse entendimento é atualmente aplicado a robôs industriais com mobilidade limitada, e que se comportam de maneiras relativamente previsíveis. Devido à sua potencialidade lesiva é que tais máquinas costumam ter suas proximidades intransponíveis aos demais operários, de modo a evitar a responsabilidade de seus operadores por falhas em manter os demais fora de perigo.³³⁵

Agentes artificiais também podem ser manuseados de forma descuidada ou perigosa; um operador pode não tomar o devido cuidado ao programar ou configurar um agente antes de disponibilizá-lo aos usuários em uma rede pública.³³⁶ É possível ainda que usuários empreguem uma IA como meio de cometer ações danosas a terceiros, ou então que alimentem um sistema de aprendizado de máquina com conteúdo inapropriado, gerando, por exemplo outputs racistas.³³⁷ Ademais, conforme destacado por Kurki, os programadores e/ou proprietários podem também ser responsabilizados caso tenham intencionalmente criado uma inteligência artificial com o intuito de cometer ilícitos.³³⁸

Uma IA projetada para certa finalidade específica pode, ademais, causar danos ao tentar alcançá-la, através de algum desenvolvimento imprevisível. Nesse sentido, ilustra-se a hipótese com o exemplo de uma torradeira inteligente que queima a casa em busca de otimizar seus resultados³³⁹, ou então com o famoso experimento mental proposto por Bostrom de uma máquina cujo objetivo é produzir o maior número possível de cliques de papel e, no processo, acabar por destruir a humanidade (*“paperclip maximizer”*).

A bem da verdade, quanto mais imprevisível for o modo de falha de uma inteligência artificial, mais difícil será responsabilizar o usuário ou programador sem recorrer a algum tipo de responsabilidade objetiva.

Na tentativa de contornar essa questão, Ryan Abbot propõe que, nos casos em que o desenvolvedor de uma IA consiga demonstrar que a mesma é efetivamente mais segura do que um homem médio, o mesmo deveria ser responsabilizado de forma subjetiva, e não

³³⁵ CHOPRA, Samir; WHITE, Lawrence. A Legal Theory ... Op. cit., p. 126.

³³⁶ Id., p. 127.

³³⁷ HUNT, Elle. Tay, Microsoft's AI chatbot, gets a crash course in racism from Twitter. *The Guardian*, 2017. Disponível em: <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>. Acesso em: 14 abr. 2020.

³³⁸ KURKI, Visa A. J. *A Theory of Legal Personhood*. Oxford: Oxford University Press, 2019, p. 181.

³³⁹ TURNER, Jacob. Op. cit., p. 89.

objetiva, pelos danos causados por ela causados, ainda que no âmbito de uma relação de consumo.³⁴⁰

A análise da negligência, contudo, fundamenta-se no conceito de previsibilidade, e levanta questões como: “era possível antever o dano?” e “quais tipos de danos eram de fato previsíveis?”. Conforme bem destaca Turner, é extremamente provável que as ações de uma IA irão se tornar cada vez mais imprevisíveis. Como consequência, responsabilizar um ser humano por toda e qualquer ação da IA mudaria o foco da culpa do agente para um sistema de responsabilidade objetiva.³⁴¹

Nesse contexto enquadra-se a teoria que compreende que a inteligência artificial constitui mera ferramenta ou instrumentalidade nas mãos de seres humanos (“*AI-as-tool*”), podendo representar apenas uma fonte de responsabilidade para outros agentes, seguindo a regra geral de que o operador de uma ferramenta é responsável pelos resultados obtidos pelo seu uso, vez que a mesma não possui volição independente própria.³⁴²

No capítulo II deste trabalho, destacou-se ser esta a opção adotada no artigo 12 da Convenção das Nações Unidas sobre a Utilização de Comunicações Eletrônicas em Contratos Internacionais, segundo o qual a pessoa – natural ou jurídica - em cujo nome um computador foi programado deve ser responsável por qualquer mensagem gerada pela máquina³⁴³.

Segundo Pagallo, a teoria da IA como ferramenta implica na atribuição objetiva de responsabilidade à pessoa física ou jurídica em nome de quem ela age, independentemente de seu comportamento ter sido planejado ou previsto, e teria as seguintes repercussões:

- (i) Uma inteligência artificial (IA) atua em nome de um agente principal (P), de modo a negociar e celebrar um contrato com a contraparte (C);
- (ii) Os direitos e obrigações estabelecidos pela IA vinculam diretamente P, uma vez que todos os atos da IA são considerados como atos emanados de P;
- (iii) P não pode evadir a responsabilidade alegando que não pretendia celebrar tal contrato, ou que a IA cometeu um erro decisivo;
- (iv) No caso de comportamento errático por parte da IA, P poderia reclamar danos contra o criador ou fabricante da IA. Entretanto, de acordo com o mecanismo

³⁴⁰ ABBOTT, Ryan. The Reasonable Computer: Disrupting the paradigm of tort liability. *The George Washington Law Review*, v. 86, n. 1, 2018, p. 1-45.

³⁴¹ TURNER, Jacob. Op. cit., p. 90-91.

³⁴² ČERKA, Paulius. Op cit., p. 384.

³⁴³ “As comunicações eletrônicas que são geradas automaticamente por sistemas de mensagens ou computadores sem intervenção direta humana devem ser consideradas como “originárias” da entidade jurídica em nome da qual o sistema de mensagens ou computador é operado.” Nota Explicativa n. 213 do Secretariado da UNCITRAL.

de ônus da prova (adotado em alguns países como os Estados Unidos), P deveria demonstrar que a IA era defeituosa, e que tal defeito já existia quando estava sob o controle do fabricante; e, além disso, que o defeito foi a causa imediata das lesões sofridas por P.³⁴⁴

A teoria da inteligência artificial como ferramenta traça um paralelo com a chamada responsabilidade vicária, utilizada no sistema contemporâneo de responsabilidade extracontratual na *common law*, e que designa a responsabilidade do superior hierárquico pelos atos dos seus subordinados ou, em um sentido mais amplo, a responsabilidade de qualquer pessoa que tenha o dever de vigilância ou de controle pelos atos ilícitos praticados pelas pessoas a quem deveriam vigiar – como a responsabilidade do patrão pelos atos dos seus empregados, ou dos pais pelos atos dos filhos menores³⁴⁵.

No ordenamento pátrio, seriam os casos de responsabilidade pelo fato de terceiro, derivada de um dever de guarda, vigilância e cuidado, nos termos do art. 932 do Código Civil, como a já mencionada responsabilidade dos pais pelos atos dos filhos menores que estiverem sob o seu poder e em sua companhia; do tutor e curador pelos pupilos e curatelados; e do patrão pelos atos dos seus empregados.³⁴⁶ Assim, para que reste configurada, deve haver uma relação entre o agente causador do dano e o indivíduo a quem o mesmo será imputado, dentro das categorias reconhecidas, e o ato ilícito deve ocorrer dentro do escopo desse relacionamento.³⁴⁷

Este modelo de responsabilidade civil desenvolveu-se a partir do deslocamento de modalidades antigas de relações de trabalho, como a servidão e a escravidão, em que o mestre respondia pelos danos causados por seus servos e escravos (*“respondeat superior”*)³⁴⁸. Ele não se baseia em uma interconexão lógica que vincula o infrator a um dano que ele próprio causou, mas em uma política de compensação de perdas, ao invés de imputar a responsabilidade a partes financeiramente incompetentes³⁴⁹.

Sob a ótica econômica, a responsabilidade objetiva de, por exemplo, empregadores por atos de seus empregados, justifica-se ainda na ausência de negligência na seleção, treinamento

³⁴⁴ PAGALLO, Ugo. Op. cit., p. 98.

³⁴⁵ PIRES, Thatiane; SILVA, Rafael. A Responsabilidade Civil pelos Atos Autônomos da Inteligência Artificial: Notas iniciais sobre a resolução do Parlamento Europeu. *Revista Brasileira de Políticas Públicas*, v. 7, n. 3, 2017, p. 239-254.

³⁴⁶ Id., p. 248.

³⁴⁷ TURNER, Jacob. Op. cit., p. 100.

³⁴⁸ ČERKA, Paulius. Op. cit., p. 385.

³⁴⁹ WEIN, Leon. Op. cit., p. 110.

e supervisão dos empregados, pois a maioria destes, por não possuir os recursos necessários para arcar com a restituição dos danos causados, não responde à ameaça da responsabilização. Do ponto de vista do empregador, a responsabilidade objetiva é eficaz porque alterações por ele promovidas, como a redução de escalas, são métodos eficientes para reduzir danos cometidos por funcionários.³⁵⁰

Para Chopra e White, o argumento econômico é também relevante no caso de agentes artificiais, que igualmente não estariam suscetíveis a ameaças de sanção enquanto não fossem considerados como sujeitos de direito e tivessem um patrimônio autônomo próprio. Segundo os autores, haveria uma justificativa para se imputar a responsabilidade objetiva a seus proprietários ou usuários, sob o fundamento de que a IA poderia representar riscos a terceiros. Assim, esse regime de responsabilidade incentivaria os indivíduos que empregassem agentes artificiais em suas atividades a apenas fazê-lo quando os benefícios superassem os custos sociais.³⁵¹

A atribuição de responsabilidade a determinado sujeito com base não no ato ilícito em si, mas sim em virtude da sua relação com aquele que o perpetrou resultaria, portanto, no caso de agentes artificiais, na responsabilização de seus proprietários ou usuários³⁵². A depender de quem está fazendo uso da IA, seria possível arguir uma responsabilidade distinta nos casos em que ela é utilizada por empresas para prestar serviços e oferecer produtos — situação em que ela age em nome de um fornecedor —, e nos casos em que ela é empregada pelo usuário para desempenhar determinadas atividades sob a supervisão deste³⁵³.

Turner exemplifica com a situação hipotética de uma força policial que fizesse uso de um robô de patrulha, e este agredisse um membro inocente do público durante sua atuação. Mesmo que a força policial não tenha criado o sistema de IA empregado no robô, tenha desejado ou permitido a sua ação, ela ocorreu dentro do escopo da função designada do robô e, portanto, poderia ser responsabilizada pela conduta do mesmo³⁵⁴.

O fato de que a responsabilidade vicária é normalmente limitada a determinada esfera de atividades realizadas pelo agente representa tanto uma vantagem quanto uma desvantagem. Isso porque nem todo ato de uma IA seria necessariamente atribuível ao seu proprietário ou operador, de tal modo que, quanto mais o seu comportamento desviasse da sua gama de tarefas delineadas, maior a probabilidade de haver uma lacuna de responsabilidade. No curto a

³⁵⁰ POSNER, Richard. *Economic analysis of law*. New York: Wolters Kluwer Law and Business, 2014.

³⁵¹ CHOPRA, Samir; WHITE, Lawrence. *A Legal Theory ...* Op. cit., p. 129.

³⁵² ČERKA, Paulius. Op. cit., p. 385.

³⁵³ PIRES, Thatiane. Op. cit., p. 248.

³⁵⁴ TURNER, Jacob. Op. cit., p. 100-101.

médio prazo, porém, em se tratando de agentes que continuam operando dentro de faixas muito limitadas (*narrow AI*), essa preocupação é menos premente³⁵⁵.

Similarmente à hipótese de responsabilidade objetiva por ato de terceiros, é possível vislumbrar também a responsabilidade objetiva dos usuários ou operadores de uma IA em analogia à responsabilidade do dono do animal pelo fato deste. Neste caso, quando um agente artificial for potencialmente perigoso e seja provável que um dano ocorra caso ele escape de seus limites físicos ou virtuais, a pessoa em custódia do agente poderia ser responsabilizada, mesmo na ausência de negligência comprovada.³⁵⁶

A escolha de responsabilizar o usuário ou o operador de um agente de IA dependeria do nível relativo de controle exercido por cada um deles, sendo, a princípio, o operador a pessoa mais adequada a prevenir danos na maioria dos casos. A falta de conhecimento do potencial de perigo apresentado por um agente artificial poderia também ser relevante no caso de agentes altamente sofisticados, cujo funcionamento interno não é transparente para seus guardiões (embora possam ser para seus programadores originais) ou quando o agente exibir um comportamento inesperado, seja como resultado de seu aprendizado ou de algum defeito³⁵⁷.

Em última instância, a responsabilização do proprietário ou usuário de uma IA não representa uma solução ideal, uma vez que eles seriam as vítimas mais prováveis de danos não antecipados e porque os fabricantes estariam, em princípio, na melhor posição para aprimorar a segurança do produto e pesar os riscos e benefícios da criação e disponibilização de novas tecnologias.³⁵⁸

Verifica-se, portanto, que, a depender do caso, os danos causados por uma inteligência artificial poderão atrair as disposições sobre a responsabilidade pelo produto, sobretudo se o fabricante não informa suficientemente o consumidor acerca dos riscos associados ao uso da tecnologia em questão, ou se os sistemas de segurança desta forem deficientes a ponto de não oferecerem a segurança esperada.

Desse modo, a recuperação pelo dano pode ser inviabilizada ou reduzida caso ele tenha sido integral ou parcialmente causado pelo modo de utilização do agente, em especial pelo usuário: ignorando os avisos ou instruções do fornecedor, usando o agente em um

³⁵⁵ Id., p. 101.

³⁵⁶ CHOPRA, Samir; WHITE, Lawrence. A Legal Theory ... Op. cit., p. 130.

³⁵⁷ Id., p. 131

³⁵⁸ ABBOTT, Ryan. Op. cit., p. 31.

contexto novo ou inesperado, imputando-lhe dados ou lhe dando instruções inadequados, ou deixando de exercer discricção ao depender fielmente no output apresentado.³⁵⁹

Uma vez cumpridos os deveres de informação e de segurança impostos ao fornecedor, e não havendo excludentes de responsabilidade³⁶⁰, permanece, ainda, a dúvida sobre a aplicação da responsabilidade civil pelos riscos do desenvolvimento. Sob a égide da Diretiva 85/374/CEE da União Europeia, por exemplo, o fabricante não poderia ser responsabilizado caso prove que o estado dos conhecimentos científicos e técnicos no momento em que o produto foi colocado em circulação não lhe permitiu detectar a existência do defeito.³⁶¹

No ordenamento pátrio, por vezes, a questão é ainda controversa, argumentando-se, de lado, pela responsabilização do fornecedor, uma vez que foi ele quem tomou a iniciativa de colocar o produto no mercado, assumindo os riscos, e não o consumidor. Por outro lado, a defesa dos riscos do desenvolvimento como excludente da responsabilidade pauta-se na impossibilidade de exigir do fornecedor o conhecimento de algo que, à época, era imensurável, e que nessa hipótese inexistia um produto defeituoso, nos termos previstos pelo Código de Defesa do Consumidor.³⁶²

Além disso, Čerka et. al chamam atenção para o fato de que, na maioria dos casos, seria difícil a aplicação da responsabilidade pelo fato do produto, pois o ônus da prova seria extremamente gravoso a quem incumbisse, tendo em vista a dificuldade em se provar um defeito na fabricação da IA e, principalmente, que esse defeito já existia quando ela deixou as mãos dos desenvolvedores. Como um sistema de autoaprendizagem, pode ser impossível traçar a linha tênue entre os danos resultantes desse fato e o defeito preexistente de fabricação do produto.³⁶³

2. Danos não diretamente atribuíveis a seres humanos

Quando um agente artificial autônomo causar um dano através de ações não diretamente rastreáveis a uma outra pessoa, uma abordagem possível para solucionar a lacuna

³⁵⁹ CHOPRA, Samir; WHITE, Lawrence. A Legal Theory ... Op. cit., p. 138.

³⁶⁰ No Código de Defesa do Consumidor brasileiro (Lei nº 8.078/1990), as principais excludentes de responsabilidade do fabricante estão elencadas no artigo 12, § 3º, segundo o qual o fabricante, o construtor, o produtor ou o importador só não será responsabilizado quando provar: que não colocou o produto no mercado; que o defeito inexistia; ou a culpa exclusiva do consumidor ou de terceiro (BRASIL, 1990).

³⁶¹ UNIÃO EUROPEIA. Diretiva 85/374/CEE do Conselho, de 25 de julho de 1985, relativa à aproximação das disposições legislativas, regulamentares e administrativas dos Estados-Membros em matéria de responsabilidade decorrente dos produtos defeituosos.

³⁶² ROSENVALD et al., Op. cit., p. 746.

³⁶³ ČERKA, Paulius. Op. cit., p. 386.

da responsabilidade seria a inferência de um defeito, baseando-se na teoria de que o acidente em si já constituiria prova da sua existência.³⁶⁴

Um defeito pode ser razoavelmente presumido quando, ocorrendo uma falha em determinado produto, ainda que seus engenheiros não sejam aptos a identificar aquilo que a justificou, esta mesma falha pode ser vislumbrada com certa frequência, seguindo um padrão comum também em outros produtos. Nessas circunstâncias, aplica-se rotineiramente a teoria *res ipsa loquitur*, concluindo que deve haver um defeito no produto, e não na atuação do usuário ou da vítima.³⁶⁵

Um exemplo recente diz respeito aos inúmeros casos envolvendo a Toyota em que se alegou que alguns veículos fabricados pela marca estariam propensos a acelerar subitamente, sem nenhuma ação correspondente do motorista, e sem a possibilidade de interrupção dessa aceleração por ele, desengatando o pedal do acelerador ou aplicando os freios.³⁶⁶

Não obstante a extensiva investigação realizada, os engenheiros da marca foram incapazes de identificar com precisão qualquer defeito que causasse a aceleração repentina. Ainda assim, devido ao número substancial de reclamações e às características comuns entre elas, a Toyota decidiu não contestar a sua responsabilidade pelos danos causados, pagando o equivalente a U\$1.3 bilhões de dólares aos proprietários dos veículos afetados.³⁶⁷

Para Vladeck, o exemplo da empresa Toyota deixa claro que o regime de responsabilidade objetiva aplicado às relações de consumo é adequado para lidar com casos em que as evidências sugerem fortemente um defeito, mas a tecnologia existente não é capaz de determinar a sua causa.³⁶⁸

Entretanto, no caso de um agente artificial autônomo causando um dano como resultado de sua atuação normal, dentro daquilo para o qual foi programado a fazer, não é possível falar na existência de um defeito ou de presunção de um defeito. Uma vez que esses agentes têm como características essenciais a autonomia e a capacidade de autoaprendizagem com base em suas experiências, seria virtualmente impossível distinguir uma ação gerada por uma falha daquela decorrente do próprio processo de aprendizado³⁶⁹.

³⁶⁴ VLADECK, David. Op. cit., p. 128.

³⁶⁵ Id., p. 142.

³⁶⁶ CARTY, Sharon. Toyota's Sudden Acceleration Problem May Have Been Triggered By Tin Whiskers. *Huffpost*, jan. 2012. Disponível em: https://www.huffpostbrasil.com/entry/toyota-sudden-acceleration-tin-whiskers_n_1221076?ri18n=true. Acesso em: 20 jun. 2020.

³⁶⁷ DYE, Jessica. Toyota Acceleration Case Settlement Gets Final OK. *Insurance Journal*, jul. 2013. Disponível em: <https://www.insurancejournal.com/news/national/2013/07/22/299154.htm>. Acesso em: 20 jun. 2020.

³⁶⁸ VLADECK, David. Op. cit., p. 143.

³⁶⁹ ČERKA, Paulius. Op. cit., p. 387.

É precisamente essa capacidade de gerar soluções inovadoras que torna o uso da IA atraente em uma variedade cada vez maior de campos, gerando um incentivo econômico para o desenvolvimento desse tipo de tecnologia. Mas, conforme bem colocado por Scherer, é importante destacar que, ainda que tais agentes possam agir de maneira imprevisível, não sendo possível antever cada uma de suas ações, no mais das vezes essa habilidade em si é intencionalmente desejada pelos seus desenvolvedores³⁷⁰ – constituindo um objetivo declarado do próprio campo da inteligência artificial³⁷¹.

Nos casos em que danos sejam causados sem que seja possível atribuí-los diretamente a um comportamento humano, caberá ao ordenamento jurídico optar pela resposta que melhor atenda aos interesses coletivos das partes afetadas. Nesse sentido, a adoção de um regime de responsabilidade objetiva, desacoplado de noções de falhas ou defeitos do produto, parece ser a alternativa mais viável.

Nesse contexto, David Vladeck destaca quatro argumentos favoráveis à adoção de um regime de responsabilidade civil objetiva para casos em que danos sejam causados por uma inteligência artificial autônoma. Em primeiro lugar, o autor destaca que fornecer reparação para pessoas que sofrem um dano sem ter tido qualquer participação para a ocorrência do mesmo é um valor importante por si só. A ideia de que os indivíduos devem arcar com danos e perdas que simplesmente lhe ocorrem, mesmo que seja impossível explicar causalmente uma falha ou defeito, contraria noções básicas de justiça compensatória e distribuição de riscos entre a sociedade.³⁷²

Maria Celina Bodin de Moraes explica que a injustiça de um dano se dá quando não for razoável que a vítima permaneça sem reparação, ponderados os interesses contrapostos.³⁷³ Nesse sentido, a operatividade do princípio da solidariedade social no âmbito da responsabilidade civil extracontratual é realizada por meio da definição e delimitação do requisito da injustiça do dano. Verifica-se, aqui, a volta paradigmática no direito de danos, de um setor que estabelecia a obrigação de indenizar atribuída àquele que age culposamente, isto é, como uma sanção a alguma violação cometida pelo sujeito (derivada da noção de justiça retributiva), para a ideia dissociação da responsabilidade da concepção de conduta culposa a ser punida (por meio da realizada de justiça distributiva).³⁷⁴

³⁷⁰ SCHERER, Matthew. Op. cit., p. 365.

³⁷¹ CALO, Ryan. Robotics and... Op. cit., p. 519.

³⁷² VLADECK, David. Op. cit., p. 146.

³⁷³ MORAES, Maria Celina Bodin. Danos à Pessoa Humana: Uma leitura civil constitucional dos danos morais. Rio de Janeiro: Editora Processo, 2000, p. 179.

³⁷⁴ MULHOLLAND, Caitlin. Op. cit., p. 334.

Poder-se-ia, claro, permitir o desenvolvimento e a aplicação de agentes artificiais autônomos e simplesmente aceitar os riscos e custos a nível social, sem desenvolver uma estrutura mais adequada para regular a sua autonomia. Essa abordagem permissiva permitiria, em um primeiro momento, muitas aplicações benéficas de IAs avançadas, mas também muitas outras prejudiciais, incluindo-se inúmeros danos pelos quais ninguém poderia ser responsabilizado, e cujas vítimas restariam sem qualquer tipo de compensação. Como efeito secundário, porém, provavelmente haveria uma reação geral contra tais usos avançados da tecnologia, vez que ela passaria a ser vista como prejudicial e oferecendo poucas opções para a restituição em caso de acidentes.³⁷⁵

O segundo argumento favorável a um regime de responsabilidade objetiva apontado por Vladeck consiste no fato de que, ao contrário da parte lesada, os desenvolvedores da inteligência artificial estão em melhor posição de prevenir a ocorrência de danos, e de absorver os custos associados, redistribuindo o ônus da perda através da definição do preço do produto. Afinal, é apenas razoável que os custos de acidentes inexplicáveis sejam suportados, ao menos em parte, por aqueles que se beneficiam de inovações redutoras de risco.³⁷⁶

Além disso, segundo o autor, um regime de responsabilidade objetiva pouparia todas as partes interessadas dos elevados custos de transação envolvidos em litigar questões de responsabilidade envolvendo agentes autônomos de IA, em que não é possível estabelecer uma falha. Na medida em que a complexidade de um produto aumenta geometricamente, os custos de um litígio envolvendo acidentes de consumo aumenta exponencialmente. Para muitas empresas, é menos custoso compensar as vítimas por eventuais danos causados do que arcar com os valores envolvidos na contratação de advogados e especialistas.

De fato, retornando ao exemplo da Toyota, Vladeck destaca que, embora não tenham havido confirmações públicas nesse sentido, é muito provável que a empresa tenha decidido pelo fechamento de um acordo nos casos de aceleração repentina - não porque temia perdê-los se fossem a julgamento, mas porque litigar cada um desses casos, ainda que com resultados favoráveis, custaria à Toyota muito mais do que aceitar um acordo. Isso, repita-se, em virtude da complexidade tecnológica envolvida nos casos.

Finalmente, o autor considera que um esquema de responsabilidade previsível pode ser melhor em estimular a inovação do que um sistema menos previsível que depende de uma análise e posterior atribuição de culpa. O sistema de responsabilidade adotado não deverá ser um empecilho à inovação, mas sim incentivar que a mesma seja feita de maneira responsável.

³⁷⁵ ASARO, Peter. Op. cit., p. 193.

³⁷⁶ VLADECK, David. Op. cit., p. 146-147.

A estabilidade fornecida por um regime de responsabilidade civil objetiva, juntamente com uma abordagem de disseminação de custos, serviria sem dúvida a esse objetivo melhor do que um sistema de responsabilidade incerto baseado na análise de elementos subjetivos.³⁷⁷

Similarmente, Turner destaca que, do ponto de vista dos desenvolvedores e fornecedores de uma inteligência artificial, a certeza da sua responsabilização na ocorrência de um dano poderia representar uma vantagem, pois permitiria a realização de cálculos atuariais mais precisos. O risco de danos poderia, desse modo, ser precificado no custo final dos produtos, bem como antecipado nas previsões contábeis das empresas e na divulgação, aos investidores, como um fator de risco em determinado prospecto.³⁷⁸

A lógica econômica por trás da responsabilidade objetiva baseia-se no fato de que os usuários (em especial, consumidores) podem fazer pouco, a um custo razoável, para evitar um raro acidente. A imposição do custo de acidentes ao desenvolvedor levará, como consequência, ao aumento de preços, resultando na substituição por alternativas menos perigosas. Desse modo, a circulação de produtos menos seguros diminuirá e, com ela, o número de acidentes. A responsabilidade objetiva efetivamente acrescenta, portanto, os riscos da tecnologia ao seu preço final, causando a substituição de tecnologias perigosas cujos riscos os usuários sequer podem ter consciência sobre.³⁷⁹

Magrani et al. destacam que a partir da análise de eficiência e maximização de riqueza (como medida alternativa de utilidade na impossibilidade de se mensurar o nível de satisfação individual), é possível que a regulação imposta quanto à alocação da responsabilidade tenha incentivos distintos na esfera jurídica tanto da vítima como do ofensor, no caso o desenvolvedor de uma IA:

Ao se estabelecer a responsabilidade objetiva, cria-se um incentivo significativo para que o ofensor atue diligentemente, de modo a reduzir os custos de expectativa de dano. De fato, no modelo econômico da responsabilidade objetiva, o ofensor responde ainda que adote um nível de precaução alto. Isso não significa que não há interesse na adoção de um comportamento precavido. Exemplifica-se: existe um nível de precaução em que o ofensor, no âmbito da responsabilidade objetiva afastará a ocorrência de dano. Nesse sentido, se a adoção do nível de precaução for inferior ao custo esperado dos danos, do ponto de vista econômico, é desejável que seja adotado o nível de precaução. Mas, registre-se, ainda que o ofensor adote um comportamento diligente, caso a vítima sofra um dano, esta será ressarcida, o que favorece, nesse caso, a posição da vítima.³⁸⁰

³⁷⁷ Idem.

³⁷⁸ TURNER, Jacob. Op. cit., p. 95.

³⁷⁹ POSNER, Richard. Op. cit., p. 136.

³⁸⁰ MAGRANI, Eduardo; SILVA, Priscilla; VIOLA, Rafael. Novas Perspectivas sobre Ética e Responsabilidade de Inteligência Artificial. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 117-147.

O regime de responsabilidade civil subjetiva, por outro lado, apresenta um quadro diverso. Como o ofensor apenas responde se agir culposamente, caso adote um comportamento diligente, o ônus do dano recairá, necessariamente, sobre a vítima, ainda que este tenha sido produzido em virtude de uma atividade potencialmente perigosa. Ou seja, o incentivo para que as vítimas adotem níveis de precaução é maior, pois, caso venham a sofrer algum tipo de prejuízo, elas o suportarão.³⁸¹

Nesse ponto, Steven Shavell aponta para a dificuldade de se valer de um padrão de negligência (*"negligence rule"*) quando os tribunais desconhecem o que efetivamente caracterizaria a culpa em razão da natureza técnica da atividade. Assim, acabam por estabelecer padrões de comportamento errôneos, permitindo que as vítimas permaneçam sem a devida reparação.³⁸² Ainda, conforme previamente ressaltado, lembre-se que no caso de danos causados por uma IA avançada, seria extremamente difícil saber como o programa chegou a certo output ou mesmo passam a gerar consequências inesperadas e possivelmente danosas, apontando para a "culpa" de um dos membros da cadeia de desenvolvimento.

Por fim, cumpre ressaltar que, embora à primeira vista o comportamento imprevisível de um agente autônomo de IA possa sugerir a aplicabilidade de responsabilidade objetiva em sua versão agravada³⁸³, isto é, fundamentando-se apenas na análise de ocorrência do dano, com a eliminação do requisito do nexo causal, existem empecilhos a serem considerados para a qualificação de um agente artificial autônomo como atividade de risco.

Na responsabilidade objetiva agravada, para que haja a imputação, exige-se que o dano possua estreita conexão com a atividade exercida pelo responsável, devendo este ser considerado inerente, característico ou típico daquela. Ainda, tal modalidade exige que o dano ocorrido afete a integridade física ou psíquica de uma pessoa, não bastando, para que haja o agravamento, que o evento danoso tenha atingido apenas objetos ou bens.³⁸⁴

A responsabilidade agravada justifica-se pelo objetivo econômico da atividade normalmente desenvolvida. Em outras palavras, aquele que se beneficia de uma atividade lícita e que seja potencialmente perigosa adquire o ônus de arcar com as eventuais consequências danosas inerentes ao processo produtivo ou distributivo.

³⁸¹ Id., p 119-120.

³⁸² SHAVELL, Steven. *Foundations of Economic Analysis of Law*. Cambridge: Harvard University Press, 2004.

³⁸³ No Código Civil brasileira, a responsabilidade objetiva agravada está preceituada na parte final do parágrafo único do artigo 927: Haverá obrigação de reparar o dano, independentemente de culpa, nos casos especificados em lei, ou quando a atividade normalmente desenvolvida pelo autor do dano implicar, por sua natureza, risco para os direitos de outrem (BRASIL, 2002).

³⁸⁴ NORONHA, Fernando. Responsabilidade Civil: Uma tentativa de ressystematização. *Revista de Direito Civil, Imobiliário, Agrário e Empresarial*. São Paulo, ano 17, n. 64, abr./jun. 1993, p. 7-47.

Através da definição, em lei, de atividades consideradas intrinsecamente de risco, impõe-se a responsabilidade objetiva agravada às atividades que não podem ser prevenidas, de maneira viável, pelo cuidado do ator ou por possíveis alterações comportamentais das vítimas. Isso porque, segundo Posner, novas atividades tendem a ser perigosas por haver pouca experiência em lidar com quaisquer perigos que elas apresentem, implicando, por vez, que existem bons substitutos para elas. O melhor método de controle de acidentes, nesse caso, encontrar-se-ia na redução de escala da atividade em questão.³⁸⁵

Entretanto, a qualificação da pesquisa e desenvolvimento de agentes de inteligência artificial autônomos como atividade inerentemente de risco é, no mínimo, problemático, pelo simples fato de que na grande maioria das vezes eles desempenharão suas tarefas de forma mais segura do que seus correspondentes humanos, e serão menos perigosos do que os produtos que virão a substituir.

Veículos autônomos ilustram bem a questão, por serem mais previsíveis do que os motoristas humanos em muitos casos, seguindo infalivelmente suas instruções e as regras de trânsito.³⁸⁶ Motoristas humanos são menos previsíveis nesse sentido, porque estão sujeitos a distrações, dirigem sob o efeito de substâncias que limitam sua atenção e tempo de resposta, ou simplesmente não seguem as regras de trânsito.³⁸⁷

Além disso, o agravamento da responsabilidade objetiva impediria, em qualquer instância, a possibilidade de que desenvolvedores dessa tecnologia pudessem se eximir de arcar com qualquer dano ocorrido, alegando, por exemplo, a culpa exclusiva da vítima ainda que ela fizesse uso de maneira completamente inapropriada do agente, ou fortuito externo.

Inexiste, aqui, o enquadramento da inteligência artificial (ao menos em cenários não futuristas de criação de uma IA geral ou de uma superinteligência) naquilo que Peter Huber descreve como risco público (*“public risk”*): ameaças à saúde ou à segurança humana que são produzidas em centralmente ou em massa, amplamente distribuídas e completamente fora do entendimento e controle diretos dos indivíduos que suportam o risco, como a tecnologia nuclear.³⁸⁸

³⁸⁵ POSNER, Richard. Op. cit., p. 180.

³⁸⁶ PANETTA, Kasey. Why Humans Are the Problem with Autonomous Cars, *EE World Online*, set. 2015. Disponível em: <https://www.ecnmag.com/blog/2015/09/why-humans-are-problem-autonomous-cars>. Acesso em: 19 jun. 2020.

³⁸⁷ RUTKIN, Aviva. Autonomous Cars Are Learning Our Unpredictable Driving Habits, *New Scientist*, ago. 2015. Disponível em: <https://www.newscientist.com/article/mg22730362-900autonomous-cars-are-learning-our-unpredictable-driving-habits>. Acesso em: 19 jun. 2020.

³⁸⁸ HUBER, Peter. Safety and the Second Best: The hazards of public risk management in the courts. *Columbia Law Review*, v. 85, n. 2, 1985, p. 277-337.

Uma vez decidido que optado por um sistema de responsabilidade civil baseado em elementos objetivos, o principal questionamento passa a ser quem deve arcar com os custos. O ônus da responsabilidade deve recair sobre o desenvolvedor, o distribuidor, o operador ou, ainda, a todos eles?

Em sua obra seminal “Sociedade de Risco”³⁸⁹, publicada na Alemanha em 1986, o sociólogo Ulrich Beck alertou para o fato de que a sociedade industrial da época moderna estava passando por profundas mudanças que constituam o próprio reflexo da modernização simples, e não resultantes de forças exteriores ao processo produtivo industrial ou de revoluções sociais. Por esse motivo, tal processo foi compreendido como uma segunda modernidade, ou uma modernização reflexiva, na qual se evidenciam as antinomias e insuficiências presentes na primeira modernidade – a era dos efeitos colaterais.

Para o sociólogo, os riscos apresentados pela modernidade reflexiva não mais constituiriam apenas um evento limitado local e temporalmente, mas sim ultrapassam as fronteiras nacionais em intensa globalização. Incerteza e irreversibilidade constituem, na sociedade de risco, elementos inerentes do dano.

Os riscos da modernidade apontados por Beck compreenderiam, por exemplo, a ameaça nuclear e os danos à natureza, ambos socializados e indiferentes às tradicionais estruturas de classe, pois deles não seria possível se subtrair.

(...) primeiro, que riscos da modernização emergem ao mesmo tempo vinculados espacialmente e *desvinculadamente com um alcance universal*; e segundo, quão *incalculáveis e imprevisíveis* são os intrincados caminhos de seus efeitos nocivos. Nos riscos da modernização, portanto, algo que se encontra conteudístico-objetiva, espacial e temporalmente apartado acaba sendo causalmente congregado e, desse modo, além do mais, colocado simultaneamente numa relação de responsabilidade social e jurídica.³⁹⁰

Ao tratar das relações, na modernidade reflexiva, de responsabilidade social e jurídica, Beck destaca que a interdependência sistêmica dos atores da modernização na economia, na agricultura, no direito e na política corresponderia à ausência de causas específicas e, principalmente, de responsabilidades isoláveis. Nesse contexto, depender-se-ia de suposições causais, a serem sempre conceitualmente adicionadas e presumidas como verdadeiras.³⁹¹

Em outras palavras, “a altamente diferenciada divisão do trabalho implica uma cumplicidade geral e esta, por sua vez, uma irresponsabilidade generalizada. Todos são causa

³⁸⁹ No original em alemão *Risikogesellschaft*.

³⁹⁰ BECK, Ulrich. *Sociedade de Risco: Rumo a uma outra modernidade*. [Trad.] Sebastião Nascimento. São Paulo: Editora 34, 2011.

³⁹¹ Id., p. 38.

e efeito, e portanto uma *não causa*”³⁹². Tem-se, aqui, o problema da atribuição de responsabilidade dentro de uma enorme cadeia de acontecimentos interligados (por exemplo, a criação de um agente artificial autônomo), e a causalidade como uma opção normativa.

A mera imputação da responsabilidade ao fabricante da IA, permitindo-lhe, posteriormente, buscar a compensação das outras partes potencialmente responsáveis, pode constituir em nada mais do que um gesto vazio. Se de fato for impossível identificar a causa do acidente, o fabricante provavelmente não terá argumentos arrazoáveis para fundamentar uma ação de regresso, ficando, desse modo, incumbido de arcar inteiramente com os custos de reparação à vítima.

Tal alternativa faz apenas sentido caso o fabricante esteja na melhor posição para suportar as perdas. Caso contrário, a opção mais justa pode ser a partilha da responsabilidade entre todas as partes que participaram na criação e manutenção do agente autônomo, sob a justificativa de que os custos associados ao dano são melhor distribuídos entre todas as partes potencialmente responsáveis ou entre as partes que poderiam, de maneira mais eficiente, arcar com ou segurar as perdas.³⁹³

Nesse ponto, é interessante a abordagem pela teoria *deep pocket* (literalmente, “bolso profundo”), conforme a denominação cunhada no direito norte-americano. Segundo ela, toda pessoa envolvida em atividades que apresentam riscos, mas que, ao mesmo tempo, são lucrativas e úteis para a sociedade, tem a obrigação de compensar os danos causados, como uma espécie de moeda de troca pelo lucro obtido. Seja o criador da IA, o fabricante de produtos que empregam IA, uma empresa ou profissional faz uso dela em sua atividade, mesmo não estando diretamente inserido na cadeia produtiva (como as empresas que utilizam *robot traders* para otimizar transações na bolsa de valores). Em suma: aquele que tem o “bolso profundo” e aproveita os lucros dessa nova tecnologia deve ser o garantidor dos riscos inerentes à sua atividade³⁹⁴.

Regimes existentes de responsabilidade solidária objetiva permitem, com base nessa teoria, que os prejudicados obtenham indenizações nos bolsos mais profundos entre as partes que compartilham uma parte da responsabilidade, sobretudo em se tratando de uma grande corporação ou do governo, que possuem capacidade para arcar especialmente com danos de valor elevado.³⁹⁵

³⁹² Id., p. 39.

³⁹³ VLADECK, David. Op. cit., p. 128-129.

³⁹⁴ PIRES, Thatiane. Op. cit., p. 251.

³⁹⁵ ASARO, Peter. Op. cit., p. 193. Para o autor, uma possível consequência negativa da aplicação da teoria seria que os fabricantes de tecnologias de inteligência artificial que provavelmente teriam que arcar com o ônus da

O efeito decorrente da impossibilidade de identificação do nexo causal entre a conduta do agente e o dano produzido – a chamada “irresponsabilidade distribuída” - foi abordada por Caitlin Mulholland em sua tese sobre presunção de causalidade. Segundo a autora, no cenário em tela a falta de clareza quanto ao nexo causal e a conduta dos agentes envolvidos reforçaria o conceito de causalidade alternativa. Este permite identificar que o dano foi causado por uma única conduta que, devido à característica de coesão do grupo, resta impossível atestar. Assim, com o intuito de buscar o ressarcimento da vítima, presume-se o nexo de causalidade. Segundo a Mulholland:

No caso de existirem várias atividades, sendo que cada uma delas, por si só, teria sido suficiente para produzir o dano, mas em que persiste incerteza sobre qual efetivamente o causou, cada uma será considerada como causa do dano até o limite correspondente à probabilidade de o ter causado. [...] Existe um único nexo causal que não pode ser identificado de forma direta. Daí a sua presunção em relação ao grupo como um todo. [...] O que se busca com a causalidade alternativa é possibilitar a reparação dos danos causados através da facilitação do ônus probatório. Ao invés de a vítima ter que provar que determinada pessoa através de sua conduta causou o dano que a afligiu, poderá contar com a presunção da causalidade, sendo suficiente que prove que sofreu um dano e que o dano foi consequência de determinada atividade realizada por um determinado grupo.³⁹⁶

Magrani et al. explicam que a referida tese vai além dos casos da responsabilidade solidária entre os imputados causadores do dano; da responsabilidade atribuída de acordo com a contribuição causal de cada agente para a obtenção do resultado danoso; e da responsabilidade atribuída somente a um dos agentes, quando for possível identificar o rompimento do nexo de causalidade entre as condutas sucessivas.³⁹⁷

No caso de um dano causado por um agente artificial autônomo, a situação torna-se ainda mais complicada por englobar um somatório de agências humanas, instituições, e entes artificiais com autonomia e poder de agencia (no sentido de agência adotado neste trabalho) próprio. Em situações como essa, porém, o foco apenas na atuação conjunta dentro de um mesmo grupo econômico pode não ser suficiente para lidar com todas as questões que possam surgir a respeito da justa atribuição de responsabilidade.

Similarmente, David Vladeck levanta a possibilidade de aplicação da “*common enterprise liability*”, teoria segundo a qual cada entidade dentro de um conjunto de empresas

responsabilidade buscariam formas de limitar a capacidade dos consumidores e usuários de modificar, adaptar ou personalizar seus produtos, a fim de manter maior controle sobre como eles são utilizados. Isso também desaceleraria o processo de inovação proveniente de comunidades de hackers, código aberto e DIY. Estas, embora sejam uma fonte poderosa de inovação, teriam meios limitados para compensar indivíduos prejudicados por seus produtos, não apenas aqueles que optaram por fazer uso dos mesmos.

³⁹⁶ MULHOLLAND, Caitlin apud MAGRANI, et al. Op. cit., 121-122.

³⁹⁷ Id., p. 122.

interconectadas pode ser responsabilizada objetiva e solidariamente pelas ações de outras empresas integrantes do grupo. A aplicação da teoria, porém, exigiria algumas adaptações, como a não exigência de que as empresas funcionassem, necessariamente, em conjunto, mas tendo como requisito apenas que elas trabalhem rumo a um objetivo comum – no caso, projetar, programar e fabricar uma inteligência artificial autônoma e seus vários componentes.³⁹⁸

Além disso, a teoria da forma como é comumente aplicada serve para garantir que, uma vez estabelecida a responsabilidade, todos os transgressores sejam responsabilizados, porém não seria possível falar em uma transgressão, no sentido de culpabilidade, na hipótese tratada. Em vez disso, haveria uma inferência de responsabilidade operacionalizada normativamente para resguardar uma parte inocente (a vítima), fazendo com que outras pessoas suportem os custos do dano.

O ponto principal, inobstante, permaneceria o mesmo: a “*common enterprise liability*” permite que a lei imponha uma responsabilidade conjunta sem ter que lidar com os detalhes de analisar e atribuir todos os aspectos do ocorrido a uma parte envolvida ou outra, sendo suficiente que, no esforço para alcançar um objetivo comum, as partes tenham ocasionado um dano.

Tal princípio poderia, segundo o autor, ser incorporado em um novo regime de responsabilidade objetiva para lidar com os danos que podem ser incutidos aos seres humanos por agentes autônomos de IA, quando é impossível ou impraticável atribuir falhas a uma pessoa específica. Da maneira por ele vislumbrada, essa responsabilidade comum serviria como uma forma de seguro obrigatório imposto pelos tribunais: grupo (“*enterprise*”) indenizaria conjuntamente a eventuais vítimas na impossibilidade de se determinar, ou ainda menos partilhar falhas ocorridas.³⁹⁹

3. A contratação de seguros para danos causados por um agente autônomo de IA

Faz-se necessário, agora, tecer alguns comentários acerca da contratação de seguros para arcar com possíveis danos causados por um agente artificial autônomo. Conforme verificou-se no capítulo II desta pesquisa, a personificação de tais agentes como mecanismo jurídico para sanar o problema da lacuna de responsabilidade teria como requisito

³⁹⁸ VLADECK, David. Op. cit., p. 149.

³⁹⁹ Idem.

fundamental a possibilidade de que eles pudessem arcar financeiramente com qualquer dano que possam causar. Dessa maneira, a imputação de responsabilidade a uma IA demandaria que a mesma pudesse possuir bens, algo que poderia ser viabilizado através da contratação de seguros.

Similarmente, na hipótese em que a responsabilidade seja imputada aos desenvolvedores da inteligência artificial, estes poderiam também recorrer à aquisição de seguros, como forma de difusão da carga indenizatória do risco econômico.

O seguro socializa a responsabilidade pela reparabilidade dos danos, diluindo-a entre um determinado grupo (como as empresas envolvidas na criação e distribuição de uma IA) ou até mesmo pela sociedade, de modo a reduzir as situações não-indenizáveis.⁴⁰⁰

A socialização dos riscos, desse modo, apresenta-se como uma das tendências no campo da responsabilidade civil, conjuntamente à sua objetivação e à expansão dos danos indenizáveis.

O seguro constitui modalidade específica de contrato, na qual uma das partes (a seguradora) concorda em pagar certas quantias em dinheiro ou, mais raramente, em tomar medidas para compensar outra parte (a segurada), se certos eventos ocorrerem. Em troca, o segurado pagará determinada quantia (prêmio) em intervalos especificados, por exemplo, mensal ou anualmente.⁴⁰¹

Trata-se de uma forma de gerenciamento de riscos, com a assunção dos mesmos pela seguradora em troca de uma taxa. De maneira geral, os segurados pagam um prêmio relativamente baixo em comparação ao valor total a ser pago pela seguradora caso o evento segurado venha a ocorrer. Quanto menor for a probabilidade de ocorrência do evento, menor será a proporção entre o valor do prêmio e o pagamento a ser feito pela seguradora. Esta, por vez, beneficia-se porque, tendo como pressuposto que seus cálculos atuários estejam corretos, o valor líquido dos prêmios pagos excederá os montantes pagos em dinheiro aos segurados.⁴⁰²

A essência do direito securitário é lidar com situações de incerteza, razão pela qual a imprevisibilidade de uma inteligência artificial pode não ser um problema para as seguradoras. Ao repassar o custo dos danos às seguradoras por um preço fixo, as partes podem planejar riscos desconhecidos com maior certeza. Assim, o custo das apólices de seguro pode, portanto, ser inserido em previsões financeiras para os investidores e repassado ao usuário final no preço pago, espalhando a carga entre os participantes do mercado.

⁴⁰⁰ SOTO, Ricardo Alonso. Responsabilidad civil y seguro. In: PANTALEÓN, Fernando (Org.). *La Responsabilidad en el Derecho*. Anuario de la Facultad de Derecho de La Universidad Autónoma de Madrid. Madrid: Universidad Autónoma de Madrid, 2001, p. 193-204.

⁴⁰¹ TURNER, Jacob. Op. cit., p. 112.

⁴⁰² Id., p. 113.

Para Curtis Karnow, a melhor maneira de lidar com a responsabilidade pelos danos causados por uma IA seria a adoção de um plano de seguro:

Assim como as seguradoras examinam e certificam os candidatos a seguros de vida, seguro automobilísticos e similares, também os desenvolvedores que buscam cobertura para um agente [artificial] podem submetê-lo a um procedimento de certificação e, se bem-sucedido, receber a cotação de uma taxa dependente aos riscos prováveis representados pelo agente. Esse risco seria avaliado ao longo de um espectro de autonomia: quanto maior a inteligência, maior o risco e, portanto, maior o prêmio a ser pago e vice-versa (tradução nossa).⁴⁰³

Para a maioria das atividades e indústrias, as apólices de seguro são voluntárias, podendo haver lacunas quando uma parte não segurada causa um dano e, posteriormente, é incapaz de satisfazer os pedidos de indenização a ela direcionados. Existem, porém, algumas exceções notáveis, como o seguro obrigatório de veículos automotores (DPVAT)⁴⁰⁴ e o seguro de acidentes nucleares⁴⁰⁵.

No primeiro caso, a obrigação imposta pelo Estado justifica-se pelo elevado número de usuários de automóveis, pela frequência de acidentes e pelo intuito de garantir que as vítimas tenham uma recuperação rápida e certa, sobretudo quando a parte causadora do dano possui recursos financeiros escassos. Considerações semelhantes podem tornar desejável que a contratação de seguros para determinados agentes autônomos de IA seja obrigatória, ao menos para cobrir os riscos impostos a terceiros.

A pergunta que deve ser respondida no caso de seguros para agentes artificiais autônomos é quem deverá realizar a sua contratação. Uma opção é que os diversos envolvidos adquiram separadamente seus respectivos seguros. No caso de um veículo autônomo, por exemplo, o seguro pode ser contratado pela empresa que produziu o veículo, pela empresa que projetou o sistema de IA, assim como pelo proprietário do veículo.

Dessa forma, se o passageiro ou um terceiro sofrer um dano resultante do comportamento da IA, existirá ao menos alguma certeza para a vítima de que receberá a compensação por estes danos, e ainda mais certeza para os segurados que deverão arcar apenas com o valor de seu prêmio. No entanto, isso não impede que as diferentes seguradoras

⁴⁰³ KARNOW, Curtis. Liability for Distributed Artificial Intelligences. *Berkeley Technology Law Journal*, v. 11, n. 1, p. 147-204.

⁴⁰⁴ A Lei 6.194/74, com alterações conferidas pelas Leis 8.441/92 e 11.482/2007, dispõe sobre a formação de um fundo que visa a ofertar uma indenização mínima às vítimas de danos advindos de acidentes automobilísticos. O artigo 5º preceitua que “O pagamento de indenização será efetuado mediante simples prova do acidente e do dano decorrente, independentemente da existência de culpa, haja ou não resseguro, abolida qualquer franquia de responsabilidade do segurado”. (BRASIL, 1974).

⁴⁰⁵ Lei 6.453/1977, que dispõe sobre a responsabilidade civil sobre danos nucleares, preceitua em seu artigo 13 que “O operador da instalação nuclear é obrigado a manter seguro ou outra garantia financeira que cubra a sua responsabilidade pelas indenizações por danos nucleares” (BRASIL, 1977).

disputem entre si quando apenas uma delas for acionada para indenizar integralmente a vítima, buscando a contribuição das demais.⁴⁰⁶

Deve-se ressaltar, ainda, que a adoção de seguros não resolve, por si só, a questão da imputação de responsabilidade, pois serve apenas para redirecionar o dever de ressarcimento daquele que causou o dano para a seguradora. Se a vítima de um dano causado por uma IA não possuir o direito de acionar o segurado, a seguradora não teria motivos para providenciar qualquer pagamento. Significa dizer que, ainda que as diversas partes envolvidas contratem apólices de seguros, ainda deverá ser decidida como deverá ser a atribuição de responsabilidade.

O seguro normalmente tem um efeito de canalização no que diz respeito ao comportamento do segurado, porque a seguradora tem interesse em minimizar o risco de danos, passando, portanto, a exigir um certo comportamento das partes seguradas para que sua apólice permaneça válida. No que diz respeito à IA, isso pode significar que as seguradoras exigirão dos segurados o cumprimento de certos padrões mínimos em seu design e utilização.⁴⁰⁷

A título de exemplo, as seguradoras podem tentar se ilidir da responsabilidade quando uma IA realizar uma atividade fora do seu padrão de alcance, como um robô de entrega sendo utilizado como concierge. Quanto mais imprevisível for a IA segurada, mais difícil será para a seguradora avaliar e, finalmente, estabelecer um preço para a probabilidade de dano.⁴⁰⁸

O grau de imprevisibilidade de um agente artificial autônomo pode vir a tornar o valor dos prêmios proibitivamente caro. A recente experiência norte americana no que tange a seguros médicos demonstra que pode ser extremamente difícil para o governo compelir as seguradoras a adentrar mercados os quais elas não consideram economicamente viáveis.⁴⁰⁹

De fato, Asaro explicita que, para que o segure funcione, é necessário que os atuários possam avaliar os riscos apresentados pelos variados agentes artificiais. A avaliação desses riscos, porém, demandará conhecimento especializado na área e, provavelmente, meios para impor ou garantir a previsibilidade nos próprios sistemas. E ainda assim, poderão haver riscos catastróficos envolvidos, impassíveis de serem administrados dentro deste cenário.⁴¹⁰

⁴⁰⁶ TURNER, Jacob. Op. cit., p. 116.

⁴⁰⁷ Idem.

⁴⁰⁸ Id., p. 117.

⁴⁰⁹ KHAZAN, Olga. Why so many insurers are leaving Obamacare: How rejecting Medicaid and other government decisions have hurt insurance markets. *The Atlantic*, mai. 2017. Disponível em: <https://www.theatlantic.com/health/archive/2017/05/why-so-many-insurers-are-leaving-obamacare/526137>. Acesso em: 17 jun. 2020.

⁴¹⁰ ASARO, Peter. Op. cit., p. 193.

Similarmente, para Patrick Hubbard, a previsão de sinistros pode ser realizada até certo ponto quando um elevado número de dados sobre prévias ocorrências existir, como na área de seguro automobilístico. No entanto, esse tipo de informação pode não estar disponível para o seguro de responsabilidade civil para um tipo específico de produto – a exemplo da hipótese em que um desenvolvedor procura segurar uma sofisticada e inovadora inteligência artificial. Seguradoras de responsabilidade por produtos podem, portanto, ter receio quanto a reivindicações potencialmente altas, e a obtenção de um seguro pode ser difícil e/ou extremamente cara.⁴¹¹

Por fim, destaca-se que a atividade securitária também acaba suportando certos ônus que, invariavelmente, acabarão sendo repassados aos próprios segurados. Isso porque, como as seguradoras terão que arcar com o pagamento referente a novas apólices, relativas a atividades já existentes ou a novas modalidades de seguro⁴¹², para que consigam obter seus lucros, elas necessariamente irão transferir os custos financeiros decorrentes dessa majoração para a própria coletividade de segurados, mediante o aumento dos prêmios, por exemplo.⁴¹³

Via de regra, as empresas encontram-se na melhor posição para prevenir danos relativos aos seus produtos, e viu-se que a responsabilidade objetiva cria um incentivo maior para que elas o façam. Ainda, empresas detêm a prerrogativa de redistribuir a responsabilidade por meio de seguros⁴¹⁴ (algo que não impede que consumidores e usuários, caso assim desejem, também adquiram seguros próprias). Isso é especialmente verdadeiro quando se leva em conta as grandes empresas vanguardistas na pesquisa e desenvolvimento de tecnologias envolvendo inteligência artificial, como Google, Amazon e Facebook⁴¹⁵, e o fato de que os primeiros agentes artificiais verdadeiramente autônomos introduzidos no mercado serão fruto dessas empresas.

É precisamente essa posição vantajosa em que empresas se encontram para lidar com os riscos advindos de suas atividades que justifica a adoção pelos tribunais, inicialmente, de regimes de responsabilidade objetiva pelo fato do produto. Danos ocasionados por IAs, no entanto, apresentam a diferença de, uma vez que as máquinas (ou sistemas) se tornam mais

⁴¹¹ HUBBARD, F. Patrick. Allocating the risk of physical injury from “sophisticated robots”: Efficiency, fairness and innovation. In: CALO, Ryan; FROOMKIN, A. Michael; KERR, Ian (orgs.) *Robot Law*. Cheltenham: Edward Elgar Publishing, 2016, p. 25-50.

⁴¹² Haja vista a já mencionada tendência, no âmbito da responsabilidade civil extracontratual, de se objetivar a análise do dano.

⁴¹³ HIRONAKA, Giselda. *Responsabilidade Pressuposta*. Belo Horizonte: Del Rey, 2005.

⁴¹⁴ ABBOTT, Peter. Op. cit., p. 15.

⁴¹⁵ Baseado em dados referentes ao ano de 2017, Ryan Calo destaca que apenas sete instituições privadas possuem capacidades no campo da IA que ultrapassam largamente todas as demais. São elas: Google, Facebook, IBM, Amazon, Microsoft, Apple e Baidu. CALO, Ryan. *Artificial Intelligence Policy: A primer and roadmap*, 2017. Disponível em: <https://ssrn.com/abstract=3015350>. Acesso em: 30 mar. 2020.

seguras do que seus correspondentes humanos, a automação resultará em ganhos líquidos de segurança.⁴¹⁶

Em um primeiro momento de inserção no mercado de agentes artificiais autônomos, é inevitável que ao menos alguns danos ocorrerão, estando os desenvolvedores na melhor posição para com eles arcar. Mas, na medida em que a tecnologia se aperfeiçoa, o nível de segurança sobe e, conseqüentemente, cada vez menos será necessário recorrer a indenizações ou acionamento de seguros, de forma a não prejudicar processos inovadores.

Com efeito, Hubbard ressalta que os ordenamentos jurídicos têm balanceado inovação e segurança há séculos, e este equilíbrio se manteve ao se deparar com máquinas automáticas como elevadores, escadas rolantes e máquinas de venda no século XX. Litígios referentes a máquinas automáticas não impediu a continuidade de processos de inovação, e nem impediu os recentes avanços na robótica.⁴¹⁷

Em termos específicos, o autor destaca que veículos têm progressivamente se tornado mais robotizados, como sistemas como: controle de cruzeiro adaptativo, que não apenas mantém a velocidade do veículo, mas também a reduz se este estiver se aproximando demais de um objeto; sistemas de frenagem ABS; e controle eletrônico de estabilidade.

À medida que o processo de automação de veículos progrediu, resultou-se em ações consideráveis. Por exemplo, com alegações de design defeituoso dos sistemas de controle de cruzeiro adaptativo, ou recalls para solucionar problemas de software. Não obstante, os fabricantes de automóveis continuam investindo na sua computorização e buscando o desenvolvimento de veículos autônomos. Hubbard conclui, desse modo, que as preocupações com a responsabilidade por danos tiveram pouco impacto negativo sobre a inovação, ao menos nessa área.⁴¹⁸

Além disso, embora não faça parte do objetivo desta pesquisa, é interessante lembrar que os governos poderiam fornecer variedade de incentivos financeiros para fabricantes com o objetivo de promover a criação de tecnologias mais seguras, algo que tem sido eficaz no fomento à inovação. Por exemplo, os incentivos podem assumir a forma de investimentos em pesquisa e desenvolvimento, empréstimos para a construção de instalações de produção, direitos de propriedade intelectual, prêmios, tratamentos tributários preferenciais ou garantias governamentais.⁴¹⁹

⁴¹⁶ ABBOTT, Peter. Op. cit., p. 22. Note-se que o autor faz menção à definição de regimes de responsabilidade objetiva pelos tribunais, e não por mecanismos legislativos, referindo-se a sistemas de *common law*.

⁴¹⁷ HUBBARD, F. Patrick. Op. cit., p. 39.

⁴¹⁸ Id., p. 40.

⁴¹⁹ ABBOTT, Peter. Op. cit., p. 32.

O aumento nos níveis de segurança e diminuição de acidentes poderia também ser promovido através da regulamentação, como a determinação de requisitos para que a indústria alcancem metas básicas de segurança, ou a imposição direta de adoção de certas tecnologias.⁴²⁰

Finalmente, cumpre ressaltar que os regimes de responsabilidade civil extracontratual e as hipóteses de contratação de seguros analisadas neste capítulo representam apenas as possibilidades já existentes na grande maioria dos ordenamentos para lidar com o tema. Não significa dizer que, a depender da necessidade, novos modelos de responsabilidade possam ser criados. Pode ocorrer que diferentes sistemas jurídicos, buscando diferentes modos de balanceamento de todos os interesses envolvidos na criação, distribuição e utilização de agentes artificiais autônomos, a atribuindo pesos diversos a cada um desses interesses, optem pela adoção de um novo modelo de responsabilidade.

Por exemplo, Matthew Scherer defende a adoção de um regime em que seja promulgada uma Lei de Desenvolvimento de Inteligência Artificial (*“Artificial Intelligence Development Act”* ou “AIDA”). Esta criaria uma agência especializada responsável pela certificação da segurança de sistemas de inteligência artificial.⁴²¹

Além disso, tal instrumento normativo instituiria, um regime de responsabilidade civil segundo o qual designers, fabricantes e vendedores de programas de IA certificados pela AIDA estariam sujeitos à um esquema de responsabilidade subjetiva. Por outro lado, aqueles que disponibilizassem programas de IA para venda ou uso comercial sem a certificação pela agência estariam sujeitos à responsabilidade objetiva e também solidária, sendo possível a potenciais vítimas reclamar os danos de qualquer entidade envolvida na cadeia de desenvolvimento, distribuição ou venda.⁴²²

⁴²⁰ Id., p. 33.

⁴²¹ SCHERER, Matthew. Op. cit., p. 393.

⁴²² Idem.

CONSIDERAÇÕES FINAIS

Partindo do pressuposto da iminente entrada no mercado de consumo de agentes artificiais autônomos, o presente trabalho teve como objetivo precípuo investigar se a eles poderia ser atribuída personalidade jurídica, como forma de se solucionar o problema da lacuna de responsabilidade existente em hipóteses de danos não passíveis de serem diretamente rastreáveis ao comportamento de um dos envolvidos em sua cadeia de produção, distribuição, ou utilização.

Verificou-se que esse tipo de agente, por ser capaz de atuar de maneira independente de instrução ou supervisão humana direta, coletando e analisando informações do seu ambiente para, a partir delas, adaptar-se a novas situações, pode proceder de maneira imprevisível. Soma-se a esta característica o fato de que, devido ao uso de técnicas de aprendizado de máquina, muitas das vezes os próprios programadores e responsáveis pelo processo de treinamento do agente não são capazes de fornecer explicações acerca de como determinado resultado foi por ele alcançado.

Desse modo, eventual dano causado por um agente artificial autônomo estará diretamente ligado à sua conduta, mas não necessariamente à conduta culposa do programador, operador ou usuário. A proposta de personificar esses agentes, portanto, vem adquirindo força em debates sobre a regulamentação da inteligência artificial.

Para analisar a sua viabilidade, foram avaliadas as razões que justificam a atribuição de personalidade a outras entidades não humanas, com foco especial na figura de entes coletivos personificados, como as corporações. Esta avaliação revelou que a decisão de se criar um novo ator legal depende, eminentemente, de uma escolha do ordenamento jurídico, frequentemente baseada nos objetivos almejados pelo mesmo com a personificação, e levando em conta fatores como a capacidade social para ação coletiva do ente.

A personalidade jurídica – abordada a partir de um ponto de vista legalista, precisamente por não estar necessariamente vinculada a uma ontologia moral ou a noções metafísicas atreladas ao ente personificado, exerce, eminentemente, uma função heurística ao simplificar um complexo de relações jurídicas em um único centro de imputação de direitos e obrigações. Dentro desse contexto, conclui-se, portanto, ser possível a atribuição de personalidade a um agente artificial autônomo, com a devida modulação da sua capacidade de direito para que se alcance a finalidade almejada pela personificação.

A análise do conteúdo da capacidade de direito desses agentes, caso sejam personificados, demonstrou-se fundamental para verificar, para além da atribuição de personalidade, se essa proposta mudança de status seria também adequada, levando-se em consideração, principalmente, as suas implicações para os entes já personificados dentro de determinado ordenamento jurídico.

Dentro deste enquadramento, viu-se que direitos principalmente de cunho patrimonial poderiam ser, sem maiores debates, concedidos ao novo ente personificado, justamente para viabilizar a articulação patrimonial e a constituição de um sistema de imputação direta. Ainda, a possibilidade de contratar e atuar em juízo, ainda que mediante representação, seriam necessárias ao funcionamento autônomo do agente dentro do sistema legal.

Outros direitos, entretanto, apresentariam não apenas maior resistência mas também riscos às pessoas humanas com as quais os agentes artificiais interagem. Diante da tendência cada vez mais presente de se estender às pessoas jurídicas direitos e garantias fundamentais, foi possível concluir que o mesmo pode acontecer na hipótese de personificação desses agentes, verificando-se o processo de expropriação da subjetividade.

Ademais, foram apresentados outros empecilhos à atribuição de personalidade, como os custos associados à criação de um sistema de registro, a dificuldade de se identificar com precisão um determinado agente para promover esse registro, e a possibilidade de utilização indevida do instituto jurídico da personalidade como instrumento de alcance da irresponsabilidade organizada das pessoas envolvidas no seu processo de desenvolvimento e na sua utilização.

Desse modo, foi possível concluir que, embora possível, a personificação de agentes autônomos de inteligência artificial não é, ao menos em um futuro próximo, desejável. Ao analisar os regimes de responsabilidade civil extracontratual existentes e passíveis de serem aplicados como forma de não deixar potenciais vítimas de danos ocasionados por esses agentes sem reparação, verificou-se a existência de alternativas viáveis, que não demandam a interposição de um novo sujeito para sanar a ausência de agência legal correspondente à agência causal exercida por esses agentes.

Dentre elas, a adoção de um regime de responsabilidade objetiva parece ser a mais adequada, pois: permite a compensação pelos danos sofridos por indivíduos que não influenciaram de qualquer maneira na sua ocorrência; os desenvolvedores da tecnologia estão em posição mais adequada para prevenir a ocorrência de danos e redistribuir os custos

incorridos; fornece estabilidade aos desenvolvedores e incentivo à criação de produtos mais seguros; e, finalmente, evitaria a ocorrência do fenômeno da irresponsabilidade distribuída.

A segunda alternativa avaliada disse respeito à contratação de seguros para arcar com eventuais danos provocados por um agente artificial autônomo, seja pelos seus desenvolvedores, seus usuários ou proprietários. Contudo, a implementação de seguros pode ser dificultada pelo caráter extremamente inovador dessa tecnologia, e a ausência de informações suficientes sobre o grau de ocorrência de danos para que se calcule, de maneira fidedigna, o valor dos prêmios. Isso contribui para que, ao menos inicialmente, a contratação de seguros seja proibitivamente cara, além de se retornar à questão inicialmente colocada acerca de quem deve arcar com os danos ocorridos.

A inteligência artificial trará consigo, inexoravelmente, uma revolução no campo jurídico. Muitas atividades legais já vem sendo, e assim continuarão, profundamente transformadas por ela, e conceitos legais basilares como o de personalidade, agência e responsabilidade, discutidos neste trabalho, deverão ser repensados e readequados à nova realidade resultante desta tecnologia.

Discussões relativas ao tema são ainda incipientes e controversas. Longe de oferecer respostas definitivas para os problemas colocados pela inteligência artificial, buscou-se por meio desta pesquisa contribuir e avançar o debate. Através de uma revisão abrangente sobre o impacto do surgimento de agentes artificiais autônomos na sociedade, espera-se que este trabalho possa gerar os subsídios teóricos necessários para o aprofundamento do tópico em questão e para a fundamentação de futuras propostas normativas nesta área.

REFERÊNCIAS

- ALLEN, Tom; WIDDISON, Robin. Can Computers Make Contracts. *Harv. JL & Tech.*, v. 9, 1996, p. 25-52.
- ALLGROVE, Benjamin D. *Legal Personality for Artificial Intellects*: Pragmatic solution or science fiction? 2004, p. 1-101. Disponível em: <https://ssrn.com/abstract=926015>. Acesso em: 12 jan. 2020.
- ANDERSON, Kenneth; WAXMAN, Matthew C. Debating Autonomous Weapon Systems, Their Ethics and Their Regulation Under International Law. *Washington College of Law Research Paper No. 2017-21*, p. 1097- 1117, 2017.
- ANDRADE, Francisco et al. Contracting Agents: Legal personality and representation. *Artificial Intelligence and Law*, v. 15, n. 4, p. 357-373, 2007.
- ANGWIN, Julia; LARSON, Jeff; MATTU, Surya; KIRCHNE, Lauren. Machine Bias. *ProPublica*, mai. 2018. Disponível em: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>. Acesso em: 26 ago. 2019.
- ARMSTRONG, Stuart; SOTALA, Kaj. How We're Predicting AI – or Failing To. In: ROMPORTL, Jan; ZACKOVA, Eva; IRCING, Pavel; POLAK, Michal; SCHUSTER, Radek (Eds.). *Beyond AI: Artificial Dreams*. Pilsen: University of West Bohemia, 2012, p. 52–75.
- ASARO, Peter M. *The Liability Problem for Autonomous Artificial Agents*. AAAI Spring Symposium Series, 2016. Disponível em: <https://www.aaai.org/ocs/index.php/SSS/SSS16/paper/view/12699>. Acesso em: 09 abr. 2020.
- BALKIN, Jack M. The Three Laws of Robotics in the Age of Big Data. *Ohio State Law Journal*, v. 78, p.1217-1241, 2017.
- BECK, Ulrich. *Sociedade de Risco*: Rumo a uma outra modernidade. [Trad.] Sebastião Nascimento. São Paulo: Editora 34, 2011.
- BEYMAN, Michael. Big Data's powerful effect on tiny babies. *CNBN*, set. 2013. Disponível em: <https://www.cnn.com/2013/09/13/big-datas-powerful-effect-on-tiny-babies.html>. Acesso em: 30 out. 2019.
- BODEN, Margaret. *AI: Its nature and future*. Oxford: Oxford University Press, 2016.
- BOSTROM, Nick. *Superinteligência*: Caminhos, perigos e estratégias para um novo mundo. Rio de Janeiro: DarkSide, 2014.
- BORKING, John; B. M. A. van Eck; SIEPEL, P. Intelligent software agents: Turning a privacy threat into a privacy protector. *Information and Privacy Commissioner Ontario (Canada)*, 1999. Disponível em: https://www.researchgate.net/publication/237219149_INTELLIGENT_SOFTWARE_AGEN TS_AND_PRIVACY. Acesso em: 30 out. 2019.

BRASIL. Lei. n. 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 26/08/2019.

BRASIL. Lei n. 8.078, de 11 de setembro de 1990. Código de Defesa do Consumidor. Brasília, DF: Presidência da República. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l8078.htm. Acesso em 20 mar. 2020.

BRASIL. Lei n. 10.046 de 10 de janeiro de 2002. Código Civil. Brasília, DF: Presidência da República. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/2002/l10406.htm. Acesso em 20 mar. 2020.

BRASIL. Lei n. 6.194 de 19 de dezembro de 1974. Dispõe sobre Seguro Obrigatório de Danos Pessoais causados por veículos automotores via terrestre, ou por sua carga, a pessoas transportadas ou não. Brasília, DF: Presidência da República. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l6194.htm. Acesso em: 21 jun. 2020.

BRASIL. Lei n. 6.453 de 17 de outubro de 1977. Dispõe sobre a responsabilidade civil por danos nucleares e a responsabilidade criminal por atos relacionados com atividades nucleares e dá outras providências. Brasília, DF: Presidência da República. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/L6453.htm. Acesso em: 21 jun. 2020.

BROUSSARD, Meredith. *Artificial Intelligence: How computers misunderstand the world*. Cambridge: MIT Press, 2018.

BRUNO, Lee. How Swarm Intelligence Helps AI Make Better Decisions. *All Turtles*, 2018. Disponível em: <https://www.all-turtles.com/2018/11/08/how-swarm-intelligence-helps-ai-make-better-decisions/>. Acesso em: 31 mar. 2020.

BRYSON, Joanna J.; DIAMANTIS, Mihailis E.; GRANT, Thomas D. Of, for, and by the people: the legal lacuna of synthetic persons: *Artificial Intelligence Law*, n. 25, 2017, p. 237-291.

BURRI, Thomas. Free Movement of Algorithms: Artificially Intelligent Persons Conquer the European Union's Internal Market. In: PAGALLO, Ugo; BARFIEL, Woodrow (orgs.) *Research Handbook on the Law of Artificial Intelligence*. Cheltenham: Edward Elgar, 2018.

CALO, Ryan. *Artificial Intelligence Policy: A primer and roadmap*, 2017. Disponível em: <https://ssrn.com/abstract=3015350>. Acesso em: 30 mar. 2020.

CALO, Ryan. Robotics and the Lessons of Cyberlaw. *California Law Review*, v. 103, n. 3, p. 513-563, 2015.

CARTY, Sharon. Toyota's Sudden Acceleration Problem May Have Been Triggered By Tin Whiskers. *Huffpost*, jan. 2012. Disponível em: https://www.huffpostbrasil.com/entry/toyota-sudden-acceleration-tin-whiskers_n_1221076?ri18n=true. Acesso em: 20 jun. 2020.

ČERKA, Paulius; GRIGIENĖ, Jurgita; SIRBIKYTĖ, Gintarė. Liability for damages caused by artificial intelligence. *Computer Law & Security Review*, v. 31, n. 3, 2015, p. 376-389.

CHINA. *Governance Principles for a New Generation of Artificial Intelligence: Develop responsible artificial intelligence*. 2017.

CHOPRA, Samir; LAWRENCE, White. *A Legal Theory for Artificial Autonomous Agents*. Ann Arbor: University of Michigan Press, 2011.

CHOPRA, Samir; WHITE, Laurence. Artificial Agents-personhood in Law and Philosophy. In: *Proceedings of the 16th European Conference on Artificial Intelligence*. IOS Press, 2004, p. 635-639.

COECKELBERGH, Mark. Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology*, v. 12, n. 3, p. 209-221, 2010.

CONNER-SIMONS, Adam; GORDON, Rachel. Using AI to predict breast cancer and personalize care. *MIT News*, mai. 2019. Disponível em: <http://news.mit.edu/2019/using-ai-predict-breast-cancer-and-personalize-care-0507>. Acesso em: 30 out. 2019.

DARLING, Kate. *Extending Legal Rights to Social Robots*. We Robot Conference, Miami, 2012, p. 1-18. Disponível em: http://robots.law.miami.edu/wp-content/uploads/2012/04/Darling_Extending-Legal-Rights-to-Social-Robots-v2.pdf. Acesso em: 28 mar. 2020.

DA SILVA, Nilton Correia. Inteligência Artificial. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 35-52.

DAVIES, Alex. Tesla's Latest Autopilot Death Looks Just a Prior Crash. *WIRED*, mai. 2019. Disponível em: <https://www.wired.com/story/teslas-latest-autopilot-death-looks-like-prior-crash/>. Acesso em: 30 out. 2019.

DEMOTT, Deborah A. Disloyal agents. *Alabama Law Review*, v. 58, p. 1049–1067, 2007.

DIOP, Lamine; CUPE, Jean. Explainable AI: The data scientists' new challenge. *Towards Data Science*, jun. 2018. Disponível em: <https://towardsdatascience.com/explainable-ai-the-data-scientists-new-challenge-f7cac935a5b4>. Acesso em: 30 out. 2019.

DUTTON, Tim. *An Overview of National AI Strategies*. Disponível em: <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>. Acesso em: 29 out. 2019.

DYE, Jessica. Toyota Acceleration Case Settlement Gets Final OK. *Insurance Journal*, jul. 2013. Disponível em: <https://www.insurancejournal.com/news/national/2013/07/22/299154.htm>. Acesso em: 20 jun. 2020.

EDWARDS, Lilian; VEALE, Michael. Enslaving the Algorithm: From a 'Right to an Explanation' to a 'Right to Better Decisions'? *IEEE Security & Privacy*, v. 16, n. 3, p. 46-54, 2018. Disponível em: <https://ssrn.com/abstract=3052831>. Acesso em: 26 ago. 2019.

EDWARDS, Lilian; VEALE, Michael. Slave to the Algorithm? Why a ‘right to an explanation’ is probably not the remedy you are looking for. *Duke Law & Technology Review*, v. 16, n. 01, p. 18-84, 2016. Disponível em: <https://ssrn.com/abstract=2972855>. Acesso em: 26 ago. 2019.

ESTADOS UNIDOS. *Executive Order on Maintaining American Leadership in Artificial Intelligence*, 2019. Disponível em: <https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/>. Acesso em: 25 nov. 2019.

ESTADOS UNIDOS. *House Resolution 2231: Algorithmic Accountability Act*. Disponível em: <https://www.congress.gov/bill/116th-congress/house-bill/2231/text>. Acesso em: 25 nov. 2019.

ESTADOS UNIDOS. *House Resolution 153: Supporting the development of guidelines for ethical development of artificial intelligence*. Disponível em: <https://www.congress.gov/bill/116th-congress/house-resolution/153/text>. Acesso em: 25 nov. 2019.

ESTADOS UNIDOS. *House Resolution 4625: Future of Artificial Intelligence Act*, 2017. Disponível em: <https://www.congress.gov/bill/115th-congress/house-bill/4625?q=%7B%22search%22%3A%5B%22the%22%5D%7D>. Acesso em: 25 nov. 2019.

ESTADOS UNIDOS. *Restatement (Third) of Agency*. 1933.

ESTADOS UNIDOS. National Transportation Safety Board: Public Meeting of November 19, 2019, Collision Between Vehicle Controlled by Developmental Automated Driving System and Pedestrian, p. 1-5, 2019. Disponível em: <https://www.nts.gov/news/events/Documents/2019-HWY18MH010-BMG-abstract.pdf>. Acesso em: 30 nov. 2019.

ESTADOS UNIDOS. Supreme Court. *Burwell v. Hobby Lobby*, 573, 2014. Disponível em: <https://www.oyez.org/cases/2013/13-354>. Acesso em: 01 abr. 2020.

ESTADOS UNIDOS. Supreme Court. *Dow Chemical Company v. United States*, 1986. Disponível em: <https://www.oyez.org/cases/1985/84-1259>. Acesso em 01 abr. 2020.

ESTADOS UNIDOS. Supreme Court. *First National Bank of Boston v. Bellotti*, 1978. Disponível em: <https://www.oyez.org/cases/1977/76-1172>. Acesso em 01 abr. 2020.

ESTADOS UNIDOS. Supreme Court. *Pacific Gas & Electric Company v. Public Utilities Commission of California*, 1986. Disponível em: <https://www.oyez.org/cases/1985/84-1044>. Acesso em 01 abr. 2020.

ESTADOS UNIDOS. Supreme Court. *Trustees of Dartmouth College v. Woodward*, 1819. Disponível em: www.oyez.org/cases/1789-1850/17us518. Acesso em 01 abr. 2020.

ESTADOS UNIDOS. Supreme Court. *United States v. Martin Linen Supply Company*, 1977. Disponível em: <https://www.oyez.org/cases/1976/76-120>. Acesso em 01 abr. 2020.

ESTADOS UNIDOS. Court of Appeals for the Ninth Circuit. *Stormans, Inc. v. Selecky*, 2009. Disponível em: <https://cite.case.law/f3d/586/1109/>. Acesso em: 01 abr. 2020.

ESTADOS UNIDOS. *Patient Protection and Affordable Care Act* (Public Law 111-148), 2010.

FLORIDI, Luciano; CATH, Carine; MITTELSTADT, Brent; WACHTER, Sandra; TADDEO, Mariarosaria. Artificial Intelligence and the ‘Good Society’: the US, EU, and UK approach. *Science and Engineering Ethics* (Springer), p. 1-24, 2017. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2906249. Acesso em: 29 out. 2019.

FORSTER, Victoria. AI Designs Radiation Therapy Treatment Plan For Cancer In Twenty Minutes. *Forbes*, ago. 2019. Disponível em: <https://www.forbes.com/sites/victoriaforster/2018/08/01/ai-can-design-radiation-therapy-treatment-plan-for-cancer-in-twenty-minutes/#2f13b8497a8c>. Acesso em: 30 out. 2019.

FRAZÃO, Ana. A Responsabilidade Civil dos Administradores de Sociedades Empresárias por Decisões Tomadas com Base em Sistemas de Inteligência Artificial In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 481-521.

GALGANO, Francesco. *Lex Mercatoria*. Bologna: Il Mulino, 2001.

GOODMAN, Bryce; FLAXMAN, Seth. EU Regulations on Algorithmic Decision-Making and a “Right to Explanation”. *AI Magazine*, v. 38, n. 03, p. 1-10, 2016. Disponível em: <https://arxiv.org/abs/1606.08813>. Acesso em: 30 out. 2019.

GRAY, Darren. No one behind the wheel: The new workforce driving Australia's mines. *The Sydney Morning Herald*, abr. 2019. Disponível em: <https://www.smh.com.au/business/companies/no-one-behind-the-wheel-the-new-workforce-driving-australia-s-mines-20190411-p51dd2.html>. Acesso em: 30 out. 2019.

GREENE, Natalia. The First Successful Case of the Rights of Nature Implementation in Ecuador. *The Rights of Nature*, 2011. Disponível em: <http://therightsofnature.org/first-rn-case-ecuador/>. Acesso em: 01 abr. 2020.

GUTIERREZ, Andrei. É Possível Confiar em um Sistema de Inteligência Artificial? Práticas em torno da melhoria da sua confiança, segurança e evidências de accountability. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 83-98.

HARARI, Yuval Noah. *21 Lições para o Século 21*. [Trad.] Paulo Geiger. São Paulo: Companhia das Letras, 2018.

HARARI, Yuval Noah. *Sapiens: Uma breve história da humanidade*. São Paulo: L&PM, 2015.

HART, H. L. A. Are There Any Natural Rights?, *The Philosophical Review*, v. 64, n. 2, 1955, p. 175-191.

HILBERT, Martin; LOPES, Priscila. How to Measure the World's Technological Capacity to Communicate, Store and Compute Information. *International Journal of Communications*, v. 6, p. 956-979, 2012. Disponível em: <https://ijoc.org/index.php/ijoc/article/viewFile/1562/742&from=32207443>. Acesso em: 30 out. 2019.

HIRONAKA, Giselda. *Responsabilidade Pressuposta*. Belo Horizonte: Del Rey, 2005.

HUBBARD, F. Patrick. Allocating the risk of physical injury from “sophisticated robots”: Efficiency, fairness and innovation. In: In: CALO, Ryan; FROOMKIN, A. Michael; KERR, Ian (orgs.) *Robot Law*. Cheltenham: Edward Elgar Publishing, 2016, p. 25-50.

HUBER, Peter. Safety and the Second Best: The hazards of public risk management in the courts. *Columbia Law Review*, v. 85, n. 2, 1985, p. 277-337.

INDIA. Bombay High Court. *Pramatha Nath Mullick v. Pradyumnakumar Mullick*, 1925. Disponível em: <https://indiankanoon.org/doc/290902/>. Acesso em: 01 abr. 2020.

KAHNEMAN, Daniel. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2013..

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

KAPLAN, Jerry. *Humans Need Not Apply: A guide to wealth and work in the age of artificial intelligence*. New Haven: Yale University Press, 2015, p. 61-63.

KARNOW, Curtis. Liability for Distributed Artificial Intelligences. *Berkeley Technology Law Journal*, v. 11, n. 1, p. 147-204.

KENDRICK, Molly. The border guards you can't win over with a smile. *BBC Future*, abr. 2019. Disponível em: <https://www.bbc.com/future/article/20190416-the-ai-border-guards-you-cant-reason-with>. Acesso em: 30 out. 2019.

KERR, Ian. Providing for Autonomous Electronic Devices in the Electronic Commerce Act. In: ULCC Uniform Law Conference of Canada, 1999. Disponível em: <https://www.ulcc.ca/en/1999-winnipeg-mb/359-civil-section-documents/362-providing-for-autonomous-electronic-devices-in-the-electronic-commerce-act-1999>. Acesso em: 28 mar. 2020.

KURKI, Visa A. J. *A Theory of Legal Personhood*. Oxford: Oxford University Press, 2019.

KURZWEIL, Ray. *The Singularity is Near: When humans transcend biology*. Nova York: Penguin Books, 2005.

LEE, Dave. Google self-driving car hits a bus. *BBC News*, fev. 2016. Disponível em: <http://www.bbc.com/news/technology-35692845>. Acesso em: 31 out. 2019.

LEITE, Renato. Existe um direito à explicação na Lei Geral de Proteção de Dados Pessoais do Brasil? Artigo Estratégico 24. Rio de Janeiro: *Instituto Igarapé*, dezembro de 2018.

Disponível em: <https://igarape.org.br/wp-content/uploads/2018/12/Existe-um-direito-a-explicacao-na-Lei-Geral-de-Protecao-de-Dados-no-Brasil.pdf>. Acesso em: 31 out. 2019.

LESTER, Caroline. What Happens When Your Bomb-Defusing Robot Becomes a Weapon. *The Atlantic*, abr. 2018. Disponível em: <https://www.theatlantic.com/technology/archive/2018/04/what-happens-when-your-bomb-defusing-robot-becomes-a-weapon/558758/>. Acesso em: 30 out. 2019.

LEVIN, Sam; WONG, Julia Carrie. Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian. *The Guardian*, mar. 2018. Disponível em: <https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe>. Acesso em: 30 out. 2019.

LOUKIDES, Mike; LORICA, Ben. *What is Artificial Intelligence?* Sebastopol: O'Reilly Media, Inc, 2016.

MAES, Pattie. Artificial life meets entertainment: Lifelike autonomous agents. *Communications of the ACM* 38 (11), 1995. Disponível em: <https://dl.acm.org/citation.cfm?id=219808>. Acesso em: 30 out. 2019.

MAGRANI, Eduardo; SILVA, Priscilla; VIOLA, Rafael. Novas Perspectivas sobre Ética e Responsabilidade de Inteligência Artificial. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 117-147.

MATTHIAS, Andreas. The responsibility gap: ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, v. 6, n. 3, p. 175-183, 2004.

MAYER-SCHÖNBERGER, Viktor; CUKIER, Kenneth. *Big Data: A revolution that will transform how we live, work and think*. New York: Eamon Dolan/Mariner Books, 2014.

MAYER-SCHÖNBERGER, Viktor; PADOVA, Yann. Regime Change? Enabling big data through Europe's new data protection regulation. *The Columbia Science and Technology Law Review*, v. 17, p. 315-335, 2017.

MARSHALL, Aarian. Starsky Robotics Unleashes Its Truly Driverless Trucks in Florida. *WIRED*, set. 2018. Disponível em: <https://www.wired.com/story/starsky-robotics-truck-self-driving-florida-test/>. Acesso em: 30 out. 2019.

MAYER, Carl J. Personalizing the Impersonal: Corporations and the Bill of Rights. *Hastings Law Journal*, v. 41, n. 3, 1990, p. 577-667.

MCARTHY, John; MINSKY, Leon; ROCHESTER, Nathanael; SHANNON, Claude. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence. 1955, p. 1-13. Disponível em: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. Acesso em 29 out. 2019.

MICHALCZAK, Rafal. Animal's Race Against the Machines. In: KURKI, Visa A. J.; PIETRZYKOWSKI, Tomasz (Eds.). *Legal Personhood: Animals, artificial intelligence and the unborn*. Cham (ZG): Springer, p. 91-101, 2017.

MINSKY, Marvin. *Computation: Finite and infinite machines*. Englewood Cliffs: Prentice-Hall, 1967.

MITTELSTADT, Brent; FLORIDI, Luciano; ALLO, Patrick; TADDEO, Mariarosaria; WACHTER, Sandra. The ethics of algorithms: Mapping the debate. *Big Data & Society*, v. 3(2), p. 1-21, 2016.

MITTELSTADT, Brent; WACHTER, Sandra; RUSSELL, Chris. Explaining Explanations in AI. *Proceedings of FAT* '19: Conference on Fairness, Accountability, and Transparency*, 2019. Disponível em: <https://arxiv.org/pdf/1811.01439.pdf>. Acesso em: 31 out. 2019.

MOOR, James H. Are there decisions computers should never make? *Nature & System*, v. 1, n. 4, p. 217-229.

MORAES, Maria Celina Bodin. Danos à Pessoa Humana: Uma leitura civil constitucional dos danos morais. Rio de Janeiro: Editora Processo, 2000.

MULHOLLAND, Caitlin. Responsabilidade Civil e Processos Decisórios Autônomos em Sistemas de Inteligência Artificial (IA): Autonomia, imputabilidade e responsabilidade. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 325-348.

MYERS, Andrew. Stanford team trains computer to evaluate breast cancer. *Stanford Medicine News Center*, nov. 2011. Disponível em: <https://med.stanford.edu/news/all-news/2011/11/stanford-team-trains-computer-to-evaluate-breast-cancer.html>. Acesso em: 29 out. 2019.

NAÇÕES UNIDAS. UNESCO's Report of COMEST on robotics ethics, 2017, p. 42. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000253952>. Acesso em: 31 out. 2019.

NAÇÕES UNIDAS. Convenção sobre a Utilização de Comunicações Eletrônicas em Contratos Internacionais, de 13 novembro de 2005. [Trad.] Bernardo Santos; Pedro Martini; Pedro Soares. Disponível em: http://www.cisg-brasil.net/doc/Traducao_convencao_comunicacoes_eletronicas.pdf. Acesso em: 01 dez. 2019.

NEGNEVITSKY, Michael. *Artificial Intelligence: A guide to intelligent systems*, 2 Ed. Harlow: Addison-Wesley, 2004.

NEGRI, Sérgio Ávila. As Razões da Pessoa Jurídica e a Expropriação da Subjetividade. *Civilistica*, a. 5, n. 2, 2016, p. 1-18.

NICOLELIS, Miguel A. L; CICUREL, Ronald. *O Cérebro Relativístico: Como ele funciona e por que ele não pode ser simulado por uma máquina de Turing*. São Paulo: Kios Press, 2015.

NOLAN, Donal. Causation and the Goals of Tort Law. In: ROBERTSON, Andrew; TANG, Hang Wu (eds.). *The Goals of Private Law* Donal Nolan, “Causation and the Goals of Tort Law”, in *The Goals of Private Law*. Oxford: Hart Publishing, 2009, p. 165-190.

NORONHA, Fernando. Responsabilidade Civil: Uma tentativa de ressystematização. *Revista de Direito Civil, Imobiliário, Agrário e Empresarial*. São Paulo, ano 17, n. 64, abr./jun. 1993, p. 7-47.

OCDE. *Recommendation of the Council on Artificial Intelligence*, 2019. Disponível em: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>. Acesso em: 25 nov. 2019.

PAGALLO, Ugo. *The laws of robots: crimes, contracts, and torts*. Heidelberg: Springer, 2013.

PANETTA, Kasey. Why Humans Are the Problem with Autonomous Cars, *EE World Online*, set. 2015. Disponível em: <https://www.ecnmag.com/blog/2015/09/why-humans-are-problem-autonomous-cars>. Acesso em: 19 jun. 2020.

PASQUALE, Frank. *The Black Box Society: The secret algorithms that control money and information*. Cambridge: Harvard University Press, 2015.

PEARL, Judea; MACKENZIE, Dana. *The Book of Why: The new science of cause and effect*. New York: Basic Books, 1. Ed., 2018.

PERLINGIERI, Carolina. *Enti e diritti della persona*. Napoli: Edizioni Scientifiche Italiane, 2009.

PERRY, Walter L; MCINNIS, Brian; PRICE, Carter C.; SMITH, Susan C.; HOLLYWOOD, John S. Predictive Policing: The role of crime forecasting in law enforcement operations. RAND, 2013. Disponível em: https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR233/RAND_RR233.pdf. Acesso em: 26 ago. 2019.

POLIDO, Fabrício. Novas Perspectivas para Regulação da Inteligência Artificial: Diálogos entre as políticas domésticas e os processos legais transnacionais. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 179-206.

POOLE, David; MACKWORTH, Alan. *Artificial Intelligence: Foundations of Computational Agents*. Cambridge (UK): Cambridge University Press, 2010.

POSNER, Richard. *Economic analysis of law*. New York: Wolters Kluwer Law and Business, 2014.

REINO UNIDO. *AI in the UK: Ready? Willing and Able?* 2017. Disponível em: <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf>. Acesso em: 25 nov. 2019.

REINO UNIDO. Court of Appeals. Bumper Development Corporation Ltd v. Commissioner of Police of Metropolis, 1991. Disponível em: <https://lawin.org/bumper-development-corp-ltd-v-commissioner-of-police-of-the-metropolis-and-others/>. Acesso em: 01 abr. 2020.

RICHARDS, Neil M.; SMART, William D. How should the law think about robots? In: CALO, Ryan; FROOMKIN, A. Michael; KERR, Ian (orgs.) *Robot Law*. Cheltenham: Edward Elgar Publishing, 2016.

RODOTÀ, Stefano. A antropologia do homo dignus. Trad. Maria Celina Bodin de Moraes. *Civilistica*, v. 6, n. 2, 2017, p. 1-17. Disponível em: <http://civilistica.com/aantropologia-do-homo-dignus/>. Acesso em: 29 mar. 2020.

RODOTÀ, Stefano. *La vita e le regole. Tra diritto e non diritto*. Milano: Fetrinelli Editore, 2007.

RODRIGUES JR, Otavio Luiz. Pessoa, personalidade, conceito filosófico e conceito jurídico de pessoa: Espécies de pessoas no direito em geral. *Revista de direito do consumidor*, v. 27, n. 118, 2018, p. 281-291.

RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 3. ed. New Jersey: Prentice-Hall, 2010.

RUTKIN, Aviva. Autonomous Cars Are Learning Our Unpredictable Driving Habits, *New Scientist*, ago. 2015. Disponível em: <https://www.newscientist.com/article/mg22730362-900autonomous-cars-are-learning-our-unpredictable-driving-habits>. Acesso em: 19 jun. 2020.

SAMUEL, Arthur. L. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, v. 3, n. 3, p. 210-229, 1959.

SAVARY, Jacques. *Le parfait négociant*. Paris: Librairie Droz, 2011.

SCHARTZ, Jacob T. Limits of Artificial Intelligence. In: SHAPIRO, Stuart; ECKROTH, David. *Encyclopedia of Artificial Intelligence*. New York: Wiley, 1987, p. 488-503.

SHAVELL, Steven. *Foundations of Economic Analysis of Law*. Cambridge: Harvard University Press, 2004.

SCHERER, Matthew U. Regulating Artificial Intelligence Systems: Risks, challenges, competencies, and strategies. *Harvard Journal of Law and Technology*, v. 29, n. 2, p. 353-400, 2016. Disponível em: <http://jolt.law.harvard.edu/articles/pdf/v29/29HarvJLTech353.pdf>. Acesso em: 29 out. 2019.

SCHMELZER, Ron. Understanding Explainable AI. *Forbes*, jul. 2019. Disponível em: <https://www.forbes.com/sites/cognitiveworld/2019/07/23/understanding-explainable-ai/#73fc6b337c9e>. Acesso em: 30 out. 2019.

SCHNEIDER, Kevin. Judge Recognizes two chimpanzees as legal persons, grants them writ of habeas corpus. *Non Human Rights*, 2015. Disponível em: <https://www.nonhumanrights.org/blog/judge-recognizes-two-chimpanzees-as-legal-persons-grants-them-writ-of-habeas-corpus/>. Acesso em: 01 abr. 2020.

SCHWAB, Klaus. *The Fourth Industrial Revolution*. Danvers: Currency, 2017.

SEARLE, John. *Mente, Cérebro e Ciência*. Tradução de Artur Morão. Lisboa: Biblioteca de Filosofia Contemporânea, 2000.

SELBST, Andrew; POWLES, Julia. Meaningful Information and the Right to Explanation. *International Data Privacy Law*, v. 7, n. 4, p. 1-20, 2017. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3039125. Acesso em: 30 out. 2019.

SHEPARDSON, David. NTSB to determine probable cause of fatal Uber self-driving crash. *Reuters*, out. 2019. Disponível em: <https://www.reuters.com/article/us-uber-crash-safety/ntsb-to-determine-probable-cause-of-fatal-uber-self-driving-crash-idUSKBN1WW2IR>. Acesso em: 30 out. 2019.

SIMON, Herbert. *The Shape of Automation for Men and Management*. New York: Harper & Row, 1965.

SINGH, Sameer; RIBEIRO, Marco Túlio; GUESTRIN, Carlos. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, 2016. Disponível em: <https://arxiv.org/pdf/1602.04938.pdf>. Acesso em: 30 out. 2019.

SNAPPER, John W. Responsibility for Computer Based Errors. *Metaphilosophy*, v. 16, n. 4, 1985, p. 289-295.

SOLAIMAN, S. M. Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy. *Artificial Intelligence and Law*, v. 25, n. 2, 2017, p. 155-179.

SOLUM, Lawrence B. Legal Personhood for Artificial Intelligences. *North Carolina Law Review*, v. 70, n. 4, 1992, p. 1231-1287.

SOTO, Ricardo Alonso. Responsabilidad civil y seguro. In: PANTALEÓN, Fernando (Org.). *La Responsabilidad en el Derecho*. Anuario de la Facultad de Derecho de La Universidad Autónoma de Madrid. Madrid: Universidad Autónoma de Madrid, 2001, p. 193-204.

STANCIOLI, Brunello. *Renúncia ao Exercício de Direitos da Personalidade: Ou como alguém se torna o que quiser*. Belo Horizonte: Del Rey, 2010.

SURDEN, Harry; WILLIAMS, Mary-Anne. Technological Opacity, Predictability, and Self-Driving Cars. *Cardozo Law Review*, v. 38, p. 121-181, 2016.

TAYLOR, Charles. *Human Agency and Language (Philosophical Papers v. 1)*. Cambridge University Press: Cambridge, 1985, p. 98.

TEPEDINO, Gustavo; SILVA, Rodrigo. Inteligência Artificial e Elementos da Responsabilidade Civil. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 293-325.

TEUBNER, Gunther. *Networks as connected contracts*. Oxford: Hart Publishing, 2011.

TEUBER, Gunther. Enterprise Corporatism: New industrial policy and the "essence" of the legal person. *The American Journal of Comparative Law*, v. 36, n. 1, 1988, p. 130-155.

TEUBNER, Gunther. Rights of non-humans? Electronic agents and animals as new actors in politics and law. *Journal of Law and Society*, v. 33, n. 4, 2006, p. 497-521.

The Telegraph. Sandra the Orangutan granted limited human rights, 2014. Disponível em: <https://www.telegraph.co.uk/news/worldnews/southamerica/argentina/11307205/Sandra-the-orangutan-granted-limited-human-rights.html>. Acesso em: 01 abr. 2020.

TUOMELA, Raimo. *Social Ontology: Collective Intentionality and Group Agents*. Oxford: Oxford University Press, 2013.

TURING, Alan. Computing Machinery and Intelligence. *Mind*, v. 49, n. 236-460, 1950. Disponível em: <https://phil415.pbworks.com/f/TuringComputing.pdf>. Acesso em: 29 nov. 2019.

TURNER, Jacob. *Robot Rules: Regulating Artificial Intelligence*. Nova York: Palgrave Macmillan, 2019.

UNIÃO EUROPEIA. *Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: Artificial Intelligence for Europe (COM/2018/237)*. Bruxelas: Comissão Europeia, 2018. Disponível em: <https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>. Acesso em: 29 out. 2019.

UNIÃO EUROPEIA. *European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103-INL)*. Estrasburgo: Parlamento Europeu, 2017. Disponível em: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html?redirect. Acesso em: 30 out. 2019.

UNIÃO EUROPEIA. *Regulamento 2016/679 do Parlamento Europeu e do Conselho, de 27 de abril de 2016*. Relativo à proteção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados e que revoga a Diretiva 95/46/CE (Regulamento Geral sobre a Proteção de Dados). Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679&from=pt>. Acesso em: 26/08/2019.

VAN PELT, Meredith. Autocephalous Greek Orthodox Church of Cyprus v. Goldberg and Feldmen Fine Arts, Inc.: A case for the use of civil remedies in effecting the return of stolen art. *Dickinson Journal of International Law*, v. 8, n. 3, 1990, p. 441-464. Disponível em: <https://elibrary.law.psu.edu/cgi/viewcontent.cgi?article=1126&context=psilr>. Acesso em: 01 abr. 2020.

VERONESE, Alexandre; SILVEIRA, Alessandra; LEMOS, Amanda. Inteligência Artificial, Mercado Único Digital e a Postulação de um Direito às inferências Justas e Razoáveis: Uma questão jurídica entre a ética e a técnica. In: FRAZÃO, Ana; MULHOLLAND, Caitlin

(Orgs.). *Inteligência Artificial e Direito: Ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 233-264.

VLADECK, David C. Machines Without Principles: Liability Rules and Artificial Intelligence. *Washington Law Review*, v. 89, n. 01, p. 117-150, 2014.

WAINWRIGHT, Martin. Robot fails to find a place in the sun. *The Guardian*, 20 jun. 2002. Disponível em: <http://www.theguardian.com/uk/2002/jun/20/engineering.highereducation>. Acesso em: 29 out. 2019.

WACHTER, Sandra; MITTELSTADT, Brent; FLORIDI, Luciano. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. *International Data Privacy Law*, p. 1-47, 2017. Disponível em: <https://ssrn.com/abstract=2903469>. Acesso em: 25 ago. 2019.

WEIN, Leon E. The Responsibility of Intelligent Artifacts: Toward an Automation Jurisprudence. *Harvard J.L & Tech*, v. 6, n. 103, p. 103-154.

WINFIELD, Alan. *Robotics: A very short introduction*. Oxford: Oxford University Press, 2012.