UNIVERSIDADE DE MARÍLIA

ANDRÉ FERREIRA MARQUES

INTELIGÊNCIA ARTIFICIAL: REGULAÇÃO ÉTICA E RESPONSABILIDADE CIVIL

MARÍLIA
2020

ANDRÉ FERREIRA MARQUES

INTELIGÊNCIA ARTIFICIAL: REGULAÇÃO ÉTICA E RESPONSABILIDADE CIVIL

Dissertação apresentada ao Programa de Mestrado Interinstitucional em Direito da Universidade de Marília e Centro Universitário U:Verse, área de concentração Empreendimentos Econômicos, Desenvolvimento e Mudança Social, como requisito parcial para obtenção do título de Mestre em Direito, sob a orientação do Prof. Dr. Rogerio Mollica.

MARÍLIA
2020

Marques, André Ferreira
Inteligência Artificial: regulação ética e responsabilidade civil / André Ferreira Marques – Marília: UNIMAR, 2020.
138f.

Dissertação (Mestrado em Direito – Empreendimentos Econômicos, Desenvolvimento e Mudança Social – Universidade de Marília, Marília, 2020.
Orientação: Prof. Dr. Rogerio Mollica

1. Danos 2. Inteligência Artificial 3. Machine Learning 4.Responsabilidade Civil I. Marques, André Ferreira

CDD – 340

ANDRÉ FERREIRA MARQUES

INTELIGÊNCIA ARTIFICIAL: REGULAÇÃO ÉTICA E RESPONSABILIDADE CIVIL

Dissertação apresentada ao Programa de Mestrado Interinstitucional em Direito da Universidade de Marília e Centro Universitário U:Verse, área de concentração Empreendimentos Econômicos, Desenvolvimento e Mudança Social, como requisito parcial para obtenção do título de Mestre em Direito, sob a orientação do Prof. Dr. Rogerio Mollica.

Aprovado pela Banca Examinadora em 11/11/2020.

_____

Prof. Dr. Rogerio Mollica
Orientador

_____

Prof. Dr. Valter Moura do Carmo

_____

Prof. Dr. Vinícius Silva Lemos

_____

Prof. Dr. Felipe Calderón-Valencia

Dedico este trabalho aos meus primos Amanda Gabriele e Pedro Henrique (*in memorian*), uma princesa e um príncipe que, em tão pouco tempo, puderam nos ensinar tanto a respeito de amor, resiliência, fé e esperança.

É difícil agradecer adequadamente a todos que contribuíram direta e indiretamente para que fosse possível a realização deste trabalho. No entanto, não poderia deixar de mencionar alguns, em especial: em primeiro lugar, a Deus, que me concedeu a sabedoria, força e parcimônia necessárias para seguir sempre avançando.

Aos professores do programa de Mestrado interdisciplinar, pela confiança no projeto, dedicação e ensinamentos dedicados à turma, sobretudo ao meu orientador, Dr. Rogerio Mollica, quem direcionou e incentivou o meu desenvolvimento acadêmico.

À minha irmã Danielle, sobrinhos Guilherme e Júlia, tio Dudu, meu amigo Péricles, meus sócios Marciano e Luiz Carlos, demais amigos e colegas de trabalho que tiveram a compreensão em virtude das constantes ausências no escritório, no convívio social e familiar, pelo companheirismo, compartilhamento deste momento sempre com palavras de incentivo e apoio.

Aos meus pais, Mariano e Socorro, que sempre foram para mim exemplos de vida, e puderam me proporcionar tudo o que de melhor eu podia esperar: educação, amor incondicional e confiança.

E por fim, à minha amada esposa Bianca, quem teve a árdua missão de conviver diariamente ao meu lado por todo este período, desde o ingresso no programa até a elaboração do presente trabalho, durante uma pandemia mundial, em isolamento social por meses a fio, me suportando e apoiando nos momentos mais difíceis e vibrando comigo com as pequenas vitórias havidas, demonstrando o acerto da decisão de compartilhar uma vida ao seu lado. Obrigado a todos.

INTELIGÊNCIA ARTIFICIAL: REGULAÇÃO ÉTICA E RESPONSABILIDADE CIVIL

Resumo: A inteligência artificial tem evoluído exponencialmente nos últimos anos diante da conjunção autorreferente de fatores como o avanço computacional, *big data* e o desenvolvimento algorítmico, gerando reflexos e repercussões significativas no sistema do direito que parecem tencionar a submissão deste à sua lógica. A partir deste contexto, o presente estudo tem por problema central analisar o tratamento que deve ser conferido aos sistemas de inteligência artificial no que se refere à responsabilização civil decorrente de danos por eles causados. A pesquisa, de caráter aplicada e descritiva, se desenvolve por meio do método hipotético-dedutivo, o qual dialoga com o método sistêmico, trazendo uma abordagem qualitativa à revisão bibliográfica e documental, a fim de compreender como se estrutura a lógica de funcionamento do processo de tomada de decisão por algoritmos baseados em *machine learning*, com vistas a aferir como se dá a responsabilidade civil, uma vez que ausente propriamente uma configuração humana que parametrize a atividade da máquina causadora do dano. São perquiridos, para tanto, conceitos técnicos e filosóficos ligados à inteligência artificial a fim de entender os principais desafios jurídicos de sua aplicação no processo de tomada de decisão, os quais os actantes internacionais integrantes da academia, indústria e governos objetivam prevenir, ou pelo menos minorar seus efeitos, por meio de um trabalho difuso de desenvolvimento de uma regulação ética *by design*. Ainda, foi avaliada a adequação da aplicação das teorias de responsabilidade civil existentes no ordenamento jurídico brasileiro para o tratamento da hipótese proposta, seguindo-se de uma avaliação dos modelos e mecanismos de organização, estruturação e responsabilização cogitados atualmente para implementação de uma inteligência artificial segura, confiável e responsável. Por fim, são analisados os projetos de lei em tramitação no Congresso Nacional brasileiro sobre o tema, os quais revelam que ainda se faz necessário um maior amadurecimento social, técnico e, por via de consequência, legislativo da matéria a fim de que seja implementada uma política nacional de inteligência artificial efetiva no Brasil.

Palavras-Chave: Danos. Inteligência Artificial. *Machine Learning*. Responsabilidade civil.

ARTIFICIAL INTELLIGENCE: ETHICAL REGULATION AND CIVIL RESPONSIBILITY

Abstract: Artificial intelligence has evolved exponentially in recent years due to the self-referential conjunction of factors such as computational advancement, big data and algorithmic development, generating significant reflexes and repercussions in the system of law that seem to intend the submission of this to its logic. From this context, the present study has as a central problem to analyze the treatment that should be given to artificial intelligence systems with regard to civil liability resulting from damage caused by them. The research, of an applied and descriptive character, is developed through the hypothetical-deductive method, which dialogues with the systemic method, bringing a qualitative approach to the bibliographic and documentary review, in order to understand how the process's operating logic is structured. decision-making by algorithms based on machine learning, with a view to assessing how civil liability occurs, since there is no human configuration that parameterizes the activity of the machine causing the damage. Therefore, technical and philosophical concepts related to artificial intelligence are investigated in order to understand the main legal challenges of its application in the decision-making process, which the international actors from academia, industry and governments aim to prevent, or at least mitigate its effects, through a diffuse work to develop an ethical regulation by design. Still, the adequacy of the application of the existing theories of civil liability in the Brazilian legal system for the treatment of the proposed hypothesis was evaluated, followed by an assessment of the models and mechanisms of organization, structuring and accountability currently considered for the implementation of a safe, reliable and responsible artificial intelligence. Finally, the bills in progress at the Brazilian National Congress on the subject are analyzed, which reveal that further social, technical and, consequently, legislative maturation of the matter is still necessary in order for it to be implemented. an effective national artificial intelligence policy in Brazil.

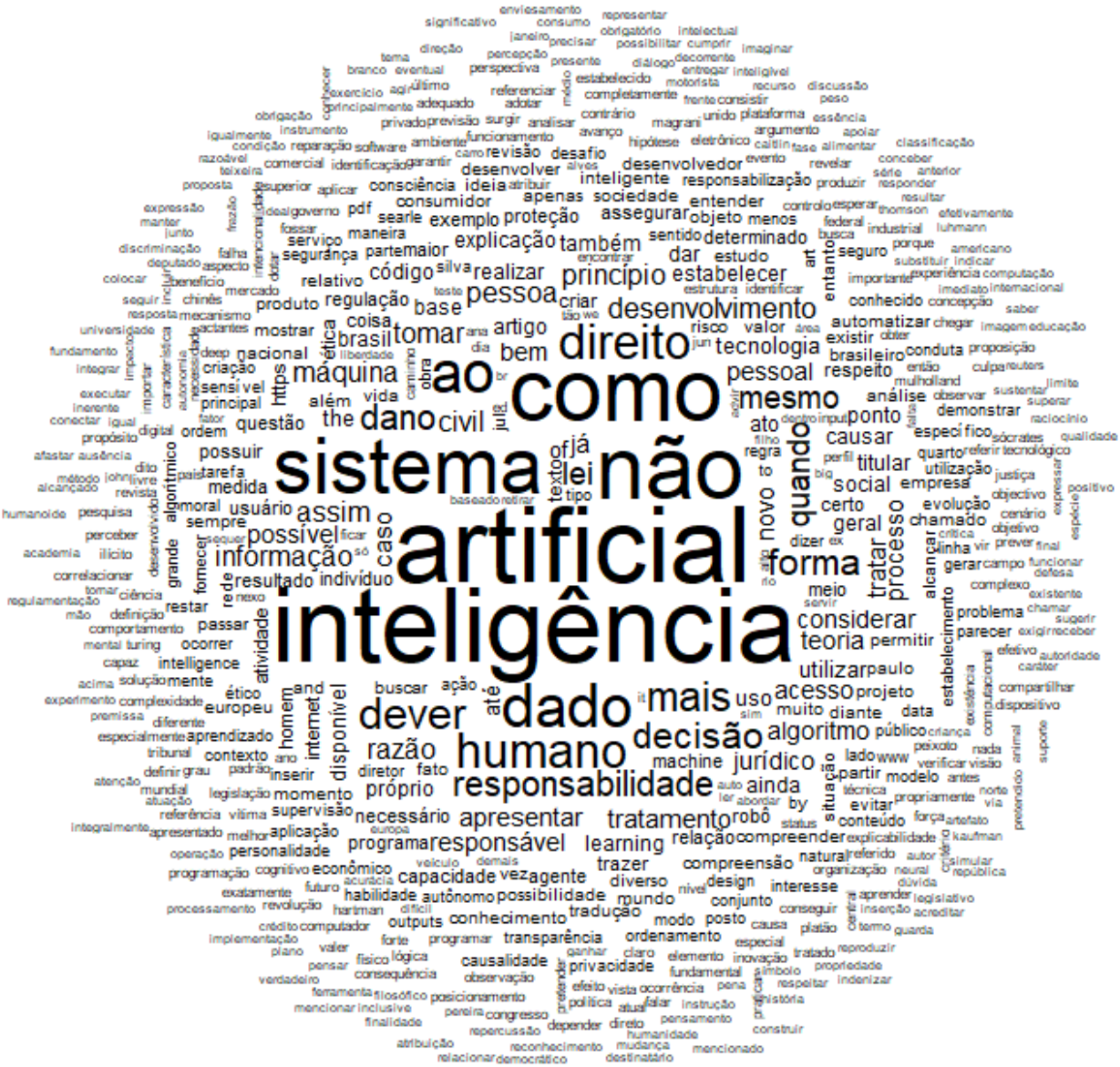
Keywords: Damages. Artificial intelligence. Machine Learning. Civil responsibility.

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AI – *Artificial Intelligence*
ANPD – Autoridade Nacional de Proteção de Dados
Art. – Artigo
CC – Código Civil
CDC – Código de Defesa do Consumidor
CN – Congresso Nacional
CNJ – Conselho Nacional de Justiça
COMPAS – *Correctional Offender Management Profiling for Alternative Sanctions*
DNA – *Deoxyribonucleic acid*
DoD – *Department of Defesense (*Departamento de Defesa dos Estados Unidos da América)
EUA – Estados Unidos da América
GDPR – General Data Protection Regulation
IA – Inteligência Artificial
IoT – *Internet of Things*
LGPD – Lei Geral de Proteção de Dados
LRP – *Layer-wise Relevance Propagation*
MCI – Marco Civil da Internet
n. – Número
OCDE – Organização para a Cooperação e Desenvolvimento Econômico
PNIA – Política Nacional de Inteligência Artificial
REsp – Recurso Especial
SPRA – *Spectral Relevance Analysis*
STF – Supremo Tribunal Federal
STJ – Superior Tribunal de Justiça
SVM – *Support Vector Machine*
TIC – Tecnologias da Informação e Comunicação
UNHRC – *United Nations Human Right Council*
V.g. – *Verbi Gratia*
xAI – *Explainable Artificial Intelligence*

NUVEM DE PALAVRAS ELABORADA PELO SOFTWARE IRaMuTeQ A PARTIR DO MATERIAL DE PESQUISA PRODUZIDO[1]



[1] Imagem formada a partir do texto produzido no trabalho, formando uma nuvem de palavras com a organização e proporção correspondente à sua incidência de aparição no texto. É uma ferramenta disponibilizada pelo *software* IRaMuTeQ (*Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*), desenvolvido por Pierre Ratinaud no Laboratório LERASS, da Universidade de Toulouse, na França.

SUMÁRIO

INTRODUÇÃO

De há muito permeia o imaginário popular a ideia de um ser não-humano inteligente, inspirando obras literárias, teatrais e cinematográficas que dão vida à uma visão de inteligência artificial presente em diversos aspectos da rotina diária das pessoas. Em todas essas obras ficcionais, o ponto convergente é a existência de seres, com maior ou menor grau de autonomia, capazes de evoluir com suas próprias experiências, interagindo com o ambiente e os humanos nos seus mais variados campos, como transportes, médico-farmacêutico, jurídico, dentre outros.

Apesar de a vida real ainda se encontrar distante do que já se imaginou em algumas das obras, sem dúvidas temos avançado significativamente em direção a um novo cenário mundial. O advento da internet e da evolução da computação pessoal, objetos centrais daquilo que foi considerada a terceira revolução industrial, recombinados entre si e somados a uma ubiquidade jamais vista em todos os aspectos e uma potencialidade sem precedentes dos artefatos tecnológicos, justificam a defesa da ocorrência de um ponto de inflexão que culminaria com a quarta revolução industrial, visto que se tratam de tecnologias emergentes que trazem novas formas de perceber o mundo, desencadeando uma modificação sistemática e substancial nos sistemas econômicos e nas estruturas sociais, quiçá o início de uma nova era na periodização clássica da história da humanidade, que se encontra na idade contemporânea desde o ano de 1789, início da revolução francesa.

Isso porque a idade contemporânea ou contemporaneidade, teve significativo destaque em razão da consolidação do capitalismo e busca das grandes nações europeias por território, mercados e matéria prima. Mas, o mundo já avançou muito desde então. Passadas duas guerras mundiais, questiona-se tanto o modelo de divisão da história, quanto quando será considerada encerrada a atual idade contemporânea.

Era do conhecimento, era da abundância, era pós-digital, era exponencial, era da colaboração e era complexa são exemplos de algumas das denominações que têm sido atribuídas ao momento atual de globalização, que se assemelha a um modelo panóptico virtual voluntário, que movimenta a economia por meio de um capitalismo de vigilância em que os indivíduos se sujeitam por conta própria, na maior das vezes, sem nem mesmo ter ideia do que estão fazendo.

As novas Tecnologias da Informação e Comunicação (TICs) estão presentes em tantos lugares no nosso dia-a-dia, que, por vezes, nem são percebidas. Plataformas de *streaming*[2] de

[2] Ferramenta de transmissão instantânea de dados utilizada para distribuir conteúdo multimídia pela internet, sem a necessidade da descarga dos dados.

música e vídeo e redes sociais que sugerem o conteúdo a ser consumido; lojas virtuais que indicam produtos; ferramenta de filtro de spam em e-mails; dispositivos eletrônicos para monitoramento de saúde, sono, atividades físicas; sistema de reconhecimento de imagens, que pode ser utilizado desde redes sociais a identificação de criminosos ou pessoas desaparecidas; veículos e aeronaves autômatos, dentre outros.

E é nesse contexto caórdico e disruptivo que a inteligência artificial ganha espaço de maneira exponencial, especialmente em razão do mencionado avanço da tecnologia computacional, bem como da internet das coisas (*internet of things - IoT*), que permitiu uma explosão de dados indexados (conhecida por *big data*) gerados por sensores e artefatos interligados à rede mundial de computadores. Com esse pano de fundo, a criação de algoritmos pôde ser sobremaneira intensificada, chegando ao ponto de desenvolvimento em que a máquina se auto programa a partir de um modo de treinamento em que lhes são entregues os dados brutos (*inputs*) e os resultados pretendidos a partir destes dados (*outputs*), não sendo possível aos seus criadores trilharem o exato caminho percorrido pela máquina para chegar de um ponto a outro. Esta tecnologia ficou conhecida como *machine learning*, e é considerada espécie do gênero inteligência artificial, a qual possui ainda uma subespécie, cujo mecanismo de aprendizagem é ainda mais intenso, chamada de *deep learning* ou *aprendizado profundo*.

Isso modifica completamente o cenário habitual em que sempre foi possível antecipar, controlar e mitigar os riscos advindos das tecnologias desenvolvidas até então, tais como armas, carros e aviões, onde era facilmente possível a identificação de uma conduta e nexo de causalidade vinculados a um responsável que tinha contra si imputada a responsabilidade pela ocorrência do dano. No entanto, a sofisticação da autonomia e complexidade dos sistemas de inteligência artificial fez surgir, indiscutivelmente, um novo e grande desafio ao controle, capacidade de previsão e compreensão dos homens sobre as máquinas, que resulta em cenários inusitados no que diz respeito à responsabilização civil decorrente de danos causados por inteligência artificial.

As ações de *machine learning* e *deep learning*, portanto, passam a ser consideradas como se não tivessem qualquer predeterminação no que diz respeito ao caminho a ser percorrido e forma de agir para se alcançar o resultado, visto que o algoritmo que é seguido não foi programado passo-a-passo pelo homem, e a máquina se utiliza de mecanismos em que a explicabilidade é inversamente proporcional à acurácia da decisão tomada. Tal fato ganha relevo na seara jurídica haja vista a possibilidade de ocorrência de danos causados pela inteligência artificial, ponto em que surge o problema a ser enfrentado por este trabalho, sobre como será (e, principalmente, como deveria ser) estabelecida a responsabilidade civil em tal

hipótese, com base no ordenamento jurídico pátrio, e se este é o bastante para assegurar uma integral reparação do dano.

Tomou proporções mundiais o acidente ocorrido com o veículo autônomo da empresa multinacional Tesla, no ano de 2016, em que o motorista veio à óbito quando se utilizava do dispositivo de direção autônoma da fabricante. O carro teria se chocado com um caminhão, enquanto o motorista assistia a um filme. A investigação concluiu que, inobstante o sistema tenha avisado ao motorista, sem sucesso, por sete vezes, para reestabelecer o controle do veículo, nem os sensores do carro – e nem o motorista – detectaram o caminhão cruzando a pista, visto que o reflexo da luz solar teria confundido o sistema, que acreditou que se tratava do céu. Em 2018 foi a vez da *startup unicórnio*[3] Uber ter um acidente fatal, quando o seu protótipo de veículo autônomo atropelou e matou uma pessoa que atravessou a rua fora da faixa de pedestres. O software não foi programado para compreender que pessoas poderiam ignorar a regra de atravessar a rua na faixa de segurança, e, portanto, não a reconheceu como um pedestre. Em casos como esses, como se daria a responsabilidade civil no Brasil? Aparentemente, a questão não se apresenta tão complexa, porém, ao explorar os casos com mais vagar, verifica-se que são várias as questões a serem equacionadas.

Imagine-se o cenário quando inseridos elementos de auto-otimização, como no caso do *chatbot* Tay, sistema de inteligência artificial desenvolvido pela empresa Microsoft, criado para conversar com pessoas de forma divertida nas redes sociais, que, em menos de 24 (vinte e quatro) horas se transformou em um robô com posicionamentos ideológicos racistas, antifeministas, nazistas e homofóbicos. Tal fato ocorreu em razão da utilização de inteligência artificial na modalidade *deep learning*, onde o sistema deveria aprender e evoluir a partir das experiências obtidas em razão da interação com os usuários da rede social. Ocorre que estes usuários estimularam o desenvolvimento do referido comportamento, que foi assimilado e reproduzido pelo *chatbot*. Nesse outro caso, como se daria a responsabilidade civil de eventuais danos causados pela Tay?

O ordenamento jurídico possui regras bem definidas acerca da responsabilidade civil, as quais poderiam facilmente ser aplicadas aos casos de danos causados por ato praticado por sistemas de inteligência artificial, dentre os quais destacam-se a teoria da causa direta e imediata, responsabilidade noxal, teoria do resultado mais grave, teoria do risco proveito, teoria do *deep pocket*, causalidade múltipla, rompimento do nexo causal, fortuito interno, presunção de causalidade, teoria do risco do desenvolvimento e causalidade alternativa. Contudo,

[3] Termo criado em 2013 por Aileen Lee, utilizado para referenciar empresas que atingem uma avaliação de preço de mercado superior a 1 bilhão de dólares, sem terem capital na bolsa de valores.

superando o juízo delibatório, nota-se que, seja em razão da imprevisibilidade algorítmica, do aprendizado da máquina a partir dos dados e do próprio usuário, e até mesmo da possível ausência de um ato ilícito ou falha do sistema, a questão não é tão simplória como parece ser.

Vários países têm caminhado precipuamente em duas frentes: de estabelecer uma regulação *by design*, que pretende, com a fixação de princípios éticos e rígidas diretrizes de segurança, não discriminação, explicabilidade, transparência, *accountability* etc., e fazendo com que as máquinas inteligentes sejam programadas de forma a respeitar tais valores, evitar a causação de danos ou que entrem no mercado sistemas de inteligência artificial que tenham uma alta probabilidade de causar danos identificada desde sua programação. De outro lado, também são buscadas formas para atribuir responsabilidade em razão do dano causado, de modo a assegurar a integral reparação da vítima, com possibilidades que vão desde conferir personalidade jurídica ao sistema de inteligência artificial, até a constituição de capital ou afetação de patrimônio específico e criação de seguros obrigatórios.

Com vistas ao desenvolvimento da investigação descritiva proposta, a qual se alinha à área de concentração do programa de mestrado, qual seja, empreendimentos econômicos, desenvolvimento e mudança social, especialmente na linha de pesquisa perseguida que trata das relações jurídicas, diante do objeto ligado às novas e necessárias relações jurídicas de uma temática nova, foi adotado o método hipotético-dedutivo, posto que considerada a hipótese de ser o ordenamento jurídico brasileiro capaz de assegurar a integral reparação de danos causados por ato praticado por inteligência artificial, partindo-se de um procedimento de revisão bibliográfica e documental das principais obras e autores sobre o tema, em uma abordagem qualitativa a respeito dos pressupostos e dados colacionados. Importante salientar também o diálogo do trabalho com o método sistêmico, que proporciona a interdiciplinariedade necessária para que seja alcançada a sua finalidade de aplicação prática.

Para tanto, o presente trabalho está dividido em três capítulos, o primeiro deles iniciando-se com um referencial de caráter técnico e teórico atinente à inteligência artificial, que pretende compreender o funcionamento axiológico dos sistemas, a lógica de funcionamento de suas espécies, e a evolução de sua percepção no contexto de um mundo complexo.

Em seguida, o segundo capítulo aborda algumas das questões jurídicas sensíveis relativas ao desenvolvimento e uso da inteligência artificial no processo de tomada de decisão, especialmente o enviesamento dos dados, falta de explicabilidade e supervisão humana, ao passo que deflagra o estudo do trabalho difuso, porém convergente, realizado pelos actantes internacionais da academia, indústria e governos, a fim de estabelecer uma tecnorregulação sobre a matéria.

Ato contínuo, a partir dessas premissas, o terceiro e último capítulo passa a tratar propriamente do problema do presente trabalho, analisando, inicialmente a implementação de uma regulação ética *by design*, com vistas à prevenção da causação do dano, enfrentando novos desafios relativos aos danos causados pelos sistemas de inteligência artificial. Prossegue aferindo a aplicação e compatibilidade das várias teorias de responsabilidade civil aos casos de danos causados por sistemas de inteligência artificial, avaliando os novos mecanismos propostos para otimizar responsabilização em tais casos, para, ao final, escrutinar as proposições legislativas correlacionadas em trâmite no Congresso Nacional.

Por fim, vislumbrado o caráter policontextual dos desafios, busca apresentar as considerações finais acerca do problema proposto, de modo a compatibilizar os avanços tecnológicos com a necessária segurança jurídica para o desenvolvimento e uso de sistemas de inteligência artificial, por meio do estabelecimento de princípios e diretrizes de governança ética, com a definição de um plano claro de responsabilidade civil, minorando, assim, as suas vulnerabilidades com o propósito de permitir o enfrentamento da situação de forma séria e responsável.

1 ONTOLOGIA DA INTELIGÊNCIA ARTIFICIAL: DO IMAGINÁRIO ÀS REDES NEURAIS ARTIFICIAIS

O ideário popular inconsciente e involuntariamente cria uma concepção futurista quando se pensa em inteligência artificial, remetendo a carros voadores, robôs humanoides e assistentes pessoais super eficientes, ignorando, a princípio, a proximidade visceral que hoje ela possui com cada indivíduo, em maior ou menor grau, nas suas mais básicas atividades diárias. Desta forma, antes de iniciar o estudo acerca dos vieses da responsabilidade civil sobre os sistemas de inteligência artificial causadores de danos, faz-se necessário explorar as premissas técnicas e filosóficas que lhe são pertinentes.

1.1 O QUE É INTELIGÊNCIA ARTIFICIAL?

Acredita-se que a vida no planeta tenha se iniciado no mar, e que posteriormente migrou para a terra firme, quando em um determinado momento uma das criaturas marinhas se alçou ao solo, forçando uma transformação evolucionária. Presume-se, ainda, que teria sido compelida a fazê-lo, diante do extremismo da situação, a qual certamente custou a vida de algumas gerações até que fossem desenvolvidas as habilidades necessárias para se viver fora da água. Vê-se, assim, que a percepção da evolução é algo natural, posto que inevitável (TOLLE, 2007, p. 25).

Perceber e encarar de frente tal evolução, portanto, pode ser o diferencial para o posicionamento ativo neste processo de transformação. Desta maneira tem sido tratado o desenvolvimento da *inteligência artificial* pela maior parte dos actantes. A referida expressão, como é conhecida mundialmente hoje, foi cunhada no ano de 1956, na conferência de verão do Darthmouth College, em New Hampshire, nos Estados Unidos (SILVA, 2019), que reuniu dez grandes estudiosos da ciência da computação, dentre eles Trenchard More e John McCarthy, para debater a possibilidade de existirem computadores que pudessem ser capazes de reproduzir ações cognitivas humanas (LOPES, 2020). No entanto, antes disso, tem-se registro de estudos e experimentos relacionados ao tema, como os de Warren McCulloch e Walter Pitts, que conceberam o primeiro modelo matemático do neurônio (VIEGAS, 2020), em 1943, e, principalmente, os feitos de Alan Turing, datados de 1950, quando, por meio do seu artigo intitulado *Computer machinery and intelligence*, instigou todo o mundo com o provocante questionamento se uma máquina poderia pensar (TURING, 1950).

Alan Mathison Turing (1912 - 1954), matemático de formação, nascido na Inglaterra, é considerado o pai da computação em razão da sua grande contribuição para a ciência da computação e formação do conceito de algoritmo. Aos 24 anos de idade, após muitos insucessos, teve sua resiliência premiada quando criou a primeira máquina que, por meio de um sistema automatizado, podia fazer leituras computacionais, manipulando símbolos de um sistema de regras próprias (HODGES, 2004).

Desenvolveu em paralelo o famoso teste de Turing, também chamado de *jogo da imitação*, que tinha como princípio geral a premissa de que uma máquina se igualaria a um humano quando não fosse possível distinguir o seu comportamento deste. O teste consistia basicamente em designar uma máquina dotada de inteligência artificial para responder perguntas de um interlocutor, sem que este soubesse que interagia com uma máquina. Se ele descobrisse que se tratava de um robô, teria falhado a máquina. Assim, ilustrou o jogo da imitação da seguinte forma: são três jogadores, um investigador (C) em uma sala, em outra sala um indivíduo (A) e em uma terceira sala, outro indivíduo (B). Nenhum dos três manteria contato com os demais, seja visual, por voz ou por outro meio que não um texto escrito, digitado em um teclado que apareceria em uma tela. As perguntas seriam formuladas pelo investigador "C", buscando identificar qual o gênero dos indivíduos "A" e "B", os quais não estavam obrigados a falar apenas a verdade. As respostas deveriam ser binárias, afirmativas ou negativas ("sim" ou "não") (TEIXEIRA, 2014, p. 30). Passaria no *teste de Turing* o robô que conseguisse se fazer confundir com um ser humano (XAVIER; SPALER, 2019), registrando-se, contudo, que, segundo o próprio Turing, passar no teste não significaria necessariamente que computadores poderiam duplicar a mente humana (AMORIM, 2014).

De outro lado, interessante e oportuna a lembrança da existência da ideia de um ser não humano, não divino, criado pelo próprio homem, dotado de certa autonomia ou mesmo inteligência, a exemplo do mito do *Golem*. Trata-se de uma lenda judia que conta que, no ano de 1580, seguindo um sonho profético de proteção ao seu povo, teria sido criado um enorme ser de argila, que, após rituais religiosos que duraram cerca de três dias, ganhou vida pelas mãos de um Grão-rabino de Praga chamado Judah Loew (1520 - 1609), de nome religioso Maharal, a quem a criatura passou a servir e proteger. Durante anos, o gigante teria mantido o povoado judeu livre das várias ameaças existentes na época, até o dia em que o Grão-rabino entendeu que sua missão tinha sido cumprida, e o destruiu em 1590. A polêmica sobre a veracidade da estória persiste na crença dos judeus mais religiosos, ratificado pela declaração do rabino Moishe New, líder chassídico do Canadá, que afirmou que "O Talmud relata momentos nos quais outros *Golems* foram criados para proteger a vida dos judeus [...] Ao corpo do *Golem* foi

dada uma alma e assim tornou-se uma espécie de anjo dentro de um corpo feito pelo homem" (KAUFMAN, 2019, p. 11).

No mesmo sentido, tem-se o que se considera a primeira obra de ficção científica, publicada em 1818, chamada de *Frankenstein ou o Prometeu Moderno* ("*Frankenstein: or the Modern Prometheus*"), romance de terror gótico criado pela escritora Mary Shelley (1797 - 1851), inspirado na história do *Golem*, em que um estudante de ciências naturais chamado Victor Frankenstein dá vida a um monstro gigantesco em seu laboratório. Ao contrário daquele, o Frankenstein, que após fugir para a floresta e aprender sozinho a se comunicar e a se alimentar, possuía, além de autonomia total, inteligência própria e até mesmo sentimentos semelhantes aos atuais robôs da ficção científica hollywoodiana (KAUFMAN, 2019, p. 12). Na obra, o jovem cientista, apesar de ter sido salvo pela criatura de hostilidades iminentes que sofrera, em um dado momento se arrepende da promessa que tinha feito de dar vida a uma companheira, com receio de que criasse uma raça de monstros que pudesse ameaçar não apenas a si, mas a toda a humanidade (SHELLEY, [1818?], p. 173).

> Se a criatura já existente abominava sua própria deformidade, poderia ver recrudescido o seu ódio quando a visse apresentar-se em forma feminina. Ela, por sua vez, poderia vir a ter aversão por ele, inclinando-se pela beleza do homem normal. Nesse caso o abandonaria e o monstro voltaria a ficar só, mais exasperado ainda pelo fato de ser desprezado por alguém de sua própria espécie. Mesmo que viessem a deixar a Europa e habitar as paragens do Novo Mundo, poderia advir que um dos primeiros resultados do relacionamento por que suspirava o monstro fosse a geração de filhos, e uma raça de demônios se propagaria pela face da Terra, espalhando o terror entre a espécie humana. Tinha eu o direito de, em meu próprio benefício, infligir tal maldição às gerações vindouras? Deixara-me levar pelos sofismas do ser que eu criara, e suas ameaças diabólicas tinham-me perturbado o juízo. Agora, porém, pela primeira vez, a incongruência de minha promessa se me revelava de chofre. Estremeci ao pensar na condenação que as gerações futuras poderiam fazer recair sobre mim, que não hesitara em comprar a própria paz ao preço, talvez, do flagelo de toda a raça humana (SHELLEY, [1818?], p. 154).

De seu turno, João de Fernandes Teixeira (2014, p. 8) ilustra o viés evolutivo da concepção de inteligência artificial em três diferentes níveis, se valendo para tanto do jogo de xadrez. A primeira delas seria o homem imitando uma máquina. A segunda, uma máquina imitando o homem, e a terceira, propriamente o que seria a inteligência artificial, quando uma máquina consegue espelhar e superar a mente humana. De início, referencia a história do barão de Kempelen, cujo título não se sabe ao certo se era verdadeiro ou falso, que, no século XIX anunciou a criação de uma máquina inteligente que jogava xadrez. Em verdade tratava-se de um anão enxadrista que era posicionado dentro de uma caixa que lhe permitia movimentar as peças do jogo, fazendo parecer que a máquina movimentava as peças. Quem olhava para a máquina, não suspeitava o que realmente acontecia. O seu inventor teria alcançado sucesso e

dinheiro, se apresentando em circos por toda a Europa, chegando a chamar atenção de Napoleão, que quis conhecer a máquina inteligente. No entanto, a "máquina pensante" cometeu o erro de começar a ganhar a partida, provocando a ira de Napoleão, que, com um chute no artefato, viu as portinholas se abrirem e o anão aparecer, revelando o truque de Kempelen.

O segundo nível, seria a tentativa de a máquina imitar o homem. Nessa hipótese tem-se a criação de sistemas que se utilizam de dois caminhos possíveis para alcançarem seu propósito: o esgotamento de todas as possibilidades viáveis do que se almeja, o que é chamado de "força bruta", que pode ser facilmente visualizado com a tentativa de se revelar um determinado código numérico de cinco dígitos, em que a máquina tenta, usando de todas as cem mil probabilidades de combinações possíveis dos dez algarismos arábicos, de zero a nove. O caminho contraposto à "força bruta" é o da *heurística*. Esta consiste, basicamente, em uma simulação de raciocínios e estabelecimento de estratégias na tentativa de descartar opções improváveis – ou iniciar por opções mais viáveis –, como se fosse buscado um atalho. A essência da mecânica heurística prima por um raciocínio mais sofisticado, seletivo. Por certo, não é indene de falhas o raciocínio heurístico, sobretudo no início de seu desenvolvimento, no século XX (TEIXEIRA, 2014, p. 9).

A terceira geração, por assim dizer, poderia ser representada por uma máquina chamada de *deep blue*, surgida no final do século XX, que, no ano de 1997, superou o então campeão mundial de xadrez, Gary Kasparov, em uma partida mundialmente conhecida. A máquina *deep blue* foi criada com base no sistema de "força bruta" e, em verdade, se tratava de um supercomputador, alimentado com um enorme banco de dados com as mais diversas possibilidades do jogo, bem como com todas as jogadas dos melhores enxadristas das últimas décadas. Tratava-se, pois, de uma máquina, que assim como as anteriores, não pensava, mas tentava imitar o raciocínio humano, contudo, desta vez, com uma capacidade superior de processamento de dados, a ponto de *parecer* pensar (TEIXEIRA, 2014, p. 10).

Percebe-se, assim, que não importa exatamente a forma da máquina, desde que ela produza – ou mesmo simule – a inteligência, entregando os resultados esperados. Por isso as máquinas podem ser acopladas a um suporte físico (hardware) com uma forma humana, o que convencionalmente seria chamado de robô humanoide, ou não. Pode a máquina pensante ser apenas um software, que realiza as funções que lhe são estabelecidas, sem nada parecer com o homem ou com o que reproduz artificialmente.

Isso já tinha sido vislumbrado por Turing, que entendia que o ponto central da inteligência é o aprendizado, então, a computação, como concebida e vislumbrada naquele momento, não se tratava de um processo inteligente, mas sim um processo lógico, que

dependeria da lista de comandos que lhe seria fornecida (VERONESE; SILVEIRA; LEMOS, 2019).

> A ideia de uma máquina que aprenda pode parecer paradoxal para alguns leitores. Como as regras de operação de máquina mudariam? Elas deveriam descrever completamente como a máquina irá reagir qualquer seja o futuro, quaisquer mudanças que ela sofra. As regras seriam, assim, impermeáveis ao tempo cronológico. Isso é verdade. A explicação do paradoxo é que as regras a serem mudadas no processo de aprendizagem seriam de um caráter muito menos pretensioso, considerando apenas a sua validade efêmera. [...] Podemos ter esperança de que as máquinas irão, eventualmente, competir com os homens em todos os campos puramente intelectuais. Porém, quais serão os mais adequados para iniciar? Mesmo essa é uma decisão difícil. Muitas pessoas poderão pensar que uma atividade muito abstrata, como os jogos de xadrez, seria a melhor (TURING, 1950, p. 458-460).

Tem-se, assim, formada uma concepção do que viria a ser chamado de inteligência artificial, muito embora não exista um conceito uniforme e aceito unanimemente (PAIVA, 2020). É possível, portanto, estabelecer um cerne bem definido, com diferentes obliquidades acerca de determinados aspectos. Para Jahanzaib Shabbir e Tarique Anwer (2015), inteligência artificial se refere à possibilidade de a máquina emular comportamentos que reproduzam a inteligência humana. Já para Miles Brundage, Olle Häggstörm, Peter J. Bentley e Thomas Metzinger (2018), inteligência artificial consistiria em sistemas aptos a desempenhar atividades para as quais se exige inteligência, quando realizada por um indivíduo.

Referencia-se, por fim, a abordagem de Ricardo Dalmaso Marques (2019), que, entendendo o constante e natural desenvolvimento do termo, trata os sistemas de inteligência artificial como máquinas que podem aprender, raciocinar e agir por si próprias quando postas diante de situações novas que guardem alguma pertinência com os padrões inseridos em sua base de dados, descobrindo padrões, identificando tendências e predizendo situações futuras, não devendo ser confundido com um software comum, com um computador ou mesmo com um robô (ANTUNES; CARMO, 2019).

Essa ausência de conceituação fechada, inclusive, entende Peter Stone (2016), fomenta o crescimento e desenvolvimento do campo de estudos por não limitar a visão dos estudiosos, sendo vista a inteligência artificial como um termo guarda-chuva, que abriga várias áreas de estudos e técnicas, dentre as quais cita-se: ciências da computação, linguística, matemática, filosofia, probabilística, neurociência e teoria da decisão.

Registra-se, ainda, o desenvolvimento de quatro categorias de pensamento acerca da inteligência artificial, por Stuart Russel e Peter Norvig (1995). A primeira delas abordou os "sistemas que *pensam como humanos*", que buscava a criação de computadores que pensassem como mentes humanas. A segunda se ocuparia de "sistemas que *agem como humanos*", que

cuidavam de realizar atividades e funções que ordinariamente dependeriam de inteligência humana. Em seguida, os "sistemas que *pensam racionalmente*", que se trata do estudo das habilidades artificiais de máquina que permitem a ela perceber, raciocinar e agir. A quarta e última vertente, seriam os "sistemas que *agem racionalmente*". Sob essa ótica, a inteligência artificial estaria inserta em agentes ou artefatos inteligentes, que teriam, pois, um desempenho inteligente (VIEGAS, 2020).

1.2 A CONSCIÊNCIA COMO ESSÊNCIA DA INTELIGÊNCIA

Tratando de inteligência artificial e do que ela viria a ser, chega-se a um esboço consistente no espelhamento ou reprodução (criação) por uma máquina artificial de uma inteligência humana. Contudo, qual a referência de inteligência humana? O que se entende por inteligência? Para o estabelecimento dessa premissa, será assumida a definição estabelecida por Dora Kaufman como "a capacidade de compreender ideias complexas, de se adaptar efetivamente ao ambiente, de aprender com a experiência, de se envolver em várias formas de raciocínio, de superar os obstáculos" (KAUFMAN, 2019, p. 16).

Partindo-se desta quadra, levantadas algumas perspectivas e características acerca da inteligência artificial, exsurge como questão axiológica central a existência (ou não) de consciência, uma qualidade humana, que é objeto de diversas ciências como filosofia da mente, psicologia, neurologia e ciência cognitiva. Yuval Harari (2016) posiciona a consciência como atributo humano, e descola a consciência da inteligência, sugerindo dois tipos de inteligência: a consciente e a não-consciente. Com essa dicotomia, impõe um limite à inteligência artificial, que acredita que nunca será dotada de consciência, razão pela qual entende que não poderá sentir e tampouco competir com a inteligência humana. Ambas coexistirão, desempenhando papéis diferentes na sociedade moderna[4].

A consciência, que pode ser entendida, então, como compreensão na acepção ora conferida, também foi objeto de estudo de John Searle (1980), um filósofo analítico norte-americano, que lhe imputou duas características: a primeira delas é que a compreensão não pode ser analisada a partir de uma lógica formal, binária ou booleana; não é possível que seja aferida apenas sob um plano cartesiano se existe ou não, de maneira estanque se há completa compreensão ou completa incompreensão (AMORIM, 2014). Existem níveis diversos de

[4] Para isso, o homem deverá desenvolver habilidades próprias, dentre as quais, persuasão, criatividade, empatia, que não podem ser desenvolvidas pela máquina, pois, caso comparados, a máquina executará muito mais rápido, e com um grau de exatidão muito superior do que o ser humano (ALVES; ALMEIDA, 2020).

compreensão, o que se apresenta como uma exceção ao princípio do terceiro excluído[5], razão pela qual de acordo com a lógica *fuzzy*, na qual não existem apenas as respostas extremas, é possível estabelecer uma relação de pertinência que atribui a cada elemento um grau, se apresentando como uma inferência mais próxima da realidade (ALVES; CORRÊA, 2019).

A segunda característica é que a compreensão não pode ser atribuída a uma ferramenta ou a um computador em razão da mera mimetização de determinadas atividades. Tais artefatos, criados pelo homem, podem ser utilizados como auxiliares para execução de tarefas para as quais foram programados, sem que por isso tenham compreensão do que estão fazendo, calculando ou medindo. O computador, *verbia gratia*, não é dotado de nenhuma compreensão, ao contrário do homem, que pode compreender algo de forma total ou parcial.

O interesse de John Searle pela *compreensão* se deu no início da década de 80, quando se preparava, a bordo de um avião, para uma palestra a ser proferida em um simpósio de inteligência artificial. Até mesmo como uma característica inata à sua formação, Searle prezava muito pelo pensamento, o que lhe induzia a uma resistência à inteligência artificial como era apresentada. Por tal posicionamento crítico, se notabilizou com seus estudos que contestam a inteligência artificial forte. Naquele momento, o estado da arte da inteligência artificial era a compreensão de estórias. Uma série de programas eram desenvolvidos com esse espeque, e o que mais intrigava o filósofo, era a alegação de que os programas podiam compreender os textos, e que funcionariam tais quais os seres humanos. Searle não aceitava tal afirmação, por entender que não havia, essencialmente, compreensão por parte dos computadores (TEIXEIRA, 2014, p. 48).

Seria o mesmo caso de uma câmera que captura imagens e possibilita sua reprodução ou de um papagaio que emite sons, sem, no entanto, ser possível que se admita que a câmera efetivamente enxerga, ou que um papagaio raciocina e fala. Para ele, na verdade, ocorre apenas uma manipulação de símbolos ou sons completamente às escuras. Para confirmar seu entendimento de que não há atividade cognitiva em uma máquina artificial[6], John Searle, segundo Paula Fernanda Patrício de Amorim (2014), se valeu da formatação de um *Gedankenexperiment*, experimento de pensamento comum na filosofia da mente que transporta a perspectiva da terceira para primeira pessoa, a quem caberá experimentar a teoria posta e responder acerca de sua aplicabilidade à uma mente humana com base na sua própria. O ensaio

[5] Uma das três leis clássicas do pensamento, representada na lógica proposicional pela fórmula ¬ (P ∧ ¬P), que estabelece que para qualquer proposição, ou é verdadeira a proposição ou é verdadeira sua negação.
[6] Utiliza-se, aqui, a expressão "máquina artificial" porque Searle acreditava que o cérebro humano ou de outros animais, desde que dotados de intencionalidade e capacidades similares, também se tratava de uma máquina.

ficou conhecido como *argumento do quarto chinês*, e buscava refutar ou invalidar teorias correlacionadas à inteligência artificial. O experimento consiste em uma situação fictícia de imaginar a si mesmo preso em um quarto, e receber um primeiro escrito chinês, que será referenciado como "texto A". O indivíduo ergastulado não teria conhecimento algum da língua e símbolos chineses, não sabendo ou identificá-los ou sequer diferenciá-los de outras línguas asiáticas.

Em seguida, é fornecido ao indivíduo no quarto um segundo conjunto de escritos, igualmente em chinês, chamado de "texto B", junto com um conjunto de instruções de como relacionar o "texto A" e o "texto B", aqui nominado de "instruções AB", este na língua nativa do indivíduo. Por fim, um terceiro montante de escritos, "texto C", também em chinês, igualmente acompanhado de um conjunto de regras no vernáculo pátrio do prisioneiro, que ensinaria a correlacionar os textos A, B e C, o qual será referenciado por "instruções ABC". Esclareça-se que ambos os conjuntos de regras (AB e ABC) apenas possibilitam uma correlação formal, visto que o indivíduo continua sem compreender o significado dos símbolos chineses constantes nos textos A, B e C.

Com base nas instruções AB e ABC, seria possível que o indivíduo conseguisse fornecer de volta determinados símbolos em chinês (*outputs*), em resposta aos símbolos apresentados no "texto C", mesmo sem compreender o que significam tais escritos. Essa é a primeira fase do experimento, circunscrita à situação da pessoa no quarto com os textos ininteligíveis e regras que demonstraram ao indivíduo como correlacionar os símbolos.

Na segunda fase, tem-se a perspectiva das pessoas que estão fora do quarto, terceiros que compreendem e têm classificados os escritos entregues ao indivíduo no quarto. O primeiro conjunto de escritos, "texto A", segundo Searle (1980), seria o script (código ou contexto), o segundo conjunto de escritos, "texto B", a história, e o terceiro conjunto de escritos, "texto C", as perguntas. O que foi devolvido pelo indivíduo, se valendo das instruções ABC, são as respostas às perguntas (texto C). E, por fim, o primeiro e segundo conjunto de regras recebidos, instruções AB e ABC, seriam o programa.

Sem dúvidas, existe uma série de críticas ao argumento do quarto chinês e até mesmo à atecnicidade computacional empregada no experimento, as quais não serão objeto de análise no presente estudo. Assim, se atendo somente à proposição analisada e à crítica realizada por Searle (1980), tem-se estabelecido, na segunda fase, sob a perspectiva das pessoas fora do quarto: um contexto no qual a história se insere, com os detalhes socioculturais e ambientais que envolvem a situação (texto A); a história propriamente dita (texto B) e as questões

formuladas (texto C). Os símbolos fornecidos pelo indivíduo seriam as respostas às perguntas, que teriam sido elaboradas com base nos programas AB e ABC (instruções).

Diante disso, para os terceiros que estão fora do quarto, são entregues perguntas em chinês e devolvidas respostas em chinês, presumindo-se que há compreensão por parte de quem respondeu. De outro lado, a partir da perspectiva em primeira pessoa do indivíduo, o que é feito no quarto nada mais é do que relacionar símbolos desconhecidos, conforme o programa que lhe foi dado, sem que haja nenhuma compreensão do conteúdo contido nos documentos.

Acrescentando uma terceira fase ao experimento, onde os textos A, B e C fossem entregues ao indivíduo preso no quarto na sua língua nativa, acredita-se que as questões seriam corretamente respondidas, independentemente do fornecimento ou utilização das instruções (AB ou ABC). Para as pessoas de fora do quarto, a conclusão será a mesma da segunda fase: que o indivíduo preso compreende a sua língua nativa tanto quanto compreende a língua chinesa. Porém, não é o que realmente acontece. Sabe-se que o indivíduo enclausurado até entende a sintaxe, mas não compreende a semântica. Não há, pois, intencionalidade. Ocorre apenas uma instanciação.

Dito isso, Searle demonstra o propósito do exercício mental do seu argumento, alcançando duas proposições, quais sejam: "intencionalidade em seres humanos (e animais) é um produto de características causais do cérebro. [...] Certos processos cerebrais são suficientes para haver intencionalidade" e "instanciar um programa de computador jamais será, por si só, uma condição suficiente para haver intencionalidade" (SEARLE, 1980, p. 417).

Há de se registrar, neste momento, os objetivos específicos do argumento searleano, que pretendem rechaçar teorias filosóficas correlatas à inteligência artificial, quais sejam: os programas, o computacionalismo, o cognitivismo, o behaviorismo, o funcionalismo e, por fim, o teste de Turing.

O primeiro ponto atacado por Searle são os programas. A questão fundamental dos programas é que eles, essencialmente, tratam de simular comportamentos humanos, mas não de duplicá-los. Mesmo um programa que "passe" no teste de Turing, não seria possível dizer que possui intencionalidade, ou que fez mais do que instanciação a fim de simular uma habilidade cognitiva humana.

O behaviorismo, teoria desenvolvida a partir da análise do comportamento (*behavior*), funda-se precipuamente na premissa de que para que seja científica a teoria, deve ser observável do ponto de vista comportamental. Os estados mentais ou filosóficos poderiam ser explicados a partir dos movimentos observáveis e características físicas, à exemplo dos *inputs* e *outputs*, em se falando de linguagem computacional. O behaviorismo foi uma das grandes bases para o

desenvolvimento da teoria funcionalista. Esta, por sua vez, não ignora aspectos internos dos estados mentais e não se apresenta de uma forma tão reducionista, buscando compreender a relação causal entre os estados mentais, inclusive com relação a outros estados mentais. O foco principal do funcionalista, como sugere o nome da teoria, é o funcionamento, que, aplicado ao estudo da mente, poderia equiparar uma máquina a uma mente humana, caso aquela mantivesse atividades internas idênticas ao cérebro humano, independentemente da matéria que a compõe. Sob essa análise, de que o funcionamento é mais importante que a substância, seria possível afirmar que se algo funciona da mesma forma que uma mente, apesar de possuir estrutura diversa, poder-se-ia afirmar que a mente foi duplicada, e não apenas simulada. A conexão do que chamou de autômato probabilístico com a máquina de Turing é clara, indo além apenas na possibilidade (probabilística) de diversos *outputs* em contraposição com o determinismo de Turing (PUTNAM, 1975).

Para a teoria computacionalista, que tem como principal referência Jerry Fodor (1935 - 2017), filósofo e cientista cognitivo norte-americano, a cognição humana se trataria apenas do cérebro processando dados sob um aspecto meramente formal, sendo irrelevante para tal o seu conteúdo. O raciocínio humano, então, seria uma questão de preservação da verdade em certas ilações, e assim, a mente se equivaleria a um programa de computador e a cognição à computação. Com essas características principais, Fodor já demonstraria, em princípio, seu posicionamento completamente oposto àquele sustentado por Searle, visto que visualiza uma sobreposição da sintaxe em detrimento da semântica. No entanto, o pensamento de ambos é mais próximo do que parece. Para Fodor, a inteligência artificial não se confunde com uma ciência, posto que consiste fundamentalmente na construção de artefatos programados para realizarem determinadas tarefas buscando *simular* a inteligência, se tratando, portanto, de uma engenharia. A ciência cognitiva, de outro lado, busca estudar e *compreender* o pensamento, seu objeto de estudo (AMORIM, 2014).

O cognitivismo, terceiro alvo do argumento do quarto chinês, em uma linha próxima da teoria computacionalista, sustenta que a manipulação de símbolos se equivaleria a pensar (ou seja, com intencionalidade). O teste de Turing comete, para Searle, um grande equívoco ao confundir epistemologia com ontologia. E isso fica claro quando modificado o epicentro do exercício de pensamento da terceira para a primeira pessoa. Para as pessoas de fora do quarto, o sistema (quarto) compreende chinês. No entanto, em uma análise mais acurada, a partir da perspectiva em primeira pessoa do indivíduo dentro do quarto, é alcançado, com o exercício proposto, que não há compreensão semântica sobre chinês, não podendo, assim, se verificar intencionalidade, e, portanto, uma inteligência artificial (forte).

A crítica de John Searle (1980) com o argumento do quarto chinês contrapõe a aplicação das supracitadas teorias à inteligência artificial, posto que, filosoficamente, entende que a instanciação, replicação de comportamento, funcionamento, manipulação de símbolos não importam na *duplicação* da cognição humana, a ponto de se alcançar uma inteligência artificial forte; E é exatamente isso que busca demonstrar com seu estudo: que pode haver instanciação sem que haja intencionalidade, e que a realização daquela não importa na compreensão ou consciência do seu agente, razão pela qual não pode ser considerada inteligente uma máquina artificial que apenas faz análise sintática, sem demonstrar qualquer compreensão da semântica. Finaliza, pois, de modo a não cerrar a possibilidade de concepção posterior de uma inteligência artificial, que seria atingida por qualquer mecanismo que conseguisse reproduzir a intencionalidade, ou seja, desenvolvesse a capacidade de compreensão semântica de seus atos, para, assim, ter poderes equiparado a um cérebro (SEARLE, 1980, p. 417).

A intencionalidade seria, por conseguinte, a propriedade que diferencia os seres dotados de inteligência, pois caracteriza seu estado mental. Ter a ciência do que se fala, não apenas emitir sons sem significação, sem conteúdo. Todos os pensamentos possuem substância e indicam coisas ou situações do mundo, mesmo que sejam apenas produto da imaginação (TEIXEIRA, 2014, p. 51).

Nesse contexto, surge o programa de Schank e Abelson (1977), criado para emular a habilidade humana de compreensão de histórias e deduzir informações adicionais não fornecidas ao programa, que alcançaria esta referida intencionalidade, não apenas simulando, mas duplicando a habilidade humana. Analisando-o, Searle apresenta um exemplo onde seriam dadas informações relativas ao modo de comportamento social do homem médio, à concepção padrão de restaurantes e à classificação de boas refeições, assim como outros elementos necessários para inferir se um alimento foi servido adequadamente e deveria ser pago. O programa deveria cruzar as informações recebidas e indicar qual o comportamento mais provável deve seguir o indivíduo, dentre as opções A ou B, em uma espécie de análise preditiva. A questão repousa em efetivamente haver compreensão da história e não apenas em formular respostas (SEARLE, 1980, p. 418).

No mesmo sentido, Roger Penrose, físico, matemático e filósofo da ciência, também apresentou uma objeção ao que se entendia por inteligência artificial, que veio a se tornar das mais conhecidas, quando apresentou o argumento do *insight*, no seu livro intitulado *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics* (PENROSE, 1991). No seu estudo, sustenta que com a ciência que lhe era contemporânea, não seria possível conceber uma inteligência artificial que pudesse ter um *insight*. Ela até poderia gerar

informações novas a partir do cruzamento dos dados de sua base, mas, tomando o *insight* como um momento "eureka", este seria um privilégio humano que a máquina não pôde ainda alcançar (TEIXEIRA, 2014, p. 52).

Nesse particular, rememora-se as conhecidas alegoria do Sol e alegoria da Linha, de Platão, em sua obra A República (2019), visto que os diálogos que descrevem ensinamentos de Sócrates por seu discípulo retratam exatamente a ideia de consciência e intelecto que ora se cogita ser alcançada ou não pela inteligência artificial.

Sócrates, filósofo ateniense que viveu no período clássico da Grécia antiga, concentrava a grande maioria de seus ensinamentos na palavra, ou seja, pela fala, posto que, segundo ele, a escrita fechava o conhecimento engessando o autor à exatidão das afirmações estáticas que escrevera. Desta forma, se contivesse algum erro nas afirmações, estes, além de perpetuados, seriam retransmitidos indefinidamente. Além disso, não se auto proclamava sábio. Ao contrário, sua célebre frase afirma que nada sabia. Assim, as fontes do conhecimento transmitido de Sócrates advêm principalmente de três alunos contemporâneos seus, Platão e Xenofonte, por meio das obras literárias marcadas pelos diálogos cujo personagem central era Sócrates, e Aristófanes, por meio de suas peças teatrais. Sócrates, então, é conhecido por relatos que fizeram a seu respeito, se apresentando os diálogos de Platão como aqueles mais abrangentes, razão pela qual popularizou-se a referência a *Sócrates de Platão*.

Não por outra razão as obras são nominadas de diálogos. A forma de representação dos textos consiste na reprodução de conversas e passagens de Sócrates com outras personalidades contemporâneas, sempre se dedicando ao ofício que ele considerava mais importante: a *maiêutica*, o parto das ideias. Sócrates teve origem humilde, sem acesso ao caro sistema de educação da época, filho de pai artesão, que esculpia colunas nos templos, e mãe parteira. Em uma das passagens que se tem relato, auxiliando sua mãe em um parto complicado, se apercebeu que o que fazia na busca da verdade, por meio do questionamento aos interlocutores, era assemelhado aos partos realizados por sua mãe, já que não era dela o filho, tampouco era ela que iria criar a criança. Tinha, contudo, a missão de realizar o parto, sob pena de vir a falecer a criança e a parturiente. Com essa analogia, conseguiu compreender sua vocação, que seria dar luz às ideias das pessoas, sempre colocando em xeque as concepções e teorias formadas a respeito de algum assunto, conduzindo-os a outras perspectivas sobre a temática. Por isso seu ofício ficou conhecido como *maiêutica*, que significa *parteira* em grego. Sócrates acreditava na busca da verdade, e que a verdade só seria conhecida pelo uso da razão. Essa é a essência do método socrático, que consiste, basicamente, na constante investigação por meio até de simplórios questionamentos, a fim de revelar eventuais contradições, bem como instigar o

interlocutor a redefinir seus valores e preconceitos por si próprio, e não por influência da sociedade.

Por certo, Sócrates não tinha alcançado a possibilidade de reprodução artificial da mente humana, mas seus ensinamentos, e, mais que isso, suas provocações, demonstram uma profundidade na análise do intelecto humano, que justificam a rápida digressão apresentada. Pelas razões acima expostas, diante da existência de alguma controvérsia na fidelidade das ideias socráticas constantes nos diálogos produzidos por Platão, em que pese seja Sócrates o personagem central da narrativa, ao se tratar das alegorias e dos significados pretendidos, será sempre referenciado o autor, responsável direto pela obra.

A mais famosa e segunda mais extensa das obras de Platão, com dez volumes, é iniciada com a narrativa do ocorrido no dia anterior, em que visitaram Pireu, um porto situado próximo a Atenas. O diálogo vai se desenvolvendo até chegar ao ponto central da obra, que é a busca pela definição do que seria justiça. Nesse espeque, fluem os diálogos, sempre conduzidos por Sócrates, até que se entende pela necessidade de uma visão holística da sociedade para a compreensão do que se pretende, pelo que é proposta a criação de uma cidade ideal, chamada de *Kallipólis*, onde seria possível a análise entre o ser humano e o seu meio social, percorrendo desde o entendimento do que seria o injusto, a temas ligados à epistemologia, metafísica psicologia, dentre outros. Nesse contexto, no volume VI d'A República, debutam a alegoria do Sol e alegoria da Linha, estórias sugeridas por Platão com o objetivo de despertar pontos de vistas e novas leituras sobre o intelecto humano, temática que particularmente nos importa, especialmente a última.

A alegoria do Sol apresenta a visão como o sentido mais desenvolvido do ser humano, correlacionando-o com o conhecimento. No plano sensível, diz que além do próprio sentido da visão e do objeto a ser visualizado, faz-se necessário um terceiro elemento para que seja possível que o sistema funcione, no caso, o Sol. O Sol, porque fornece a luminosidade necessária para que os olhos possam enxergar o objeto, o qual, sem isso, seria invisível ao olho humano. No plano inteligível proposto pela metáfora, o Sol seria o Bem que iluminaria o conhecimento (objeto), fornecendo a verdade (luz) ao sujeito cognoscente, que lhe permite a faculdade e exercício da razão, tornando-o apto a conhecer (enxergar), sem que a ideia do Bem integre o objeto do conhecimento, posto que estaria acima disso (PLATÃO, 2019, p. 313).

A alegoria da Linha, em complemento à do Sol, prossegue na divisão entre visível (sensível) e inteligível. Sugere que um pedaço de uma linha seja dividido ao meio, e colocadas lado a lado. A primeira parte representaria o mundo sensível, e a segunda metade da linha o mundo inteligível. Cada metade, então, deveria ser mais uma vez dividida ao meio, resultando

em quatro segmentos da linha, sendo os dois primeiros do mundo sensível e os dois últimos do mundo inteligível, os quais representariam os níveis de cognição, consistentes em *imagem*, *objetos sensíveis*, *razão* e *consciência*, em ordem cognitiva crescente. O primeiro deles, a *imagem*, está inserida no campo do visível, e diz respeito aos reflexos no espelho, reflexos nas águas e às sombras dos objetos. Estes, os *objetos sensíveis*, são a representação material de todas as coisas que vemos na natureza, tais como árvores, plantas, pedras etc. Na outra metade da linha, que trata do mundo inteligível, tem-se em primeiro lugar a *razão*, local em que ocorre o entendimento, seguido, no último quarto de Linha, das ideias, Formas, e do belo, que representam a *consciência*. Essa última camada de cognição não se alimenta diretamente de imagens ou objetos sensíveis advindos do mundo visível, mas sim de ideia para ideia, dentro do mundo inteligível, materializando o pensamento dialético.

> Então entende também que pela outra subseção do inteligível entendo aquilo que a própria razão aprende por meio do poder da dialética, não considerando essas hipóteses como primeiros princípios, mas literalmente como hipóteses, pontos de apoio para o impulso que servem para o ponto de partida que permita alcançar o primeiro princípio *não hipotético* de tudo. Uma vez apreendido esse princípio, ela reverte a si mesma, e, retendo o que disso resulta, aporta a uma conclusão sem fazer qualquer uso do sensível, mas somente das próprias *Formas* (ideias), movendo-se de Ideias para Ideias e terminando em Ideias. (PLATÃO, 2019, p. 320)

Com essa proposição, Platão, por meio da figura de Sócrates, correlaciona os quatro segmentos apresentados com as quatro operações da alma sugeridas, em ordem crescente de profundidade cognitiva: suposição (imagem), fé (objetos), entendimento (razão) e inteligência (consciência).

A apresentação da inteligência por Sócrates de Platão é instigante quando se cogita a possibilidade de sua replicação de forma artificial. No decurso de sua obra, Platão demonstra, com fortes argumentos, a necessidade da participação do homem filósofo na gestão da cidade, que era naquele contexto a representação da sociedade. Revela que somente uma alma filósofa, liberta das amarras, poderia ascender do mundo sensível para o mundo inteligível e lançar luzes aos demais acerca da ilusão que se apresenta por meio das imagens, aparências e sombras da realidade vistas. Poderia, então, a máquina artificial alçar esse campo de intelecção do mundo inteligível, para que fosse considerada propriamente inteligente?

Não se pode negar o alto poder dos sistemas de inteligência artificial, os quais possuem acurácia muito superior à humana e continuam em exponencial evolução, apesar de restritos a tarefas específicas, por enquanto. No entanto, sob o aspecto filosófico, muito bem desenhado por Searle e mais abstratamente por Platão, surge o questionamento se a consciência é a essência humana? Ou, ainda, se poderá a máquina reproduzir o intelecto humano? Por certo, a partir de

uma visão racionalista, o diferencial humano é a sua capacidade de imposição de suas vontades em relação à natureza (TOCCHETTO; GRUBBA, 2018), ou seja, o agir com intencionalidade. Para que sejam adequadamente abordadas essas reflexões não devem ser consideradas as essências das leituras behavioristas, funcionalistas ou computacionalistas, que, em geral, correlacionam a existência de inteligência artificial à *mimetização* do comportamento humano, da própria mente humana ou do que se entende por inteligência, sob pena de se encerrar o diálogo prematuramente. Sem dúvidas, ainda que alcançada a chamada inteligência artificial forte, ou até mesmo a superinteligência (singularidade), essa permanecerá como uma provocação retórica, que nunca encontrará consenso, visto que, a depender das bases filosóficas do sujeito, da ausência delas ou do que ele entende por inteligência, poderá se inclinar para uma resposta positiva ou negativa, sendo importante ressaltar que a unanimidade não é necessariamente o objetivo da discussão, mas sim a própria discussão em si.

Ressalte-se que essa análise filosófica abstrata e generalista não afasta a possibilidade de que em algum momento seja atribuída inteligência (intencionalidade, compreensão) à uma máquina artificial. Contudo, apenas será inteligente, filosoficamente falando, aquela máquina que tiver capacidade de desenvolver intencionalidade, para além da instanciação, momento em que duplicaria os poderes causais do cérebro humano, que passaria a ser dispensável, alcançando o que se entende por inteligência artificial forte.

Existem duas acepções de inteligência artificial: inteligência artificial fraca (*narrow*, limitada, restrita, estreita ou superficial) e inteligência artificial forte (geral ou *strong*). A primeira delas é vista quando a máquina artificial, por meio de programas computacionais, pretende compreender as habilidades e capacidades da mente humana, simulando-as em um ambiente real ou virtual. Assim, a inteligência artificial fraca pode agir *como se fosse* inteligente, agir *como se tivesse* mente, mas não pode se dizer que seja propriamente inteligente. Trata-se de uma emulação do comportamento da mente humana, que não tem raciocínio e nem vontade. A máquina artificial, nesta concepção, sempre seria orientada por um programa que foi necessariamente desenvolvido em linhas gerais por um humano (SEARLE, 1997).

É importante que se diga que a inteligência artificial fraca representa o tipo de inteligência efetivamente desenvolvida pelo homem até o presente momento. São sistemas que possuem propósitos específicos, que foram criados para solucionar um determinado problema, ou uns determinados problemas, mas são incapazes de resolver problemas de outras áreas de forma autônoma. É o exemplo do já citado *deep blue*, sistema de inteligência artificial que foi criado para desafiar o enxadrista Garry Kasparov. Apesar de fundar-se em força bruta, e ter sido ensinado a emular todas as possibilidades existentes de movimentos a ponto de vencer o

campeão mundial do jogo, caso fosse designado para jogar damas ou pegar um copo com água, falharia miseravelmente. Os sistemas inteligentes desenvolvidos até os dias atuais, dotados de menor ou maior complexidade, se ocupam de tarefas específicas, que na maioria dos casos são resolvidos com base nos comandos previamente estabelecidos pelo homem.

Demandam, necessariamente de que seja projetado um algoritmo que estabeleça como deve processar os dados recebidos (*inputs*), permitindo a entrega de resultados pretendidos (*outputs*), o que os torna supostamente inteligentes. No entanto, a bem da verdade, não há propriamente uma atividade cognitiva que seja desempenhada pelo sistema artificial. Ela não compreende o processamento de dados que faz. Na verdade, a inteligência artificial pode ser programada por meio dos algoritmos, para rotular dados, executar tarefas, e, em um grau mais elevado, ser programada para que, além de reagir a situações novas, possa adquirir novos conhecimentos e otimizar sua própria programação.

A explicação mais rasa, básica e funcional utilizada para ilustrar a essência dos algoritmos é compará-lo a uma receita culinária em um nível de detalhes muito mais descritivos, vez que deverão ser explicitados como são os ingredientes a serem identificados pelo sistema, caso sejam utilizados dados não rotulados, por exemplo. Além disso, todas as informações relativas ao modo de preparo, manuseio, ordem das ações, devem ser analiticamente descritas, sob pena de não se alcançar o resultado esperado. Com tal programa apresentado ao sistema (algoritmo) os dados se apresentarão como os ingredientes (*inputs*) a serem tratados, para que se alcance o prato final (*outputs*). A captura de novos dados a partir dos *outputs* que sirvam para retroalimentar o sistema, pode ao fim e ao cabo, melhorar o programa, permitindo o ajuste fino com base no que a inteligência artificial conseguir constatar como desalinhado com o resultado esperado.

De outro lado, a inteligência artificial forte é aquela comumente vislumbrada pelo inconsciente popular, semeada por obras ficcionais e pela indústria cinematográfica, em uma espécie de polinização cruzada, na medida em que cria fantasiosamente proposições, que acabam vindo a chamar atenção de estudiosos e cientistas que passam a desenvolver as tecnologias imaginadas, que servem de subsídio para novas obras ficcionais. Esse tipo de inteligência artificial não se limita a resolver tarefas específicas, imitar a capacidade humana, ou mesmo lhe servir como ferramenta ou instrumento, mas sim de espelhar a mente humana a ponto de substituí-la completamente pelos programas computacionais. Seria capaz de compreender e emular estados cognitivos, se apresentando como um sistema complexo que teria a capacidade de abstrair informações que julgasse desnecessárias e tomar decisões com intencionalidade. Aplicando a inteligência artificial a diversas áreas distintas e de maneira

autônoma, ela viria facilmente a superar a capacidade humana. Para a inteligência artificial forte, a mente humana estaria para o cérebro assim como o software está para o hardware. Segundo Searle (1997, p. 26), não existe nada propriamente biológico na mente humana que a impossibilite de ser substituída perfeitamente e por inteiro pelo programa certo, instalado no hardware certo. Seria, pois, um sistema que poderia ser chamado de consciente.

Ressaltando-se, mais uma vez, que quando se trata de inteligência artificial em qualquer de suas modalidades, por assim dizer, não se está vinculado à ideia do humanoide, ou seja, robô que busca imitar a forma humana. É possível sim que seja desenvolvido um hardware que dê suporte a um sistema tal, que imite visualmente a aparência dos humanos. Mas, o que importa essencialmente é o software, é o programa instalado em qualquer que seja o "corpo físico", desde que adequado e capaz para recebê-lo, podendo, assim, a inteligência artificial nem mesmo possuir representação física, já que se trata de um programa que na maior das vezes somente terá uma estrutura material apta a entregar os *outputs* pretendidos.

Portanto, em um campo menos filosófico e mais pragmático, fica claro que não são visões estanques, mas concepções diferentes onde tem-se, de um lado, a inteligência artificial forte que seria aquela capaz de duplicar e substituir completamente o raciocínio humano em diferentes situações, com competência ampla em todas as áreas, enquanto os sistemas baseados em inteligência artificial fraca se ocupariam de tarefas específicas e predeterminadas. Longe de se acreditar que não há interesse e desenvolvimento significativo em torno da inteligência artificial geral, é certo que a crescente evolução da complexidade da inteligência artificial específica em diversos sentidos, como por exemplo no reconhecimento de discurso, aprendizagem e até raciocínio propriamente dito, importam na solução de subproblemas que quando combinados, certamente contribuirão para o atingimento de uma inteligência artificial geral.

Especula-se, ainda, acerca de um terceiro tipo, a superinteligência, onde a inteligência artificial ascenderia a um outro patamar, em que superaria exponencialmente a inteligência humana, alcançando o que se convencionou chamar de *singularidade*, palavra que representa a ideia de uma "explosão de inteligência", produto do auto desenvolvimento não supervisionado e constante dos sistemas de inteligência artificial, a fim de criarem sistemas cada vez melhores (RUSSEL, 2016). Contudo, considerando o estado da arte em que a tecnologia se encontra, não se tem conhecimento do desenvolvimento de uma inteligência artificial forte, quiçá de superinteligência, inobstante os discursos alarmistas e apocalípticos propagados (BOSTROM, 2018). É destacado, outrossim, uma projeção de futuro no campo de inteligência artificial mais coerente e razoável, ligada ao desenvolvimento de aplicações úteis à humanidade, como o uso

de *machine learning* em larga escala, *deep learning*, aprendizado por reforço, robótica, visão computacional, processamento de linguagem natural, sistemas colaborativos, *crowdsourcing* e computação humana, internet das coisas, escolha computacional social, teoria dos jogos algorítmica e computação neuromórfica (HARTMAN PEIXOTO; SILVA, 2019, p. 81), campos os quais têm evoluído significativamente, sobretudo em razão da trinca de fatores a seguir apresentados.

1.3 PONTO DE INFLEXÃO DA INTELIGÊNCIA ARTIFICIAL: EXPLOSÃO DE DADOS (*BIG DATA*), AUMENTO DA CAPACIDADE COMPUTACIONAL E EVOLUÇÃO DOS ALGORITMOS

A explosão de dados produzidos, o aumento da capacidade computacional e a evolução dos algoritmos foram os principais motores de propulsão da ruptura que tem sido considerada a quarta revolução industrial. Isso se deu, fundamentalmente, pela confluência desses fatores, que se complementam entre si. O avanço computacional e da tecnologia como um todo, permitiu o desenvolvimento de artefatos eletrônicos que, funcionando como sensores, passaram a acumular bilhões de dados sobre as pessoas (MAGRANI, 2018, p. 49). Esse avanço das novas tecnologias atua de duas formas: uma delas clara e explícita, quando seduz os indivíduos para que forneçam deliberada e voluntariamente seus dados em troca de pequenas facilidades no seu dia-a-dia, inspirando confiança nas aplicações (CARINI; MORAIS, 2019), como aferição de batimentos cardíacos e monitoramento de exercícios físicos por pulseiras ou relógios eletrônicos, acesso a operações bancárias etc.; ou de forma velada, onde a captura de informações ocorre de maneira imperceptível aos indivíduos, como se nota no estabelecimento de perfil do usuário em plataformas de *streaming*, que captam suas preferências com base no conteúdo consumido, nas redes sociais que, alimentadas pelos rastros de pesquisas realizadas, oferece produtos de interesse do usuário, ou, ainda, monitora a sua própria experiência a fim de lhe proporcionar mais conteúdo personalizado (MAGRANI, 2018, p. 49).

Um usuário de uma rede social que recorrentemente acompanha as postagens de determinada conta, terá esse perfil sempre apresentado nas primeiras posições, quando acessar a sua plataforma. E esse acompanhamento se dá de várias formas, não apenas pela “curtida” na postagem, mas desde a aferição do tempo que o usuário passou lendo um texto, vendo uma foto, assistindo a um vídeo, a ampliação da tela por aplicação de *zoom*, ou mesmo onde o cursor do mouse “transitou” na postagem. Até o cruzamento de dados relativos ao horário frequente de navegação na internet e redes sociais conforme o conteúdo buscado é realizado, permitindo que

o algoritmo realize uma apresentação dirigida de conteúdo sobre o que aprendeu a respeito do indivíduo, que pela manhã costuma ler jornais, pela tarde informações esportivas e à noite receitas culinárias, por exemplo, que passará a ter mais notícias em seu *feed* pela manhã, conteúdo esportivo pela tarde e publicidade sugerida de gastronomia pela noite. A troca de informações de usuários entre plataformas pode fazer surgir um anúncio de determinado produto ou serviço em uma rede social que havia sido pesquisado há poucos minutos em um navegador de internet, ou que o usuário mencionou ter interesse em uma conversa de áudio ou texto. Não sem razão, pode se dizer que mais do que consumir informação, quando se está conectado na rede mundial de computadores, o usuário é, na verdade, o produto consumido, que se apresenta como personagem passivo da situação, vez que tem os seus dados de experiência capturados e utilizados pelos algoritmos para lhe condicionar a viver em uma bolha (AMARAL; BOFF, 2018).

Casas inteligentes que possuem assistente pessoal para gerir o sistema de som, segurança, iluminação, aparelhos domésticos como televisão, geladeira, ar-condicionado e uma série de outros utensílios conectados já não são mais objetos exclusivos de ficção cinematográfica. A internet das coisas é resultado da hiperconexão e interatividade dos objetos à internet, que permitiu uma mudança de cenário onde os dados eram "passivos", ou seja, eram produzidos de forma não indexada e que não tinham extraídas suas melhores aplicações, para um contexto em que os dados passaram a conduzir a operação de forma ativa (KAUFMAN, 2019, p. 26). A hiperconectividade de objetos já vem sendo construída desde a década de 90, quando Bill Joy, cientista da computação norte americano, cofundador da Sun Microsystems, imaginava a possibilidade de realizar a conexão de dispositivo para dispositivo (*device to device - D2D*), e em 1999, Kevin Ashton, pesquisador britânico, cunhou a expressão *internet das coisas* (*internet of things - IoT*), sustentando, acertadamente, que o seu uso seria útil para a economia de recursos naturais e energéticos, além das facilidades pessoais e de saúde já existentes de forma embrionária naquela oportunidade. Assim é iniciado um novo mundo de oportunidade com a possibilidade de máquinas (hardwares) se conectarem diretamente, trocando comandos entre si para a execução de tarefas (REZER; FORTES, 2018). O ascendente desenvolvimento tem sido possível principalmente em razão de tecnologias como *wi-fi*, *bluetooth* e radiofrequência (MAGRANI, 2019, p. 30).

Os *smartphones* são a releitura moderna dos diários – não mais secretos – de outrora. Informações pessoais em notas, conversas por aplicativos, fotos, vídeos, dados sobre atividades físicas, sono, compromissos, agenda, interesses futuros, perfil consumidor, detalhes de relações interpessoais. Atualmente tudo isso está exposto, voluntariamente, pelos próprios usuários, para

que tenham facilidade de tomar notas em um bloco eletrônico; evitar ligações por chamada de voz ou encontros presenciais; ou o monitoramento do seu bem estar ao alcance de sua mão. Apesar da imprecisa sensação de segurança causada em razão da utilização de senha de acesso ao dispositivo por identificação biométrica, pela identidade digital ou até mesmo por reconhecimento facial, tem-se, ao contrário, mais e mais dados sensíveis que podem ser expostos, entregues voluntariamente pelo usuário às empresas e governos.

Essa enxurrada de dados cada vez mais estruturados, que foi chamada de *big data*, aprimorou a possibilidade de analisar e gerar grandes quantidades de informações sobre um tópico específico, sem limitação de amostragem, com um prestígio maior por correlações ao invés de causalidades, que outrora resultou em abrir mão da exatidão, para visualizar a tendência geral (KAUFMAN, 2019, p. 33). Se por um lado esse novo posicionamento a respeito da utilização dos dados de forma ativa ainda não foi completamente percebido pelas pessoas, de outro lado, são crescentes os avanços comerciais a esse respeito nos últimos anos, com um uso intensivo de algoritmos de inteligência artificial por parte de empresas, formando o que hoje é chamado de sociedade da informação e sociedade do conhecimento (ITO, 2020, p. 24). A bandeira de cartão de crédito Mastercard, há alguns anos, teve um de seus executivos abrindo uma palestra com a fala de que a multinacional não era mais uma empresa financeira, mas sim uma empresa de dados. A divisão *Mastercard Advisors* foi criada para analisar sessenta e cinco bilhões de transações de um bilhão e meio de titulares de cartões em mais de duzentos países, buscando a identificação de padrões para que as informações fossem vendidas para terceiros (KAUFMAN, 2019, p. 58).

Conhecendo mais os hábitos dos indivíduos, é possível personalizar e customizar automaticamente os conteúdos nas plataformas digitais que serão para ele dirigidos ampliando as chances de aquisição dos produtos ofertados, e favorecendo determinados nichos de mercado a se expandirem, explorando a teoria da cauda longa[7]. De certo, existem diversos pontos negativos no uso imoderado da internet das coisas, que vão desde um extrapolo no desenvolvimento de coisas inutilmente conectadas (MAGRANI, 2018, p. 51), o incremento na produção de lixo advindo do descarte de produtos obsoletos (*e-waste*), aumento no preço dos dispositivos conectados face àqueles não conectados, maior vulnerabilidade em razão da exposição de dados captados pelos sensores na rede, dentre outros. De igual maneira, o

[7] A teoria da cauda longa é utilizada em estatística para referenciar uma projeção de volume de dados decrescente. Fatores como baixo custo de armazenagem e distribuição permitem a negociação de uma grande variedade de produtos para uma grande variedade de pessoas, deixando de restringir o maior enfoque aos produtos mais procurados para obtenção de maior lucro. Ganhou grande notoriedade a partir do artigo homônimo de Chris Anderson, publicado na revista Wired, em outubro de 2004.

direcionamento de nicho de mercado pode, independentemente de ser assertivo ou não, compartimentalizar as preferências e percepções dos usuários, já que lhe distanciaria de visões, opiniões e posicionamentos contrários aos seus, fomentando uma sociedade de hábitos e ideias bitoladas, no que ficou conhecido como *filtro bolha*.

Contudo, mesmo diante do viés negativo acima citado, houve uma grande aposta no potencial da implementação de tecnologia em objetos do cotidiano (BLUM, 2019), a qual, somada com pesadas estratégias de marketing, convencem a aquisição de produtos que se mostram úteis em um primeiro momento, mas, com o uso efetivo é percebida sua inutilidade, criando, segundo Jenny Judge (2015, p. 2), uma escravidão tecnológica.

> Mas, mesmo que as empresas de tecnologia não estejam realmente tentando nos escravizar, ou nos fazer sentir inadequados, isso não significa que a situação atual seja um caso de boas intenções que deram errado. Não há maior razão para pensar que a tecnologia é intrinsecamente boa, mas ocasionalmente dá errado, do que há para pensar que ela é uma vilã extremamente bem sucedida. [...] Nós amamos elogiar a tecnologia, e nós amamos condená-la. Nós a equiparamos ao caos, ao poder, ao amor, ao ódio; à democracia, à tirania, ao progresso e à regressão – nós a louvamos como nossa salvação enquanto a lamentamos como nosso flagelo. Como qualquer tecnologia que veio anteriormente, a tecnologia digital é tudo isso. Mas não é essencialmente nada disso[8].

Há de se obtemperar, outrossim, que apesar da manifesta infinidade de benefícios proporcionados pelos avanços tecnológicos e da sua inevitabilidade, é certo que nem todos os avanços importam em resultados positivos para a humanidade.

De seu turno, interessante a menção de que há, também, um conceito de *Internet de todas as coisas* ou *Internet de tudo* (*Internet of Everything - IoE*), utilizado pela empresa Qualcomm com emprego semelhante ao da internet das coisas, embora a empresa Cisco sustente que a internet das coisas seria um estágio preliminar à internet de tudo (WEISSBERGER, 2014).

Neste momento, é importante esclarecer que esse *boom* de dados, precisou ser adequadamente trabalhado para que fossem geradas informações, e, por conseguinte, conhecimento. O entendimento dessa compreensão importa, visto que dados são considerados os elementos puros, sem qualquer tratamento, que são produzidos em quantidades absurdas nos dias atuais. Com a obtenção deste dado, sua inserção em um contexto para que seja possível

[8] Texto original: *But even if tech companies aren't really trying to enslave us, or make us feel inadequate, that doesn't mean that the current situation is a case of good intentions gone awry. There's no more reason to think that tech is intrinsically good, but occasionally getting it wrong, than there is to think that it's a remarkably successful villain. […] We love to praise tech, and we love do condemn it. We equate it with chaos, power, love, hate; with democracy, with tyranny, with progress and regress – we laud it as our salvation, while lamenting it as our scourge. Like and technology that has come before it, digital technology is all of these things. But it's essentially none of them*.

sua análise, tem-se uma informação. E a partir de uma informação, é possível o estabelecimento de um "modelo mental que descreva o objeto e indique as ações a implementar, as decisões a tomar" (REZENDE, 2003, p. 5), que é o que se entende por conhecimento.

Assim, para além da enorme produção de dados, o seu adequado tratamento foi fundamental para que fossem geradas inúmeras informações, que combinadas e transformadas, permitiram a criação de conhecimento a partir do raciocínio algorítmico (ROVER, 2001). A distinção entre dados, informação, conhecimento e a forma que a transformação de um para o outro ocorre, importa à percepção de que os sistemas artificiais inteligentes, assim considerados aqueles que são aptos a compreender, analisar e sintetizar informações para a tomada de decisão, estariam *raciocinando* quando, por meio do uso de algoritmos cada vez mais sofisticados, se prestarem a produzir conhecimento na resolução de problemas, cumprir tarefas, bem como quando se aproveitarem de associações e referências cruzadas para solucionarem desafios complexos, num contexto nunca antes imaginado por John Searle.

Esclareça-se que dados por si só não representam uma informação. É necessário que haja uma organização lógica para transformar os dados em informação (ITO, 2020, p. 21). Para isso, o *big data* e a internet das coisas foram forças motrizes necessárias para o entendimento do aperfeiçoamento dos algoritmos pelo homem e por eles mesmos. Uma base de dados repleta de conteúdo, de todos os mais variados matizes, devidamente estruturada e apta a produzir informações e gerar conhecimento, é o campo ideal para o desenvolvimento de algoritmos cada vez mais sofisticados, exatamente em razão da abundância de dados, que propicia uma rica gama de possibilidades de cruzamento de informações, assim como de um *database* de teste, para o treinamento do algoritmo.

Tecnicizando a definição retratada anteriormente por uma simples receita de bolo, para Ethem Alpaydin (KAUFMAN, 2019, p. 35) algoritmo é uma sequência de instruções que são realizadas para transformar a entrada (*input*) na saída (*output*). Essa alusão à receita culinária, diga-se, é rechaçada por Pedro Domingos (2017, p. 26), para quem o algoritmo deve ser muito mais preciso do que pode ser a receita, como já dito, por ter que descrever analiticamente tudo o que deverá ser realizado pelo computador, desde a identificação dos ingredientes, à indicação do que seria "uma pitada" ou "fogo alto". A programação algorítmica demanda o detalhamento de todas essas informações previamente para que o computador tenha o conhecimento dos *metadados* (dados sobre os dados) para executar corretamente a tarefa.

Os professores Fabiano Hartman Peixoto e Roberta Zumblick Martins da Silva (2019, p. 72), trazem à lume, nesse particular, a descrição mais detalhada de Horowitz, que aponta características essenciais para concepção do algoritmo. Em linhas gerais, seria um conjunto

finito de instruções que, seguidas, realizam uma tarefa específica, com cinco características necessárias, quais sejam: 1) *input*: fornecido externamente; 2) *outputs*: quantificável produzido; 3) *definiteness*: clareza nas instruções; 4) *finiteness*: o encerramento com um número de etapas finitos; e 5) *effectiveness*: com instruções executáveis por pessoas.

A evolução algorítmica substituiu o inicial modelo de posicionamento preciso de pinos em imensos painéis para carregamento de determinado programa, para, em primeiro lugar, o que se pode resumir como o uso de cartões perfurados carregados na sua memória que, ao depois, foi substituído exclusiva e integralmente pelo software (PEREIRA, 2019). Superadas as limitações físicas iniciais, e até de concepção e organização da escrita de algoritmos, passa-se à sua segunda fase, em que os programadores, que tinham como missão antever todas as contingencialidades futuras possíveis, e escrever a cadeia de comando a ser ensinada ao computador, começam a ser substituídos pelo aprendizado de máquina (*machine learning*).

Assim são substituídos os algoritmos clássicos, integralmente desenhados pelo programador, pelos algoritmos aprendizes, que, assim como um bebê humano, vêm dotados de uma capacidade incrível e genérica de construir modos de reação frente a situações inusitadas e não programadas (MACHADO, 2011, p. 211). Os algoritmos aprendizes são auto programados para evoluírem com base nas suas experiências, e, mesmo sem uma diretriz básica sobre como solucionar um determinado problema, conseguem, a partir do universo de dados que lhe são apresentados, extrair padrões, regularidades, conexões sistemáticas que implicam em conhecimento. Isso na ordem regular ou inversa, sendo alimentado seja com *inputs* para gerar *outputs*, seja com *outputs* para se chegar aos *inputs*, realiza-se o treinamento algorítmico a fim de outorgar capacidade ao sistema de gerar os resultados pretendidos diante dos novos dados que se apresentem.

À capacidade de aprendizado de máquina é dado o nome de *machine learning*, que apesar poder parecer, a primeira vista, ser um novo ramo da inteligência artificial, sempre foi, em verdade, um dos pontos centrais do desenvolvimento da inteligência artificial, desde os primeiros artigos de Turing, em 1950 (HARTMAN PEIXOTO; SILVA, 2019, p. 88). Seu conceito é definido pela sua capacidade de aprendizado por si só, que lhe possibilita a identificação de padrões, construção de proposições e elaboração de previsões sem regras e modelos pré-programados (MAINI; SABRI, 2017).

Buscando explanar o funcionamento de uma inteligência artificial dotada de *machine learning*, Tom Mitchell (2017, p. 3) identifica os elementos necessários ao processamento, quais sejam: fonte de experiência, classe de tarefas e medida de performance a ser desenvolvida. Assim, tem-se que bem delineados os elementos, o programa poderá melhorar a *medida de sua*

performance na execução da *classe de tarefas* por meio da *experiência* obtida sozinho. Aplicando-se a um exemplo, podemos citar o caso do robô Leo (SCHUITEMA, 2012), que passou pelo processo de aprendizado de andar. Inicialmente, lhe foi fornecida uma programação manual pelo desenvolvedor, que quando executada pelo robô cumpriu o objetivo corretamente em cinco minutos. Na sequência, em um novo teste, desta vez sem qualquer diretriz inserida, ao robô foi dado o objetivo de andar, sem que tivesse sido programado detalhadamente como fazê-lo. A princípio ele não conseguiu se sustentar de pé, vindo a cair por diversas vezes até começar a alcançar o equilíbrio para movimentar suas pernas para que caminhasse, inclinando o corpo na medida certa para se manter de pé. Após diversas e sucessivas quedas, conseguiu aperfeiçoar o movimento, progredindo no seu aprendizado, até que, quatro horas depois, conseguiu caminhar corretamente sem cair.

Uma série de algoritmos de *machine learning* tem dado suporte aos mais variados tipos de atividades básicas do cotidiano atual, sem que sejam notados. Tem-se, por exemplo, o *naive bayes*, que realiza a triagem de e-mails de spam e se aperfeiçoa com o passar do tempo, na medida em que é retroalimentado com o descarte de um e-mail pelo usuário que não tenha sido identificado e tratado automaticamente. Da mesma forma é o revés da hipótese: o algoritmo aprende que a estrutura de um determinado e-mail identificado incorretamente como se fosse spam, não o é, quando o usuário o recupera. Tais erros cometidos pela máquina servem como substrato para o seu aprendizado, que vai absorver o padrão dos e-mails válidos e dos spams, aprimorando sua acurácia a partir de então.

O machine learning é comumente dividido para fins meramente didáticos, em três tipos: supervisionado, não supervisionado e por reforço. Importante pontuar que não existe uma definição formal bem delimitada sobre cada um dos tipos, que podem inclusive ser combinados dentro de um mesmo sistema. O primeiro deles ocorre quando se tem um *dataset* com elementos já rotulados pelo homem (WOLKART, 2020) e a inteligência artificial se utilizará desses elementos anotados para executar, via de regra, atividades de classificação e regressão. Para validação desse tipo de algoritmo, são necessários pelo menos dois conjuntos de dados rotulados, cujo primeiro deles será fornecido integralmente à máquina para alimentar sua base, enquanto o segundo conjunto de dados teriam os seus rótulos ocultados (não informados) para a máquina, os quais servirão de gabarito quando confrontados com a rotulação realizada pelo algoritmo. Nesse modelo, pode ser mencionado o reconhecimento de rostos em fotos. O sistema parte de uma base de dados alimentada pelo usuário, que faz a marcação naquele rosto identificado pela inteligência artificial como seu, e a partir daí a máquina estrutura o padrão, e

consegue identificá-lo, mesmo quando não marcado pelo usuário, sendo aperfeiçoado na medida em que é corrigido manualmente.

Os algoritmos não supervisionados são aqueles programados sem um problema ou métrica especificamente definidos. Ele busca identificar por si só o padrão existente nos dados brutos. Esse tipo de algoritmo pode ter como objetivo exatamente a identificação do padrão, ou, a utilização do padrão identificado para o atingimento de alguma outra tarefa secundária (WOLKART, 2020).

A aprendizagem por reforço, defendida por Murphy (2012, p. 2), tem como diferencial o aprendizado com base em retroalimentação de recompensa ou punição. Nestes casos, o *dataset* não é fixo, e vai sendo ajustado de acordo com a sua interação com o ambiente. É o que mais se assemelha com o aprendizado humano, na medida em que o feedback positivo ou negativo dos seus *outputs* são fundamentais para o aprimoramento algorítmico (WOLKART, 2020). É como uma criança que aprende que queima a mão se encostar no forno quente. Se distingue do aprendizado supervisionado na medida em que neste, os sistemas têm dados com exemplos rotulados, e cada um deles aponta qual a ação correta a ser tomada. No aprendizado por reforço, pretende-se que haja uma atuação para além daquela pré-definida, de forma a conseguir executar ações em situações que não foram objeto do conjunto de treinamento. De outro lado, também difere do aprendizado não supervisionado, vez que o mote principal destes é encontrar as estruturas existentes nos dados apresentados, e, muito embora o aprendizado por reforço também possa fazê-lo, o faz como parte do processo, que tem como objetivo maior a sua otimização por um sinal de recompensa.

No entanto, diante do maior uso da inteligência artificial, que sempre se apresentou extremamente eficiente para a resolução de problemas que lhe eram submetidos, surgiram, em contraposição, diversos outros problemas extremamente simples aos seres humanos e que, diante de sua elementariedade, se mostravam difíceis de parametrizar algoritmos para resolvê-los. Isso porque um ser humano para realizar essas simples tarefas se utilizam de conhecimentos diversos acumulados sobre uma infinidade de situações fáticas que nem conseguiria identificá-las, descrevê-las ou mesmo relacioná-las.

Para reproduzir artificialmente essa gama de conhecimentos correlacionados, buscou-se, por meio do desenvolvimento de uma subespécie de *machine learning* chamada *deep learning*, a estruturação de redes neurais artificiais que permitiram a organização do conhecimento em camadas, sendo possível à máquina o mapeamento de um vetor de entrada e um vetor de saída. Se engana quem imagina que é recente a utilização do aprendizado profundo. Em verdade, o uso de redes neurais já foi cunhado de outros nomes, à exemplo de *cybernetics*

e *connectionism*, mas foi com a atual nomenclatura, *deep learning*, que se desenvolveu mais fortemente em razão do já mencionados dilúvio de dados e avanço do poder de processamento computacional. O sistema de redes neurais tem seu funcionamento inspirado nas sinapses dos neurônios humanos, apesar da distância real entre o sistema artificial e o biológico. De qualquer modo, à exemplo do cérebro humano, a rede neural é composta por diversas unidades individuais, estruturadas em camadas, que são direcionadas para receber *inputs* e enviar *outputs*.

De se registrar que toda a conceituação em torno do *machine learning*, suas divisões e ramificações se entrelaçam, na medida em que é perfeitamente possível que sejam combinadas as técnicas, como já dito, podendo existir em funcionamento um sistema de inteligência artificial que se utilize de *deep learning* de reconhecimento facial que se valha de algoritmos supervisionados. Assim, cresce o volume de estudos de desenvolvimento de algoritmos de aprendizado contínuo, também chamados de *lifelong learning algorithm* (*LLA*) ou *lifelong machine learning* (*LML*) (SILVA, 2019), que buscam cada vez mais alcançar uma inteligência artificial mais generalista (inteligência artificial forte) (HARTMAN PEIXOTO, 2020, p. 35).

As respostas dos sistemas de redes neurais podem variar de acordo com a experiência, pesos, número de camadas e tamanho delas. Sua capacidade de processamento é cada vez maior, ao passo que são ampliadas essas estruturas. Entrementes, esse alto poder de processamento de dados, que já possui uma acurácia superior aos demais sistemas de inteligência artificial, traz consigo um dos maiores riscos envolvidos com a temática: a explicabilidade. Ou, em verdade, a falta dela. Toda a profundidade para que a máquina se auto desenvolva tem como custo a interpretabilidade dos processos por ela utilizados. Na medida em que são acrescentadas camadas, amplia-se a assertividade do sistema, e diminui-se a transparência dos algoritmos, chegando a uma ininteligibilidade que ficou referenciada como "caixa-preta da otimização", ou, simplesmente *blackbox* (MAINI; SABRI, 2017, p. 76).

É possível identificar, portanto, três classes de sistema: compreensíveis (*whitebox*), interpretáveis e opacos (*blackbox*) (ALMEIDA, 2020). Os sistemas compreensíveis (alta interpretabilidade) podem ter sua decisão e processo de tomada de decisão compreendidos por qualquer indivíduo, mesmo sem conhecimento técnico algum. Inclui algoritmos tradicionais de regressão, árvore de decisão e classificadores baseados em regras. Os sistemas interpretáveis (média interpretabilidade), de seu turno, demandam certo nível de compreensão técnica para serem lidos, sendo composto por algoritmos mais avançados, dentre os quais, modelos gráficos; e os sistemas opacos (baixa interpretabilidade) são ininteligíveis a respeito dos processos adotados e motivação para o atingimento do resultado. São assim considerados os que se

utilizam das mais avançadas técnicas de aprendizado de máquina, *support vector machine* (SVM), *ensemble methods* e redes neurais profundas (SILVA, 2019).

Considerando que o atual estado da arte tem apresentado a acurácia em contraposição à compreensão, na medida em que revela que quanto mais simples os modelos utilizados, maior a sua inteligibilidade, porém, menor a sua assertividade, assim como quanto mais sofisticado e profundo o aprendizado de máquina do algoritmo, maior a sua assertividade e menor a sua inteligibilidade, tem sido apostadas muitas fichas na *Explainable Artificial Intelligence* (*xAI*) como o novo graal científico mundial (SILVA, 2019).

1.4 A INTELIGÊNCIA ARTIFICIAL, HOMEM E O NOVO MUNDO COMPLEXO

A inteligência artificial, para o bem ou para o mal, se tornou onipresente. Claramente, pode-se observar três posturas frente aos inegáveis avanços da inteligência artificial: a negacionista, que, a pretexto de uma suposta segurança, tenderia a obstacularizar o seu desenvolvimento, por meio de normativas para impedir, proibir, reprimir e até criminalizar o uso da inteligência artificial; Aquela no sentido de não se preocupar atentamente com os riscos, explorando o enfoque comercial e proveito econômico naturalmente advindo do uso das novas tecnologias; e, por último, uma intermediária, mais equilibrada, que se apresenta como ideal, pois compreende a importância de se adaptar ao novo contexto de evolução tecnológica, busca caminhos e mecanismos para reduzir as vulnerabilidades e riscos, abordando-a de forma ética e responsiva (HARTMAN PEIXOTO, 2020, p. 10).

Por isso, impõe-se sua análise correlacionada ao direito, uma vez que possui diversas implicações que devem ser observadas diante desse novo mundo digital *online*, cada vez mais integrado com o mundo real *offline* (FORTES; CELLA, 2016). A aplicação da inteligência artificial no direito deve primar pela otimização das atividades jurídicas mecânicas e, com as devidas cautelas, até mesmo na tomada de decisão, assegurando-se sempre a possibilidade de revisão humana. Tal posicionamento, invariavelmente forçará os operadores do direito a desenvolverem novas habilidades e campos de atuação onde a máquina não poderá superar o homem, a saber, persuasão, criatividade, empatia, dentre outras (ALVES; ALMEIDA, 2020). Isso porque, é certo que se o homem se limitar à realização de atividades mecânicas e repetitivas, por exemplo, inevitavelmente falhará diante da acurácia e velocidade da máquina artificial.

Não por outra razão que diversos estudos são realizados abordando a temática, trazendo as habilidades e profissões do futuro, ou, em sentido inverso, sinalizando as profissões que

tendem ao perecimento em razão da substituição da mão-de-obra humana por inteligência artificial. É o caso do *papper* de Carl Benedikt Frey e Michael A. Osborne (2013), intitulado "*The future of employment: how susceptible are jobs to computerisation*?[9]" onde buscam os autores analisar os impactos sobre a taxa de emprego diante do crescimento exponencial da tecnologia. Adotam uma metodologia para estipular as possibilidades de informatização de setecentas e duas ocupações, cruzando os dados indexados com o grau de escolaridade e a média do valor dos salários pagos para a ocupação.

A percuciente análise, na linha do que a literatura já sinalizava (FREY; OSBORNE, 2013), ratifica o acima exposto no sentido de que as atividades que importam em uma rotina intensiva, seguindo procedimentos que podem ser bem definidos, serão facilmente executadas por algoritmos sofisticados. Assim, diante da maior acessibilidade às novas tecnologias em razão da queda de preços, já é possível perceber o crescimento da oferta de vagas para funções que demandam um alto esforço cognitivo, com altas faixas salariais, bem como de trabalhos manuais e demasiadamente simplórios de baixa renda. Em contraponto, as atividades repetitivas de remuneração média estão em declínio, especialmente em razão de sua substituição por máquinas (FREY; OSBORNE, 2013).

O reflexo da tentativa de realocação da mão-de-obra é percebido no incremento dos índices de retorno à educação visto que atividades como caixas, balconistas e operadores de telemarketing, bem como a maioria dos trabalhadores em atividades de transporte e logística, juntamente com os trabalhadores de escritório e de apoio administrativo provavelmente serão substituídos pelo capital computacional, tendendo a desaparecer do mercado tais ocupações, uma vez que não exigem alto grau de inteligência social e inteligência criativa (FREY; OSBORNE, 2013).

De outro lado, ao passo que se diminuem determinados postos, são ampliados outros tantos campos de trabalho para ocupações administrativas, comerciais e financeiras, visto que demandam uma incisiva atuação generalista, e uma grande necessidade de inteligência social e criativa. De igual maneira, crescem as ocupações em áreas de educação, saúde, bem como trabalhos de artes e mídia. Cientistas, matemáticos e até mesmo advogados – não seus assistentes e paralegais, mas aquele profissional que atua de forma especialmente cognitiva –, estão na categoria de baixo risco, e possuem, ao contrário, até mesmo o surgimento de novas áreas de atuação (FREY; OSBORNE, 2013).

[9] Tradução livre: O futuro dos empregos: quão suscetíveis são os empregos frente à computadorização?

O mapeamento das atividades foi realizado com base em informações binárias sobre o uso de cognição ou habilidades manuais, e repetitividade ou não das tarefas executadas. Tem-se, assim, quatro possibilidades: tarefas cognitivas rotineiras; tarefas cognitivas não-rotineiras; tarefas manuais rotineiras; e tarefas manuais não-rotineiras. Além de uma série de chamados "gargalos de informatização", quais sejam, percepção, manipulação, inteligência criativa e inteligência social, que podem, assim, ser consideradas como as habilidades do futuro.

Uma forma interessante de organizar essas habilidades do futuro é dividindo-as em cinco eixos, quais sejam: inteligência intrapessoal, inteligência interpessoal, inteligência criativa, inteligência interartificial e inteligência aprendedora ou educadora (GUN, 2020). Para Isabella e Priscilla (2020), deve-se procurar desenvolver habilidades como persuasão, criatividade e empatia, uma vez que não podem ser realizadas por uma máquina artificial. Já Mariana Amaro (2019) na mesma linha, antevê que estarão inseridos nesse novo contexto de automatização aplicada ao mercado de trabalho aqueles que atuam com resolução de problemas, criatividade, imaginação, interação interpessoal e pensamento crítico.

Como se vê, ainda que com a utilização de alguns nomes diferentes, o caminho que se espera parece ser o mesmo, com um distanciamento de atividades braçais, repetitivas que serão substituídas por automação e inteligência artificial, devendo o ser humano focar no que lhe faz *ser* humano.

A primeira delas diz respeito à capacidade de se conectar consigo mesmo, de compreender sua própria essência, gerindo suas emoções e controlando seus medos. É postada como a primeira delas por ser o ponto de partida, deter um poder de autoconhecimento para que então sejam desenvolvidas as demais habilidades, em um solo fértil. A inteligência interpessoal é aquela que envolve as relações entre as pessoas, a capacidade de se relacionar e principalmente de se comunicar com as outras pessoas, compreendendo sonhos e desejos, dialogando sobre emoções e pensamentos comuns.

A inteligência criativa guarda relação com a inovação, o *insight*, a *combinatividade* entre os *inputs* que integram o repertório do agente, para a criação de novos *outputs*, não necessariamente alinhados com o padrão que ordinariamente se espera. A inteligência interartificial dá suporte para a relação entre homens e a tecnologia, a fim de que o agente se utilize de suas potencialidades de uma forma saudável, e não seja usado pela máquina. Por fim, a inteligência aprendedora ou ensinadora quando bem desenvolvida permite ao agente uma condição de adaptabilidade diante das mudanças, aprendendo a aprender.

Todas essas habilidades são ligadas a atividades cuja sensibilidade humana, intuitividade, a compreensão da relação entre causa e efeito são essenciais, em maior ou menor

grau, de modo que o seu desenvolvimento firme é o que assegurará a importância do homem frente aos avanços da tecnologia, uma vez que são tipos de inteligência ainda não alcançadas de maneira artificial, e que dificilmente poderão ser reproduzidas.

Essas são algumas das bases que sustentam a proposição de mudança de era por estudiosos como Marc Haley, Dee Hock, Walter Longo (GUN, 2020), apoiada na quarta revolução industrial (SCHWAB, 2016), que trouxe a hiperconectividade entre os seres humanos e as coisas (*internet of things - IoT*), permitindo a abundância de dados e desenvolvimento algorítmico, em um contexto de complexidade interrelacional até então nunca visto. Esta intrincada complexidade que se apresenta como um tecido de constituintes heterogêneas, associadas inseparavelmente, onde o todo que é mais, e ao mesmo tempo menos, importante que as partes, a depender do momento de observação (MORIN, 2006, p. 13).

Dando mais profundidade teórica à complexidade abordada de forma superficial até agora, vê-se que tal percepção, mais encorpada a partir do século XX, teve como um de seus maiores expoentes o sociólogo alemão Niklas Luhmann, com a sua teoria geral dos sistemas sociais. Nascido em 1927, graduado em direito e pós-doutor, ganhou notoriedade quando se dedicou à sociologia, apesar das significativas contribuições de seus estudos para política, religião, artes, economia e sistemas comunicacionais, tendo como obra principal *Die Gesellschaft der Gesellschaft*[10], de 1997 (TRINDADE, 2008).

Sua teoria geral dos sistemas, que modifica toda uma linha lógica-dedutiva pós-newtoniana, propõe uma ideia de estruturalismo funcional (em oposição ao funcionalismo estrutural), voltada para o "sistema" e para o "entorno", em contraposição à "unidade" e ao "todo". Pressupõe que o *mundo* seria o pano de fundo, onde se encontra toda a complexidade, e por tal, não poderia ser concebido como sistema e nem como entorno. Não é sistema, visto que não teria entorno que naturalmente o é mais complexo, assim como também não é entorno, por não ter um interior que pudesse lhe ser diverso e organizado pelo sentido. Dentro dessa referência suprema de mundo é que está a complexidade. E a complexidade, desta forma, não pode ser compreendida pela consciência humana, considerando todas as variáveis e circunstâncias no mundo.

Entre a extrema complexidade e a limitada capacidade da consciência humana, existe, pois, uma lacuna que é preenchida pelos sistemas sociais (NEVES; NEVES, 2006). Os sistemas, então, são criados para simplificação desta complexidade por meio do sentido, e seriam selecionados a partir da diferença e não mais da semelhança, ou seja, a partir do

[10] A Sociedade da Sociedade.

momento em que a conexão entre elementos não se fizesse mais possível, tal elemento não integraria o sistema, sempre privilegiando a variação e contingência por meio da experimentação, rejeitando o modelo anterior de uniformidade e previsibilidade (organização estática). Algo é complexo, pois, quando envolve mais de uma circunstância. E na medida em que aumentam as possibilidades, são ampliadas igualmente a quantidade de relações, e por via de efeito, a própria complexidade.

Por isso, uma análise sistêmica permite uma visão geral sobre o todo, bem como das suas partes e inter-relações comunicativas existentes entre os sistemas, facilitando a compreensão da complexidade interna, e identificação da diferenciação funcional (TACCA; ROCHA, 2018).

Nas palavras de Luhmann, complexidade é verificada "quando num conjunto interrelacionado de elementos já não é possível que cada elemento se relacione em qualquer momento com todos os demais, devido a limitações imanentes à capacidade de interconectá-los" (LUHMANN, 1990, p. 69). Este é o momento em que precisa ocorrer a seleção: "a complexidade significa obrigação à seleção, obrigação à seleção significa contingência e contingência significa risco" (LUHMANN, 1990, p. 69).

Essa visão paradigmática remove o antropocentrismo social, colocando o homem no entorno, e os sistemas, em especial as comunicações e as relações, em seu epicentro. Apropria-se da ideia de *autopoiese*, de Humberto Maturana e Francisco Varela (1995, p. 136), concebida na biologia, para aplicá-la na sociologia, com a auto-organização e manutenção do equilíbrio dinâmico do sistema com o meio (LUHMANN, 2016, p. 87). Toda a teoria geral de sistemas de Luhmann, e a sua proposição de complexidade são fundamentais para a conexão da teoria com a inteligência artificial. Isso porque, para que seja compreendida a complexidade, é introduzida a figura do observador, bem como a teoria desenvolvida a partir deste ponto, que é a observação de segunda ordem (MATURANA; VARELA, 1995, p. 40).

A teoria da observação de segunda ordem importa na compreensão de que todos os sistemas têm uma contingência, ou um ponto-cego, que somente outro sistema tem a capacidade de perceber. Isso porque, a operação, assim entendido pelo seu processo de reprodução do sistema, difere da observação, que consiste no ato de diferenciar com vistas à criação de informações. Os sistemas autopoiéticos, que mais tarde foram chamados de autorreferentes, são, portanto, capazes de, para além da sua capacidade de operação, observarem a si mesmos. E a observação de segunda ordem consiste na observação do observador. Enquanto a observação de primeira ordem consiste na observação de um evento, a observação de segunda

ordem importa não na observação do evento, mas na observação de como observador observa o evento, em uma perspectiva mais voltada para o *como* do que o *que*.

Esse foi um dos principais motes para a evolução algorítmica exponencial. A observação de segunda ordem foi decisiva para o desenvolvimento da capacidade de compreender a hiperconectividade, introduzindo técnicas de observação dos esquemas de observação (observação de segunda ordem) para internalizar para si estruturas operativas que reproduzem tais esquemas de observação de primeira ordem (PEREIRA, 2019).

Além desta perspectiva de contribuição ao desenvolvimento algorítmico, a teoria luhmanniana apesar de ter sido idealizada numa perspectiva de dois sistemas psíquicos, também se mostra perfeitamente aplicável à relação entre sistema psíquico e artificial, consoante alcançado por Habermas, para quem é clara a sua estratégia originada da percepção de que

> se a operação do símbolo linguístico se esgota na articulação, na abstração e na generalização de processos da consciência e de contextos de sentido pré-linguísticos, a comunicação realizada com os meios da linguagem não pode ser explicada com base nas condições de possibilidade especificamente linguísticas (HABERMAS, 2000, p. 527).

Prossegue Habermas (2000, p. 528) aduzindo que, "os aspectos da intersubjetividade linguisticamente gerada precisam ser derivados, como artefatos autoproduzidos, das reações recíprocas dos sistemas elaboradores de sentido". Resta, portanto, evidenciada a insuficiência da interpretabilidade de texto com base nas regras rígidas da linguística. Os artefatos produzidos pela inteligência artificial, com suas novas técnicas e ferramentas para extrair significação dos textos, tornou desnecessária a figura do especialista humano, densificando exponencialmente a capacidade de mineração de dados, especialmente dos textos, dos algoritmos, que, passaram a ameaçar mais ainda o ser humano em razão da sua nova e abissal capacidade de leitura e absorção de informações.

Todavia, tem-se que ainda há muito campo para o ser humano se reposicionar em um admirável mundo novo[11], dependendo de si próprio para, adaptando-se, conseguir se utilizar da inteligência artificial para potencializar ainda mais suas habilidades pessoais ímpares e não se prender somente ao discurso filosófico ontológico (KASTER; ROVER, 2019), se preparando para um eventual cenário de avanço potencializado da tecnologia, já que não se trata de estar de um lado (pessimista) ou de outro (entusiasta), mas de aceitar essa evolução, assim como o avanço da inteligência artificial, que integrará grande parte dos segmentos de trabalho, entretanto, com uma postura irrequieta acerca de diretrizes éticas, e do avanço responsável.

[11] Em alusão ao romance futurista homônimo (HUXLEY, 2014).

De se ressaltar, ainda, que apesar do crescimento vertiginoso da chamada quarta revolução industrial no cenário mundial, onde seu alcance tem sido exponencialmente superior às revoluções industriais anteriores, ainda há um déficit populacional que não alcançou sequer as benesses das primeiras revoluções, já que mais de um bilhão de pessoas no mundo inteiro ainda não têm acesso à eletricidade. Além disso, cerca de quatro bilhões de pessoas, o equivalente a mais da metade da população mundial, não têm sequer acesso à internet, apesar de muitas delas viverem em países em desenvolvimento. Tais dados são relevantes neste ponto de inflexão, pois, se de um lado tem-se um desenvolvimento indiscutivelmente mais rápido, considerando que a internet (substrato principal da terceira revolução industrial) se espalhou pelo mundo em menos de dez anos, enquanto o tear mecanizado (objeto que ilustra a primeira revolução industrial) levou mais de um século para alcançar proporção continental, de outro lado, é certo que as revoluções não atingem a sociedade de forma homogênea e igualitária, se reservando às camadas populacionais mais abastadas (SCHWAB, 2016, p. 17), mantendo-se a salvos os postos daqueles que persistirem em realizar atividades rotineiras, manuais ou cognitivas.

Vale mencionar, em linhas pretéritas, a respeito do efetivo alcance da quarta revolução industrial a quem tem acesso a ela. É perceptível que ainda se mostra necessária, ao setor privado, governo, instituições públicas e até mesmo à população, a criação de uma maior consciência e percepção para esse momento disruptivo que se apresenta, de forma que se afastem de uma polarização dirigida por filtros que os aproximam apenas de comunidades que compartilham de iguais ideologias e de um analfabetismo digital (MAGRANI, 2014, p. 148) que em nada contribui para a necessária reformulação de todo o cenário social, político e econômico que deve, ao contrário, conduzir o caminho que essas fundamentais mudanças devem seguir, estabelecendo uma narrativa ponderada e propositiva, que identifique as suas potencialidades e vulnerabilidades, acalmando os argumentos apocalípticos (SCHWAB, 2016, p. 17) que tratam a singularidade da inteligência artificial como uma verdadeira espada de Dâmocles.

2 DIREITO E A INTELIGÊNCIA ARTIFICIAL

Abordadas muitas das premissas técnicas e filosóficas que circundam a inteligência artificial, passa-se a tratar de algumas das aplicações jurídicas concernentes a esse novo mundo de relações complexas entre o homem e a máquina. Tratamento, enviesamento e proteção de dados, (falta de) explicabilidade no processo de tomada de decisão, supervisão humana sobre as decisões automatizadas e tecnorregulação são questões sensíveis que precisam ser enfrentadas pelos actantes a fim de avançar responsavelmente com o desenvolvimento da inteligência artificial.

Evidentemente as questões acima referenciadas não são os únicos riscos e desafios do uso e desenvolvimento exponencial da inteligência artificial, dos quais podem ser lembrados, ainda, a identificação e re-identificação de indivíduos, isolação e discriminação, filtro bolha, reuso de dados em aplicações distintas, armazenamento perpétuo de dados, coleta de dados irrelevantes, consentimento falho etc. (MASSENO; SANTOS, 2019), mas foram aqueles com maior viés jurídico e social escolhidos para serem desenvolvidos.

2.1 TRATAMENTO DOS DADOS

Como visto, a quarta revolução industrial tem como um dos fatores principais o chamado *big data*. Ao mesmo tempo em que o incremento na produção e captação de dados permitiu um significativo avanço no que diz respeito à sofisticação algorítmica, também colocou a sociedade no que se referenciou anteriormente como um panóptico virtual voluntário. A alusão ao modelo de vigilância de Bentham (FOUCAULT, 1987, p. 226), que inicialmente pode até induzir que a sociedade se encontra em uma posição ativa de vigilância, em verdade, desponta rapidamente a demonstrar que os indivíduos estão em uma posição passiva, e que todos são, ao contrário, vigiados pelos rastros digitais que são coletados à sua revelia ou entregues voluntariamente. Criou-se um ilusório cenário de liberdade, em que na verdade, as pessoas passaram a ficar aprisionadas nas estruturas coercitivas da sociedade, agora baseadas na positividade, que, em excesso, tem se mostrado mais patológica do que a negatividade (HAN, 2015, p. 17). Descabe, portanto, a associação da internet com liberdade, posto que, apesar de aparentar se tratar de um campo livre para expressão individual, a rede mundial de computadores conta com uma estrutura organizacional de funcionamento que permite aos seus responsáveis controlar as informações disponíveis, limitando a liberdade de manifestação (GALINDO; CARMO, 2017).

Merece especial atenção, portanto, a questão da proteção de dados, já que a evolução algorítmica tem se apresentado cada vez mais complexa e distante de uma desejada neutralidade, o que revela um flanco exposto para a geração de resultados viciados quando da alimentação de sistemas com dados enviesados. É o que se viu, por exemplo, no famoso caso do sistema norte-americano *Correctional Offender Management Profiling for Alternative Sanctions* (COMPAS) (MARQUES, 2019), em que eram avaliados diversos dados pessoais dos condenados a fim de estabelecer o seu risco de reincidência, se utilizando de *evidence-based* (dados de experiência), que repercutiriam na sua progressão de regime, concessão de liberdade e cálculo da pena a ser fixada.

Para além do desafio da opacidade do sistema, que será tratado adiante, da análise dos resultados do COMPAS, percebeu-se que os índices de alerta a respeito de pessoas integrantes de grupos de minorias étnicas, em especial da raça negra, eram sobremaneira elevados quando comparados com os índices de pessoas brancas em situação semelhante. O sistema apontava para penas e regimes de cumprimento mais duros para os primeiros, mesmo quando os crimes cometidos e as suas circunstâncias eram idênticas. Tal enviesamento do sistema de inteligência artificial teve grande repercussão, especialmente porque não era inserida no sistema nenhuma informação com relação à raça do apenado. Foi por isso que se descobriu que algumas das informações que se mostravam mais recorrentes em minorias étnicas, como, por exemplo, o domicílio dos condenados, o fato de o próprio apenado, algum familiar ou amigo já ter sido preso, tinham peso significativo – e completamente desconhecido, diga-se – nessa avaliação, e, dada à característica particular dos Estados Unidos de possuir uma grande divisão racial em bairros, temos um discrímen indevido, que aponta para um direito penal do autor, e não um direito penal do fato (LARSON; MATTU; KIRCHNER, ANGWIN, 2016).

Essa situação é emblemática para demonstrar a necessidade da utilização de dados não viciados para alimentar um sistema de inteligência artificial, especialmente quando se fala de *machine learning* e, principalmente, *deep learning*, onde as bases do algoritmo serão construídas ou aprimoradas apoiadas nessas informações. Não foi inserido nenhum comando que determinasse que negros deveriam ter penas mais duras do que pessoas brancas, mas, o fato de existir uma concentração de pessoas negras em bairros com maiores índices de violência, fez com que o sistema, ao identificar que o condenado mora em determinada região, ou possua histórico de relações com pessoas em situações de violência, o tratasse de uma forma diferente de como trataria uma pessoa branca que tivesse cometido o mesmo crime (EDWARDS; VEALE, 2017). Esse tratamento diferenciado tem sido firmemente combatido, sobretudo neste momento que voltou à tona nos Estados Unidos o enfrentamento ao racismo, inobstante a

Suprema Corte do Wisconsin não tenha dado provimento à apelação de Eric L. Loomis, que atacava a ausência de informações acerca do peso dado às respostas das questões analisadas pelo sistema. O Tribunal entendeu que o algoritmo estaria protegido pela propriedade intelectual da empresa desenvolvedora, não podendo, portanto, ser violado. Declarou, ainda, que a utilização do sistema deveria se dar de forma não exclusiva para o suporte ao magistrado, e que não poderá o juízo tê-lo como fundamentação única (ESTADOS UNIDOS DA AMÉRICA, 2016).

Quando se fala em dados enviesados ou *bias*, deve ser alcançado que não se trata de o dado conter algum tipo de discriminação em si, no entanto, quando analisados sob uma visão holística, pode-se constatar que se tratam de dados obtidos com base em uma seleção discriminatória ou puramente heurística, que restará perpetuada no DNA do algoritmo de aprendizagem de máquina, uma vez que esse será estruturado a fim de replicar os resultados que alimentaram o *dataset* (MULHOLLAND; FRAJHOF, 2019). É o que ocorreria, por exemplo, em um sistema de inteligência artificial de seleção de candidatos a uma vaga de emprego que se utiliza de *deep learning*, em que a base de dados utilizada para treinar o algoritmo tiver uma predominância de homens brancos de meia idade em comparação com mulheres, negros ou jovens. Mesmo que não exista nada expressamente que indique qual o perfil de empregado deva ser contratado no que se refere à raça, gênero ou idade, é muito provável que os resultados se mostrem semelhantes ao estereótipo criado com base nos dados, apontando homens brancos de meia idade como o perfil de candidato ideal, caso não tenha sido realizada uma configuração expressa para rechaçar esse viés. Essa falsa sensação de objetividade, neutralidade, racionalidade e imparcialidade dos algoritmos tem ruído em razão dessa constatação testada em diversas pesquisas (MULHOLLAND; FRAJHOF, 2019).

Outra questão sensível relativa aos dados é a sua proteção após conectados na rede. Como visto, a internet das coisas trouxe um cenário de hiperconectividade em que diversos artefatos eletrônicos estão captando dados de forma estruturada a cada instante, de todos os aspectos da rotina das pessoas, animais, meio ambiente etc. Alguns desses sensores sequer são conhecidos pelos indivíduos vigiados, muito embora a ciência de que o preço do benefício pretendido seria o fornecimento voluntário de dados provavelmente não diminuiria a quantidade de pessoas dispostas a pagá-lo. Esse é um dos frutos da liquidez da modernidade (BAUMAN, 2013), que remodelou a sociedade após a segunda guerra mundial, e trouxe uma superficialidade às relações pessoais, que deixaram de ser sólidas, firmes e duradouras, e passaram a ser líquidas, efêmeras e frágeis, formando um cenário de imprevisibilidade, incertezas e insegurança que demonstra que nada foi feito para durar. O impacto das redes

sociais nessa metamorfose é sobremaneira significativo, visto que a lógica do consumo passou a se sobrepor à lógica da moral. O alto poder de consumo sempre foi associado a *status*, porém, na modernidade líquida, o consumismo se tornou imperativo. A exploração capitalista deixou de ser vista como exploração, já que o controle está na mão do indivíduo, que se submete voluntariamente à lógica capitalista de consumo. As pessoas transformaram relações e relacionamentos outrora duradouros, em meras conexões, que podem ser substituídas a qualquer tempo, alterando o desejo de qualidade para uma busca incessante de quantidade, e com uma necessidade de auto exposição visceral e ao mesmo tempo irreal e superficial, para uma reafirmação pessoal na sociedade, de uma maneira nunca antes vista. Até mesmo a estrutura organizacional dos negócios possui intrinsicamente elementos de desorganização, em uma concepção de que "quanto menos sólida e mais fluida, melhor" (BAUMAN, 2001, p. 194).

Exemplo disso são os modismos que ganharam visibilidade mundial nas redes sociais, que podem servir para um aperfeiçoamento algorítmico, como o *10 years challenge*, em que as pessoas faziam montagens com fotos pessoais de seu rosto com intervalo de tempo de dez anos entre cada foto. A brincadeira, que teve grande repercussão nas redes sociais, instigava a curiosidade das pessoas para perceberem o quão diferente se encontravam depois de uma década. Longe de encampar uma postura paranoica de que se tratava de um plano sofisticado para aprimorar softwares de reconhecimento facial, é certo de que a obtenção de dados limpos e rotulados, é extremamente conveniente para tal fim, quer tenha sido criado propositalmente ou não. O que deve ser objeto de ponderação, como dito, é a consciência do custo envolvido em razão de tal exposição. E que, além de exigirmos que as empresas e governos tratem adequadamente os dados sensíveis, que entendamos que os titulares precisam se dignar a tratar seus próprios dados adequadamente.

Outro caso que reascendeu a discussão, foi o aplicativo de origem russa *FaceApp*. No referido software é possível que a fotografia de uma pessoa seja manipulada simulando como ela seria se fosse do gênero oposto, tivesse um corte de cabelo ou acessórios diferentes, e até mesmo seu envelhecimento ou rejuvenescimento, quem sabe, se utilizando de tecnologia desenvolvida com os dados do desafio dos dez anos. Mas a curiosidade dessa situação não era exatamente o que o aplicativo fornecia; essa era apenas a parte atrativa que fez com que ele tivesse uma grande adesão por parte dos usuários. A controvérsia estava nos termos de uso do software, que concedia à empresa um acesso amplo e irrestrito a uma série de informações dos dispositivos que vão muito além do que ordinariamente precisaria para a aplicação dos filtros nas fotografias dos usuários, dentre as quais, informações sobre as atividades online em

aplicativos e sites, sites visitados, tempo de permanência em cada página, histórico de compra, informações essas que poderiam ser, inclusive, compartilhadas com parceiros da empresa.

A vulnerabilidade decorrente do compartilhamento inconsciente dos dados é um dos efeitos colaterais que esse contexto de modernidade líquida trouxe, como pontuado nos dois casos acima mencionados. Não existe mais a percepção de divisão das esferas privada e pública, no que se refere ao compartilhamento de informações pessoais. As pessoas compartilham voluntaria e gratuitamente suas vidas inteiras, sem se aperceber dos

> [...] riscos terminais à privacidade e à autonomia individual, emanados da ampla abertura da arena pública aos interesses privados, e a sua gradual, mas incessante transformação numa espécie de teatro de variedades dedicado à diversão ligeira (BAUMAN, 2013, p. 113).

Por isso a adoção de medidas de proteção aos dados se apresenta não apenas como necessária, mas urgente e obrigatória, visto que no contexto de hiperconectividade, os dados são a matéria prima da inteligência artificial (PAIVA, 2020), e não protegê-los adequadamente para o seu tratamento pelos sistemas inteligentes, implica em uma construção enviesada dos algoritmos.

Não por outro motivo, na última década houve significativo avanço no que diz respeito à proteção de dados, principalmente os chamados dados sensíveis (MAGRANI, 2019, p. 57), assim considerados aqueles que guardam relação com a origem racial, opiniões políticas, convicções religiosas, questões sobre saúde ou vida sexual das pessoas. A Europa como um todo, assim como países como Estados Unidos e Canadá, têm estado a frente no desenvolvimento de estudos, na tentativa de que seja estabelecida uma regulação da mineração e utilização de dados, e, por via de consequência, da própria inteligência artificial.

2.2 (FALTA DE) EXPLICABILIDADE NO PROCESSO DE TOMADA DE DECISÃO

O processo de tomada de decisão é, talvez, um dos pontos mais sensíveis da resistência para aceitação irrestrita da inteligência artificial. É consabido que muitos não fazem questão da explicabilidade acerca da decisão tomada pelo sistema, mas, essa não pode ser a tônica da situação, mormente quando essa postura despreocupada será fatalmente alterada quando uma decisão contrária aos interesses do indivíduo que a ignorou for tomada sem que reste perfeitamente esclarecida a racionalidade do iter procedimental que resultou na decisão.

É certo que o uso de dados para a predição de resultados já era objeto de estudo da jurimetria, que consiste basicamente na disciplina do conhecimento que utiliza a metodologia

estatística para investigar o funcionamento de uma ordem jurídica (NUNES, 2019, p. 111). Possui, assim como muitos aspectos da inteligência artificial, um propósito de identificação de padrões com base em dados que a princípio não pareceriam mensuráveis, deslocando o eixo gravitacional do estudo do plano abstrato, de normas gerais como a lei, para o plano concreto, de normas individuais e específicas, como sentenças, acórdãos, contratos e demais instrumentos jurídicos em sentido concreto (NUNES, 2019, p. 108). Não se trata, entretanto, de ignorar o plano abstrato, mas de alçar o plano concreto à referência principal do objeto do direito, para, assim, compreender a relação de causa e efeito da norma se utilizando da estatística para investigar os múltiplos fatores sociais, econômicos, psíquicos, éticos etc., que influenciam o comportamento dos partícipes. A tomada de decisão realizada por inteligência artificial, portanto, compartilha das mesmas raízes da jurimetria, na medida em que parte da utilização de um *database* de diversas outras situações pretéritas, e demais informações relevantes ao caso para decidir o que fazer, se valendo, entretanto, de uma capacidade computacional sobre-humana.

A tomada de decisão por sistema dotado de inteligência artificial, quando apoiada em uma análise baseada na experiência, até pode ter um alto grau de interpretabilidade. No entanto, na medida em que a predição de resultados advém de uma (re)combinação de *inputs* com métricas e pesos desconhecidos, é muito provável que sejam alcançados resultados assertivos, porém, com uma opacidade indesejada inserta no processo decisório. O desafio é ampliado na medida em que esses critérios são estabelecidos pela própria inteligência artificial e não pelo desenvolvedor, que pode não conseguir compreender a racionalidade da heurística algorítmica (MANGETH, 2019). O direito à explicação deriva do princípio da transparência, e se apresenta como uma ferramenta de *accountability* que permite saber em qual medida determinado *input* influenciou o resultado (BIONI; LUCIANO, 2019). Já se nota que o conflito surge em razão de os desenvolvedores e principalmente as empresas que investem no desenvolvimento de inteligência artificial preservarem o seu direito à propriedade intelectual, segredos comerciais e industriais, que de fato não podem ser ignorados.

É possível se admitir um grau de explicabilidade que demonstre o impacto e a forma que determinados dados sensíveis modificam a sintonização necessária para a geração dos *outputs*, sem que haja uma completa exposição do algoritmo. Porém, também, a depender do sistema em si, essa configuração pode ser exatamente a essência do sistema inteligente, que poderia ser rapidamente replicado por empresas concorrentes, caso conhecido (BIONI; LUCIANO, 2019). Esse é o desafio maior dessa relação entre inteligibilidade algorítmica e

assertividade. Como já visto, o uso de métodos que se mostram mais compreensíveis contrasta inversamente com a assertividade do sistema inteligente.

Por isso a falta de explicabilidade se apresenta como uma questão mais preocupante que o enviesamento de dados no caso do sistema de inteligência artificial norte-americano utilizado para dosimetria da pena. Como visto, a inserção de dados com viés preconceituoso é refletida nos *outputs* do COMPAS. Mas, o principal motivo que levou o apenado a recorrer não foi a existência de *bias* no algoritmo, mas o resultado que esse enviesamento proporcionou, em razão do desconhecimento de como se dava o processo de tomada de decisão do sistema para atribuição de notas, diante das informações prestadas.

É o mesmo caso ocorrido no Brasil acerca do serviço de *score* de crédito, que teve o idêntico desfecho permissivo por parte da corte judicial. Uma empresa de análise de crédito passou a realizar, sem o conhecimento ou consentimento das pessoas cadastradas, antes mesmo da previsão da lei do cadastro positivo, havida em 2019[12], um banco de dados em que atribuía uma pontuação (*score*) à pessoa cadastrada, assim como indicava a possibilidade de fraude na documentação, estipulando até mesmo uma faixa de valor até onde seria segura a concessão de crédito a ser disponibilizado. Não demorou para que os consumidores se insurgissem em larga escala em face dessa rotulação obscura realizada pela empresa, que cobrava pelo fornecimento destas informações, as quais eram geradas sem que fossem esclarecidos quais os dados utilizados como *inputs*, tampouco a sua forma de processamento. Diante da multiplicidade de ações e posicionamentos diversos em torno de tal tema – que chegou a mais de duzentas e cinquenta mil ações no país –, o Superior Tribunal de Justiça (STJ), ao receber o primeiro recurso especial, o afetou como representativo da controvérsia, determinou a suspensão da tramitação de todas as ações que versavam sobre o assunto (BRASIL, 2014). Após a realização de audiências públicas, entendeu a corte quando do julgamento do Recurso Especial (REsp) n. 1.419.697, pela legalidade do sistema de score de crédito, não havendo a necessidade de a empresa desenvolvedora revelar a fórmula utilizada para obter a pontuação, devendo, contraditoriamente, "haver transparência e clareza no que diz respeito aos dados utilizados" (BRASIL, 2014), como já estabelece o Código de Defesa do Consumidor (CDC), no artigo 4º[13]

[12] Lei complementar n. 166/19, que altera a lei complementar n. 105, de 10 de janeiro de 2001, e a lei n. 12.414, de 9 de junho de 2011, para dispor sobre os cadastros positivos de crédito e regular a responsabilidade civil dos operadores (BRASIL, 2019e).

[13] Lei n. 8.078/1990. Código de defesa do consumidor. Art. 4º A Política Nacional das Relações de Consumo tem por objetivo o atendimento das necessidades dos consumidores, o respeito à sua dignidade, saúde e segurança, a proteção de seus interesses econômicos, a melhoria da sua qualidade de vida, bem como a transparência e harmonia das relações de consumo, atendidos os seguintes princípios: [...] (BRASIL, 2017).

e artigo 6º[14], para que o consumidor possa retificá-los, caso se apresentem incorretos ou desatualizados.

Ambas as situações são típicos casos em que, apesar de os tribunais buscarem firmar posicionamento de que existe um dever de transparência, este resta limitado face ao segredo comercial do algoritmo, que faz com que de pouco, ou nada adiante a transparência relativa aos dados. Os interessados não têm a ciência de como são processadas suas informações, qual o peso que foi dado a cada um dos dados, o que revela tratar-se de uma verdadeira *blackbox*, que contraria essencialmente o princípio da transparência e o direito à explicação.

Situações como essas demonstram que o romance kafkiano "*O Processo*" (KAFKA, 1925), em que o personagem principal é perseguido, julgado em um processo cujo acesso nunca lhe fora franqueado, tampouco lhe fora dado conhecimento à acusação que pendia contra si, não se tratava de uma ficção tão distante da realidade. A depender de interesses políticos, econômicos e sociais, resta claramente demonstrada a disposição em flexibilizar o direito à explicação, sob os mais variados argumentos, relegando ao indivíduo uma condição de mera aceitação, diante de uma estruturação jurídica sistêmica que supostamente deveria ter por escopo assegurar uma convivência social justa da coletividade.

Por isso uma regulamentação firmando não apenas a obrigatoriedade de transparência e explicação, mas o nível de detalhamento de como deverão ser materializados se faz tão importante. Na França, o direito à explicação em face de decisões automatizadas já tinha amparo legal desde a lei de informática e liberdades, datada de 06/01/1978. De igual maneira, a Diretiva n. 95/46/CE (UNIÃO EUROPEIA, 1995) também veio a prever tal direito, até que fosse substituída pelo Regulamento Geral de Proteção de Dados Europeu (*General Data Protection Regulation - GDPR*) (UNIÃO EUROPEIA, 2016), que replica o mesmo dever de explicação e oposição, além de referenciar a observância aos termos da Carta dos Direitos Fundamentais da Europa e Convenções do Conselho da Europa (VERONESE; SILVEIRA; LEMOS, 2019).

Ao se analisar o artigo 22 (3)[15], da lei geral de proteção de dados europeia (*General Data Protection Regulation - GDPR)* (UNIÃO EUROPEIA, 2016), tem-se a determinação ao

[14] Lei n. 8.078/1990. Código de defesa do consumidor Art. 6º São direitos básicos do consumidor: [...] III - a informação adequada e clara sobre os diferentes produtos e serviços, com especificação correta de quantidade, características, composição, qualidade, tributos incidentes e preço, bem como sobre os riscos que apresentem; IV - a proteção contra a publicidade enganosa e abusiva, métodos comerciais coercitivos ou desleais, bem como contra práticas e cláusulas abusivas ou impostas no fornecimento de produtos e serviços;[...] (BRASIL, 2017).

[15] *Artigo 22.º (Decisões individuais automatizadas, incluindo definição de perfis)* [...] 3. Nos casos a que se referem o n. 2, alíneas a) [decisão necessária para um contrato] e c) [decisão obtida com o consentimento do titular], o responsável pelo tratamento aplica medidas adequadas para salvaguardar os direitos e liberdades e legítimos

responsável pela manipulação dos dados de que assegure a fruição dos direitos, liberdades e interesses do seu titular, lhe garantindo, ainda, o direito à manifestação, oposição e revisão humana no que se refere à decisão. Observe-se que não há um rol exaustivo, de forma que seja sustentada a inexistência do direito à explicação no sistema. Ao contrário, sua presença implícita se alinha exatamente com o que parece pretender o legislador, sobretudo quando analisado o artigo em conjunto com o *considerando* n. 71[16], que afirma textualmente que o titular dos dados tem a garantia de informação específica, bem como de obter uma explicação sobre uma decisão já tomada, portanto, *a posteriori*, em duas hipóteses: quando a decisão afetar significativamente a vida do indivíduo, ou quando operar efeitos na esfera legal, estabelecendo um nível de salvaguarda bem mais alto do que efetivamente o é reconhecido. Contudo, é preciso ponderar que o quanto determinado nos *considerandos* deve servir tão somente de orientação à interpretação dos artigos estabelecidos na legislação, não possuindo força vinculante, pelo que se mostra controvertida a concepção do direito à explicação nos termos estabelecidos neste *guideline*.

Os artigos 13 e 14, de seu turno, determinam ao responsável pelo tratamento dos dados que notifique quando realizar a sua coleta, seja diretamente ou por meio de terceiros, informando ao titular dos dados a sua submissão aos processos de tomada de decisão de forma automatizada, bem como acerca da *lógica subjacente*, importância e consequências previstas de tal tratamento.

> Artigo 13.º (Informações a facultar quando os dados pessoais são recolhidos junto do titular) 1. Quando os dados pessoais forem recolhidos junto do titular, o responsável pelo tratamento faculta-lhe, quando da recolha desses dados pessoais, as seguintes informações: a) A identidade e os contatos do responsável pelo tratamento e, se for caso disso, do seu representante; b) Os contatos do encarregado da proteção de dados, se for caso disso; c) As finalidades do tratamento a que os dados pessoais se destinam, bem como o fundamento jurídico para o tratamento; d) Se o tratamento dos dados se basear no artigo 6.º, n. 1, alínea f), os interesses legítimos do responsável pelo tratamento ou de um terceiro; e) Os destinatários ou categorias de destinatários dos dados pessoais, se os houver; f) Se for caso disso, o fato de o responsável pelo tratamento tencionar transferir dados pessoais para um país terceiro ou uma organização internacional, e a existência ou não de uma decisão de adequação adotada pela Comissão ou, no caso das transferências mencionadas nos artigos 46.º ou 47.º, ou no artigo 49.º, n. 1, segundo parágrafo, a referência às garantias apropriadas ou adequadas e aos meios de obter cópia das mesmas, ou onde foram disponibilizadas. [...] (UNIÃO EUROPEIA, 2016).

interesses do titular dos dados, designadamente o direito de, pelo menos, obter intervenção humana por parte do responsável, manifestar o seu ponto de vista e contestar a decisão.

[16] *Considerando n. 71*. [...] Em qualquer dos casos, tal tratamento deverá ser acompanhado das garantias adequadas, que deverão incluir a informação específica ao titular dos dados e o direito de obter a intervenção humana, de manifestar o seu ponto de vista, de obter uma explicação sobre a decisão tomada na sequência dessa avaliação e de contestar a decisão. [...] (UNIÃO EUROPEIA, 2016).

> Artigo 14.º (Informações a facultar quando os dados pessoais não são recolhidos junto do titular) 1. Quando os dados pessoais não forem recolhidos junto do titular, o responsável pelo tratamento fornece-lhe as seguintes informações: a) A identidade e os contatos do responsável pelo tratamento e, se for caso disso, do seu representante; b) Os contatos do encarregado da proteção de dados, se for caso disso; c) As finalidades do tratamento a que os dados pessoais se destinam, bem como o fundamento jurídico para o tratamento; d) As categorias dos dados pessoais em questão; e) Os destinatários ou categorias de destinatários dos dados pessoais, se os houver; f) Se for caso disso, o facto de o responsável pelo tratamento tencionar transferir dados pessoais para um país terceiro ou uma organização internacional, e a existência ou não de uma decisão de adequação adotada pela Comissão ou, no caso das transferências mencionadas nos artigos 46.º ou 47.º, ou no artigo 49.º, n. 1, segundo parágrafo, a referência às garantias apropriadas ou adequadas e aos meios de obter cópia das mesmas, ou onde foram disponibilizadas [...] (UNIÃO EUROPEIA, 2016).

Imputam, igualmente, aos responsáveis pelo tratamento de dados, o dever de fornecer ao seu titular, para além das informações já mencionadas: o prazo de conservação dos dados pessoais; os interesses legítimos para sua utilização; direito à acesso, retificação, apagamento, limitação, oposição e portabilidade dos dados; retirada de consentimento; apresentação de reclamação dirigida à autoridade de controle, dentre outros.

Aqui resta clara a intenção de fornecimento de uma explicação *ex ante*, sobre a parametrização do sistema artificial e como espera-se que ele opere, o que não se confunde com uma explicação sobre como uma decisão foi tomada em concreto (GOODMAN; FLAXMAN, 2017). Os *considerandos* n. 60[17] e 61[18] corroboram a ideia de uma explicação prévia visto que a notificação deve ser realizada logo quando da coleta dos dados, ou em prazo razoável, informando as implicações que daí advêm.

Por fim, fechando o arcabouço jurídico que dá suporte ao direito à explicação na lei geral de proteção de dados europeia, ao contrário dos artigos retrocitados cujo mote principal é estabelecer prescrições positivas ao responsável pelo tratamento dos dados, o artigo 15 assegura ao titular o direito de obter a confirmação se os seus dados são objeto de tratamento, bem como informações sobre quais dados, a origem da coleta, sua exatidão, apagamento, retificação,

---

[17] *Considerando n. 60*. Os princípios do tratamento equitativo e transparente exigem que o titular dos dados seja informado da operação de tratamento de dados e das suas finalidades. O responsável pelo tratamento deverá fornecer ao titular as informações adicionais necessárias para assegurar um tratamento equitativo e transparente tendo em conta as circunstâncias e o contexto específicos em que os dados pessoais forem tratados. O titular dos dados deverá também ser informado da definição de perfis e das consequências que daí advêm [...] (UNIÃO EUROPEIA, 2016).

[18] *Considerando n. 61*. As informações sobre o tratamento de dados pessoais relativos ao titular dos dados deverão ser a este fornecidas no momento da sua recolha junto do titular dos dados ou, se os dados pessoais tiverem sido obtidos a partir de outra fonte, dentro de um prazo razoável, consoante as circunstâncias. Sempre que os dados pessoais forem suscetíveis de ser legitimamente comunicados a outro destinatário, o titular dos dados deverá ser informado aquando da primeira comunicação dos dados pessoais a esse destinatário. Sempre que o responsável pelo tratamento tiver a intenção de tratar os dados pessoais para outro fim que não aquele para o qual tenham sido recolhidos, antes desse tratamento o responsável pelo tratamento deverá fornecer ao titular dos dados informações sobre esse fim e outras informações necessárias [...] (UNIÃO EUROPEIA, 2016).

finalidade, a quem foram disponibilizados, se servirão de base para tomada de decisões automatizadas e sua lógica adjacente, além das consequências futuras que poderão ser por ele invocado, como bem sinalizado no *considerando* n. 63[19].

> Artigo 15.º (Direito de acesso do titular dos dados) 1. O titular dos dados tem o direito de obter do responsável pelo tratamento a confirmação de que os dados pessoais que lhe digam respeito são ou não objeto de tratamento e, se for esse o caso, o direito de acessar aos seus dados pessoais e às seguintes informações: a) As finalidades do tratamento dos dados; b) As categorias dos dados pessoais em questão; c) Os destinatários ou categorias de destinatários a quem os dados pessoais foram ou serão divulgados, nomeadamente os destinatários estabelecidos em países terceiros ou pertencentes a organizações internacionais; d) Se for possível, o prazo previsto de conservação dos dados pessoais, ou, se não for possível, os critérios usados para fixar esse prazo; e) A existência do direito de solicitar ao responsável pelo tratamento a retificação, o apagamento ou a limitação do tratamento dos dados pessoais no que diz respeito ao titular dos dados, ou do direito de se opor a esse tratamento; f) O direito de apresentar reclamação a uma autoridade de controle; g) Se os dados não tiverem sido recolhidos junto do titular, as informações disponíveis sobre a origem desses dados; h) A existência de decisões automatizadas, incluindo a definição de perfis, referida no artigo 22.º, n. 1 e 4, e, pelo menos nesses casos, informações úteis relativas à lógica subjacente, bem como a importância e as consequências previstas de tal tratamento para o titular dos dados (UNIÃO EUROPEIA, 2016).

Não há o estabelecimento de um limite temporal para o exercício do direito assegurado pelo artigo 15, da GDPR (UNIÃO EUROPEIA, 2016), ao titular dos dados, o que possibilita que determinado questionamento ocorra após a tomada de decisão de forma automatizada, e seja criada a expectativa de que lhe sejam explicadas as suas razões em concreto. Mas, o tempo verbal consignado no dispositivo indica uma orientação futura: "consequências previstas" (*envisaged consequences*), que asseguraria apenas o fornecimento sobre a funcionalidade geral de um sistema de inteligência artificial de tomada de decisão, o que se amolda à precedência histórica da interpretação da Diretiva n. 95/46/CE, regulação que tratava do tema até a entrada em vigor da GDPR.

---

[19] *Considerando n. 63*. Os titulares de dados deverão ter o direito de acessar aos dados pessoais recolhidos que lhes digam respeito e de exercer esse direito com facilidade e a intervalos razoáveis, a fim de conhecer e verificar a tomar conhecimento do tratamento e verificar a sua licitude. Aqui se inclui o seu direito de acessarem a dados sobre a sua saúde, por exemplo os dados dos registos médicos com informações como diagnósticos, resultados de exames, avaliações dos médicos e quaisquer intervenções ou tratamentos realizados. Por conseguinte, cada titular de dados deverá ter o direito de conhecer e ser informado, nomeadamente, das finalidades para as quais os dados pessoais são tratados, quando possível do período durante o qual os dados são tratados, da identidade dos destinatários dos dados pessoais, da lógica subjacente ao eventual tratamento automático dos dados pessoais e, pelo menos quando tiver por base a definição de perfis, das suas consequências. Quando possível, o responsável pelo tratamento deverá poder facultar o acesso a um sistema seguro por via eletrônica que possibilite ao titular acessar diretamente aos seus dados pessoais. Esse direito não deverá prejudicar os direitos ou as liberdades de terceiros, incluindo o segredo comercial ou a propriedade intelectual e, particularmente, o direito de autor que protege o software. Todavia, essas considerações não deverão resultar na recusa de prestação de todas as informações ao titular dos dados. Quando o responsável proceder ao tratamento de grande quantidade de informação relativa ao titular dos dados, deverá poder solicitar que, antes de a informação ser fornecida, o titular especifique a que informações ou a que atividades de tratamento se refere o seu pedido (UNIÃO EUROPEIA, 2016).

Desta forma, a ausência de previsão expressa e clara do direito à explicação na legislação europeia tem gerado dúvidas acerca de seu alcance e abrangência, em razão da linguagem ambígua e pouco clara utilizada. Sua concepção resulta da interpretação sistêmica que pode ser extraída da leitura dos artigos acima citados, especialmente quando analisados em conjunto com os respectivos *guidelines* que, mesmo sem força vinculante, revelam o chamado *espírito da lei*, ou seja, o objetivo do legislador. Nesse cenário, o direito à explicação se limitaria a uma explicação restrita ao funcionamento do sistema e às consequências previstas; da lógica preliminar à tomada de decisão, não alcançando, pois, um direito à explicação posterior de uma tomada de decisão individual e específica, com razões e circunstâncias concretas que possibilitariam uma explicação *ex post*, apesar da existência de argumentos em sentido contrário (LOPES, 2018).

Essa é uma percepção importante de ser compreendida, visto que diante de como é popularmente entendido, o direito à explicação sugere um *olhar para trás*, de modo que se espera que seja explicado como determinada decisão foi tomada, e não como de fato se apresenta, como um *olhar para frente*, que garante ao titular dos dados o direito ao recebimento de informações gerais sobre como o sistema funciona, quais os resultados serão esperados diante dos *inputs* fornecidos, assegurando o direito ao segredo comercial do autor do software, o que torna a questão bem mais complexa do que se mostra inicialmente.

O estudo mais aprofundado acerca do princípio da transparência, revela que o que se considera como direito do titular dos dados ao conhecimento de como se dá o processo, suas consequências e resultados possui particularidades específicas que permitem sua classificação em dois desdobramentos, um direito à informação e um direito à explicação propriamente dito (FLORIDI; MITTELSTADT; WACHTER, 2017). O primeiro deles se encontraria expressamente previsto na legislação europeia, e consiste exatamente nessa informação sobre a funcionalidade do algoritmo, sua lógica adjacente, os resultados esperados, o que se referiu alhures como explicação *ex ante*. De outro lado, o direito à explicação, em essência, seria mais ligado à racionalidade da decisão tomada, à demonstração de quais dados foram utilizados, qual peso que lhes foi atribuído, em uma explicação *ex post*.

Seja qual for a classificação adotada, de se tratarem de níveis distintos de explicação, *ex ante* e *ex post*, ou de um direito à informação e um direito à explicação, ambos derivam do princípio da transparência, de modo que resta indubitável que existem perspectivas diferentes em cada um deles, e que a legislação não é clara sequer sobre a existência de um direito à explicação, quiçá o detalhamento de como este deve se dar. Considerando que esteja assegurado de forma implícita, não há indicativo de qual o seu tipo, e, principalmente, de como deve se dar

a explicação sobre o tratamento de dados, lembrando que, sem o estabelecimento de parâmetros mínimos para essa explicação prometida, certamente restará esvaziado o propósito de assegurar a compreensão, seja do funcionamento ou da racionalidade algorítmica, diante da possibilidade de fornecimento de códigos e explicações técnicas igualmente ininteligíveis.

No Brasil, o tratamento do direito à privacidade tem raízes constitucionais, ligadas à proteção da dignidade e à personalidade humana (MAGRANI, 2019, p. 86) estabelecidas no inciso X[20], do artigo 5º, que trata dos direitos à intimidade, à vida privada e à inviolabilidade de dados. Estes são os fundamentos constitucionais de validade no ordenamento jurídico que, somados com o Código Civil (CC)[21], Código de Defesa do Consumidor (CDC) e Marco Civil da Internet (MCI), dão suporte ao direito à explicação. Além de já ter sido previsto no artigo 5º da chamada lei de cadastro positivo, n. 12.414/2011[22], entende-se que o direito à explicação também pode ser extraído da interpretação do texto base da Lei Geral de Proteção de Dados pessoais (LGPD), de n. 13.709/2018. À exemplo da sua equivalente europeia, apesar de não ter a menção expressa à explicabilidade, parece clara a intenção de o legislador assegurar o acesso ao titular dos dados às informações relativas aos critérios e procedimentos utilizados em uma decisão automatizada, referenciada como explicação *ex ante* ou direito à informação. Contudo, os dispositivos da legislação brasileira consistem mais em normas atributivas de direitos, que asseguram o direito à informação[23], direito à revisão[24] e direito de petição[25] para realização de auditoria, pela autoridade nacional, mas pouco tratam a respeito de normas proibitivas ou impositivas de obrigações aos desenvolvedores de inteligência artificial, como ocorre no ordenamento europeu.

---

[20] Constituição da República Federativa do Brasil de 1988. Art. 5º Todos são iguais perante a lei, sem distinção de qualquer natureza, garantindo-se aos brasileiros e aos estrangeiros residentes no País a inviolabilidade do direito à vida, à liberdade, à igualdade, à segurança e à propriedade, nos termos seguintes: [...] X - são invioláveis a intimidade, a vida privada, a honra e a imagem das pessoas, assegurado o direito a indenização pelo dano material ou moral decorrente de sua violação (BRASIL, 2016).

[21] Lei n. 10.406/2002. Código civil brasileiro. Art. 21. A vida privada da pessoa natural é inviolável, e o juiz, a requerimento do interessado, adotará as providências necessárias para impedir ou fazer cessar ato contrário a esta norma (BRASIL, 2020e).

[22] Lei n. 12.414/2011. Art. 5º São direitos do cadastrado: [...] IV - conhecer os principais elementos e critérios considerados para a análise de risco, resguardado o segredo empresarial (BRASIL, 2019e).

[23] Lei n. 13.709/2018. Lei geral de proteção de dados pessoais. Art. 20. [...] § 1º O controlador deverá fornecer, sempre que solicitadas, informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados para a decisão automatizada, observados os segredos comercial e industrial. (BRASIL, 2020f).

[24] Lei n. 13.709/2018. Lei geral de proteção de dados pessoais. Art. 20. O titular dos dados tem direito a solicitar a revisão de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, incluídas as decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade. (Redação dada pela Lei nº 13.853, de 2019) (BRASIL, 2020f).

[25] Lei n. 13.709/2018. Lei geral de proteção de dados pessoais. Art. 20. [...] § 2º Em caso de não oferecimento de informações de que trata o § 1º deste artigo baseado na observância de segredo comercial e industrial, a autoridade nacional poderá realizar auditoria para verificação de aspectos discriminatórios em tratamento automatizado de dados pessoais. (BRASIL, 2020f).

A explicabilidade se conecta diretamente ao dever de *accountability*, expressão que a sua tradução literal (prestação de contas) não consegue alcançar a significação que lhe é atribuída quando se trata de *transparência qualificada* de sistemas informacionais inteligentes (BIONI; LUCIANO, 2019). A bem da verdade, como visto, tem-se um conjunto de princípios que se entrelaçam como fundamentos de validade do direito à explicação. O princípio da transparência, princípio da *accountability*, e até mesmo o princípio da precaução estão correlacionados com a necessidade de se garantir compreensão e conhecimento à forma de processamento da inteligência artificial.

Para estabelecer o que é abrangido pelo direito à explicação, é preciso compreender o que é uma decisão totalmente automatizada, quais delas afetam a esfera jurídica dos titulares e quais os graus de transparência e explicação poderão ser exigidos. Diante da ausência de previsão expressa na legislação, a academia, indústria e o próprio governo têm tentado suprir essa lacuna com doutrina, autorregulação e orientações (*softlaw*), sem caráter vinculativo, as quais buscam, precipuamente, o desenvolvimento *by design* de uma inteligência artificial explicável, responsável e transparente, que não se apega somente ao direito à explicação como um remédio universal e eficaz para a combater a discriminação e a opacidade inerentes do *machine learning*, posto que, diante das suas diversas inconsistências, esse foco unidirecional pode se apresentar como uma verdadeira falácia (EDWARDS; VEALE, 2017).

2.3 SUPERVISÃO HUMANA SOBRE AS DECISÕES AUTOMATIZADAS

A falibilidade humana é o principal fundamento que justifica o direito ao duplo grau de jurisdição, existente em diversos países como Brasil, Espanha, Itália, Portugal, França, Rússia, Angola, Croácia, Austrália, Chile e Dinamarca, de forma explícita ou não em seus respectivos ordenamentos jurídicos (SÁ, 1999, p. 103). Essa premissa, de o julgador ser falível quando toma determinada decisão, também guarda relação com o inconformismo inerente ao ser humano de não se resignar quando é contrariado, acreditando de maneira mais veemente nas razões que lhe convém.

A simples previsão de que a decisão pode vir a ser revista em uma segunda instância, por si só, já faz com que o julgador se cerque de maiores cuidados para que seja acertada sua decisão, evitando, pelo mesmo motivo, abusos na tomada de decisão. Por mais altruísta que seja o julgador, ele deve ter ciência de que seus atos são verificáveis, sob pena de restar configurada uma indesejada irresponsabilidade monárquica, já antevista por Montesquieu

(1994, p. 78), que alertava que um juiz poderia se tornar um déspota quando tivesse ciência de que não existiria, sobre suas decisões, qualquer tipo de controle.

Desse modo, na medida em que um ordenamento jurídico é estruturado desta forma, além de assegurar a independência do julgador, que deverá julgar conforme seu livre convencimento motivado, satisfaz o inconformismo próprio do ser humano ao lhe proporcionar a possibilidade de um novo julgamento sobre a mesma situação, onde poderão ser expostos e atacados pontos falhos ou premissas equivocadas em que foi apoiada a primeira decisão.

A natural evolução dos processos manuais para os mecanizados, e da automação para a automatização, com a tomada de decisões por softwares de inteligência artificial, diga-se, com uma velocidade e acurácia comprovadamente superior quando comparada às decisões tomadas por seres humanos, trouxe um fator novo de desconfiança sobre as referidas decisões. Cita-se, exemplificativamente, o caso do projeto Victor, do Supremo Tribunal Federal (STF) brasileiro. O sistema, ainda em desenvolvimento, tem como escopo inicial realizar a triagem dos processos, identificando os temas de repercussão geral de forma mais célere que hoje é realizada por técnicos e analistas judiciários. O trabalho de classificação e análise que é realizado por um servidor em um tempo médio de três horas, seria igualmente desempenhado pelo sistema de inteligência artificial em cinco segundos, reduzindo em cerca de dois anos o tempo médio de tramitação dos processos nesta fase de reconhecimento de repercussão geral (ANDRADE *et al*., 2020). A precisão dos testes com o Victor alcançou acurácia de 84% (FREITAS; BARDDAL, 2019). A questão reside exatamente na provável inexistência de um conformismo do indivíduo médio contrariado, que, apegado às razões de sua crença, apenas a fim de ganhar tempo ou, ainda, tentar a sorte uma outra vez, poderá argumentar que estaria dentro dos 16% que compreendem a margem de erro prévia confirmada pelos desenvolvedores, justificando, assim, a sua pretensão de uma revisão da decisão tomada, já que somente seria possível aferir o acerto do sistema acaso realizada uma nova análise. Por óbvio, o sistema não indica os *outputs* falhos que gerou. Por isso, é majorada a sensação de insegurança quando essa decisão é integralmente automatizada, sem existir a possibilidade de uma revisão humana.

Sobre o tema, a lei geral de proteção de dados europeia é restritiva de direitos na medida em que proíbe expressamente que os titulares de dados sejam sujeitos a tratamento de dados por meio de decisões exclusivamente automatizadas, com o regular registro das exceções legais, mas lhe sendo assegurado, outrossim, a revisão por pessoa natural, no artigo 22(3), bem como um direito ao contraditório pleno. Há de se obtemperar, outrossim, que, para além das exceções legais, algumas outras situações também podem ser tratadas por decisões exclusivamente automatizadas: quando a decisão é necessária para celebrar ou executar um contrato, ou se a

pessoa deu o seu consentimento explícito (GALINDO, 2019a). São recomendadas, igualmente, a submissão a decisões exclusivamente automatizadas questões ligadas a atividades burocráticas e repetitivas (MARQUES, 2019), as quais devem se posicionar como ferramenta de apoio à tomada de decisão por um ser humano (CARMO; GERMINARI; GALINDO, 2019).

No entanto, diferente do modelo europeu, onde há expressa previsão de que não deverá ser submetido à decisão exclusivamente automatizada, e poderá exigir a revisão com supervisão humana, tal direito não é assegurado ao titular de dados no Brasil. A LGPD trazia, na redação original de seu artigo 20, *caput*[26], a possibilidade de revisão, por pessoa natural, de decisões tomadas unicamente com base em tratamento automatizado que lhe trouxesse repercussões, aí incluídas aquelas destinadas à definição de seu perfil. Porém, antes mesmo de entrar em vigor[27], por ocasião da Medida Provisória n. 860, de 2018, a redação do *caput* do artigo 20 foi modificada a fim de retirar a menção à pessoa natural, permitindo, assim, que a revisão requerida pelo titular dos dados seja realizada integralmente por um sistema automatizado.

Como já visto, a aparente impressão de neutralidade algorítmica é relativa, para não dizer equivocada, especialmente quando se trata de inteligência artificial que se utiliza de *machine learning*, visto que o algoritmo se desenvolve principalmente com base no *dataset* inserido na máquina, o que poderá, na melhor das hipóteses, perpetuar os estereótipos e vieses enraizados na sociedade. A recorribilidade com a participação humana é, portanto, uma segurança para que seja evitado esse tipo de tratamento discriminatório, vez que no atual estado da arte, a inteligência artificial não está apta a desenvolver uma capacidade de análise crítica que perceba essas obliquidades, e tenha, de fato, uma tomada de decisão neutra.

Então surge o questionamento sobre o porquê das sucessivas alterações legislativas a fim de afastar a obrigatoriedade da supervisão humana no processo de tomada de decisão automatizada. Pontuando cronologicamente a sucessão de eventos, ocorreu inicialmente a publicação do texto base da lei, em 14/08/2018, em que havia a previsão expressa de revisão por pessoa natural, no *caput* do artigo 20. Além disso, a Autoridade Nacional de Proteção de Dados (ANPD), que estava idealizada no texto aprovado pelo Congresso Nacional, teve sua criação vetada integralmente pelo Presidente da República quando da promulgação da lei, com a justificativa de que o processo legislativo teria incorrido em inconstitucionalidade formal.

[26] Lei n. 13.709/2018. Lei geral de proteção de dados pessoais. Art. 20. O titular dos dados tem direito a solicitar revisão, por pessoa natural, de decisões tomadas unicamente com base em tratamento automatizado de dados pessoais que afetem seus interesses, inclusive de decisões destinadas a definir o seu perfil pessoal, profissional, de consumo e de crédito ou os aspectos de sua personalidade. [Redação original] (BRASIL, 2020f).

[27] Entrada em vigor inicialmente prevista para 15 de agosto de 2020, modificada para 3 de maio de 2021, pela Medida Provisória nº 959, de 2020.

Assim, para reorganizar o ordenamento jurídico, foi apresentada a Medida Provisória n. 869, em 27/12/2018, criando, finalmente, a Autoridade Nacional de Proteção de Dados (ANPD), que teria o propósito de ser o órgão responsável pela regulação, controle e fiscalização da aplicação da lei geral de proteção de dados pessoais. No entanto, nesta mesma ocasião, foi modificado o texto do artigo 20, da LGPD, afastando a supervisão humana originalmente prevista, sem qualquer menção a tal modificação na exposição de motivos da medida provisória.

Com a necessidade de conversão em lei pelo Congresso Nacional, referida medida provisória originou o texto da lei n. 13.853/2019. Antes da promulgação do Presidente da República, o texto aprovado pelo parlamento reinseria a obrigatoriedade de supervisão humana removida à revelia do parlamento por ato exclusivo da Presidência da República, inserindo o §3º[28], do artigo 20, a ser realizada conforme regulamentação da autoridade nacional. Essa regulamentação deveria levar em consideração dois critérios: a natureza e o porte da entidade ou o volume de operações de tratamento de dados. O estabelecimento destes critérios surgiu durante os debates no parlamento, a fim de que não restassem inviabilizadas as estratégias e modelos de negócios inovadores das empresas que se utilizam de inovações tecnológicas para otimizar o mercado.

O Congresso Nacional documentou no parecer n. 01/2019 (BRASIL, 2019a), que instrumentaliza os estudos e ponderações sobre a estrutura da legislação discutida, os argumentos que justificavam o retorno da supervisão humana, defendendo que

> com a popularização do uso da Inteligência Artificial e outros mecanismos automatizados para a prestação de serviços e a consequente retirada da pessoa humana, o exercício dos direitos humanos, de cidadania e do consumidor (previstos no art. 2, VI e VII) são dificultados e, por consequência, enfraquecidos. Ademais, a inexistência de humanos dificulta em sobremaneira a interação com controladores por parte de pessoas que possuam deficiência de julgamento ou experiência, o que poderia levar a práticas abusivas.
> Outro ponto a ser ressaltado é que os desenhos dos algoritmos que processam esses dados são baseados em probabilidade e estatística. Como tal, as implementações não englobam o universo dos titulares e seus comportamentos, e sim uma amostra, baseada em intervalos de confiança, erros e desvios padrões naturais dessa ciência. Ademais, assim como as demais ferramentas das Tecnologias das Informações, estão sujeitos a ocasionais incorreções e imprevistos quando executados.
> Ainda neste aspecto, consideramos que a retirada vai de encontro ao disposto no art. 22 da LGPD europeia, o que poderá dificultar o entendimento comercial entre as partes e dificultar a integração comercial e geração de oportunidades e de investimentos.

[28] Lei n. 13.709. Lei geral de proteção de dados pessoais. Art. 20. [...] §3º A revisão de que trata o caput deste artigo deverá ser realizada por pessoa natural, conforme previsto em regulamentação da autoridade nacional, que levará em consideração a natureza e o porte da entidade ou o volume de operações de tratamento de dados. (VETADO) (BRASIL, 2020f).

Porém, o Presidente da República, após a oitiva prévia e concordância dos Ministérios da Economia, da Ciência, Tecnologia, Inovações e Comunicações, da Controladoria-Geral da União e do Banco Central do Brasil, vetou o texto aprovado pelo Congresso Nacional que inseriria o §3º ao artigo 20, da LGPD, sob o argumento de que a submissão à revisão humana de toda e qualquer decisão que tivesse sido baseada exclusivamente em tratamento automatizado contrariaria o interesse público, na medida em que restariam inviabilizados diversos modelos de negócios de empresas que se utilizam de tecnologias de inteligência artificial, impactando negativamente na oferta de crédito aos consumidores, inflação e condução da política econômica monetária (BRASIL, 2019b).

É certo que tais considerações tinham sido ponderadas pelo legislador quando da reinserção da supervisão humana no processo de tomada de decisão automatizada como visto, porém, as conclusões do parlamento não foram compartilhadas pelo chefe do poder executivo, a quem é assegurado o poder do veto para exercê-lo quando convencido de que deve fazê-lo. Assim, a atual configuração legal no Brasil desde sua entrada em vigor é pela supervisão humana no processo de tomada de decisão como mera faculdade do responsável pelo tratamento de dados, que deverá, pelo menos, assegurar que as decisões tomadas de forma autônoma por sistemas de inteligência artificial sejam revisadas, ainda que esta revisão também seja por um sistema de inteligência artificial. Difícil é crer no desenvolvimento de tecnologias distintas e autônomas para que a revisão executada não seja tão somente *pro forma*, com a análise realizada por sistema idêntico, uma segunda vez, com a mesma formatação algorítmica da primeira análise, o que provavelmente geraria resultados iguais.

Com esta tônica dos órgãos governamentais de buscarem uma compatibilização entre o desenvolvimento da inovação e os direitos fundamentais dos indivíduos, em que não há uma expressa garantia de um direito à explicação, ou sequer a definição de parâmetros mínimos para seu exercício efetivo, assim como uma postura comedida em relação ao estabelecimento de mecanismos de salvaguarda à exposição ao enviesamento de dados, como a supervisão humana, parece claro que as questões econômicas têm se sobressaído nesse choque de interesses, o que se mostra ainda mais evidente quando as principais diretrizes têm se resumido a uma tímida formatação de conjuntos de regras meramente orientativos, de caráter recomendatório, que outorgam aos partícipes desta nova sociedade de vigilância de dados (*dataveillance*) o poder decisório de escolher os parâmetros a serem seguidos.

2.4 A NECESSIDADE DE UMA TECNORREGULAÇÃO E AS PRIMEIRAS REGULAMENTAÇÕES

Diante da pouca ou nenhuma política pública governamental organizada capaz de combater os desafios e riscos inerentes de uma maior utilização da inteligência artificial, os quais revelam um alto nível de dificuldade e complexidade para sua reparação, tem se mostrado um caminho viável o estabelecimento de normas voltadas para a prevenção, a serem implementadas quando do desenvolvimento da inteligência artificial, que tem sido chamada de *Explainable Artificial Intelligence* (*xAI*).

A inteligência artificial explicável tem como objetivo a projeção de sistemas inteligentes que possuam diretrizes em seu DNA que, ainda que diminuam sua acurácia, permitam a explicação lógica de como se chegou às decisões. Esta, inclusive, tem sido a estrutura preferida para o desenvolvimento de um sistema de decisões judiciais fundamentadas computacionalmente (CAMARGO, 2019), pois, não basta à sociedade que se alcance o resultado pretendido, que seja tomada a decisão, sem que seja possível conhecer os métodos pelos quais o sistema se valeu. Nem mesmo a transparência total, revelando os códigos e metadados importam, caso não demonstrem as inferências e heurísticas do sistema. Somente desta forma será possível exercer uma *accountability*, que permita o estabelecimento de uma explicabilidade e interpretabilidade, gerando confiança do homem no sistema de inteligência artificial.

Essa transparência qualificada se faz necessária visto que, em razão da sua incapacidade de compreensão de uma relação de causa e efeito, e mais, do próprio tratamento lógico do raciocínio não-monotônico (CELLA; WOJCIECHOWSKI, 2014), os sistemas de inteligência artificial podem se valer de caminhos indesejados para alcançarem os resultados que lhes foram determinados, se distanciando dos objetivos para os quais foram desenvolvidos. É certo que tal preocupação não se faz presente em toda e qualquer situação. Não se imagina que seja objeto de questionamento a recomendação de um produto por um site, ou de um filme por uma plataforma de *streaming*. No entanto, quando passamos a confiar a sistemas de inteligência artificial a direção de veículos autônomos, a realização de diagnósticos de doenças e recomendação de tratamento, avaliação e seleção de trabalhos e pessoas, ou os termos e condições da liberdade de um apenado, a questão se mostra mais delicada. Deve se ter a certeza de que as máquinas fornecerão as respostas certas, pelos motivos certos.

Os riscos de confiar cegamente na heurística desenvolvida pelas redes neurais que se valem de *deep learning* restaram evidenciados nos testes realizados pelo professor de *machine*

learning da Universidade Técnica de Berlim, Klaus-Robert Müller, em conjunto com o Instituto Heinrich Hertz, na Alemanha e da Universidade de Tecnologia e Projetos de Singapura. Na oportunidade, a equipe de cientistas criou um sistema de computador que tinha a capacidade de aferir e quantificar, de forma automatizada, os resultados de um algoritmo de inteligência artificial. O método ficou conhecido como *Layer-wise Relevance Propagation* (LRP)[29]. Com a extensão desse método[30], foi possível a identificação e quantificação de um amplo espectro de comportamento de tomada de decisão desenvolvidos por um programa de *machine learning*. Com esse sistema de análise, foi possível perceber que nem sempre o "raciocínio" desenvolvido pelos programas de aprendizado de máquina se baseava em premissas minimamente razoáveis, não se mostrando tão inteligentes quanto se esperava. As soluções encontradas pelos sistemas inteligentes, se apresentam simplórias e rasas para os humanos. Foi o que verificaram no caso de um programa de inteligência artificial de classificação de figuras que ganhou várias competições internacionais: ele se utilizava mais do contexto em si, do que da própria imagem principal na sua atividade de rotulação; imagens que tinham um objeto rodeado por muita água eram classificadas como "navio", enquanto imagens que apresentavam trilhos eram classificadas na categoria "trem" (LAPUSCHKIN *et al*., 2019).

Também foi possível verificar o funcionamento de sistemas de inteligência artificial baseados em redes neurais, os quais imaginava-se que não estariam sujeitos a tais riscos. No experimento, dois sistemas tinham como objetivo designado a identificação de cavalos em uma vasta biblioteca de fotografias. Enquanto operavam, o sistema de inspeção LRP desenvolvido pelos pesquisadores tinha por objetivo realizar uma observação de segunda ordem sobre o funcionamento dos sistemas de inteligência artificial na sua tomada de decisão de rotulação. Apesar de ambos os sistemas atingirem uma maioria de resultados corretos, foi possível observar que, enquanto um deles se atinha às características do animal que lhe foram compiladas para sua correta identificação, o outro concentrava sua atenção para pixels no canto inferior esquerdo das imagens, onde posteriormente foi percebido que apenas as fotos dos cavalos possuíam um selo de direitos autorais (SAMPLE, 2017). Ambos os sistemas entregaram os *outputs* pretendidos, porém, um deles com base em informações indevidas, razão pela qual resta evidenciada a impossibilidade de uma confiabilidade plena e irrestrita no tratamento de dados pelos sistemas de inteligência artificial.

Isso demonstra que a inteligência artificial não é exatamente o que parece ser. Estima-se que aproximadamente a metade dos sistemas de inteligência artificial se utilizam de

[29] Em tradução livre: propagação de relevância sensível a camadas.
[30] *SpRA*: *spectral relevance analysis*, em tradução livre: análise de relevância espectral.

estratégias do "*Hans esperto*"[31], que mais parecem truques sofisticados do barão de Kempelen, do que inteligência propriamente dita. Nesse sentido, Meredith Broussard, professora da Universidade de Nova Iorque, alinhada com os resultados da pesquisa acima mencionada, sustenta que a inteligência artificial nada mais é do que matemática, e nem todos os problemas podem ser resolvidos pela matemática (LAPUSCHKIN *et al*., 2019).

Observando todos esses fatores de risco, que se entrelaçam em diversas questões éticas relativas à criação em si de sistemas inteligentes, bem como da exata compreensão do que é e como deve ser tratado um sistema de inteligência artificial, os organismos sociais interessados, em especial a academia, indústria e governo, começaram a trabalhar em diversos estudos em busca da definição de princípios éticos, diretrizes e recomendações a serem observadas pelos desenvolvedores de inteligência artificial, os quais, ainda que desprovidos de obrigatoriedade, orientariam como deve ocorrer o processo de criação de sistemas automatizados de forma segura, ética e responsável. Antes de tratar propriamente dos desafios para assegurar que a inteligência artificial opere de forma segura, é preciso entender um pouco mais sobre essa coisa. Dos primeiros estudos que buscaram entender e construir uma base principiológica essencial relativa à existência de um *status* moral na inteligência artificial que merece referência é o trabalho de Nick Bostrom, filósofo contemporâneo, nascido na Suécia, em 1973, que ficou conhecido por seu trabalho a respeito da superinteligência dentre outros. Em 2011 ele escreveu um texto junto com Eliezer Shlomo Yudkowsky, um pesquisador norte-americano de inteligência artificial (BOSTROM; YUDKOWSKY, 2011), sobre os desafios futuros da inteligência artificial relacionados à ética, com uma abordagem voltada também acerca de como devem ser considerados os sistemas artificiais quando avaliados como potenciais sujeitos de direito, assim como nas suas relações com os seres humanos. A possibilidade de um sistema de inteligência artificial reclamar um *status* moral por possuir direitos que lhe são próprios, passa a ser relevante diante do nível de desenvolvimento tecnológico, senciência e sapiência que lhe podem ser atribuídos.

No entanto, difícil considerar que sistemas inteligentes possuam tais características. Assim como uma pedra, certamente os sistemas não possuem qualquer *status* moral, visto que completamente desprovidos de qualquer desses atributos. Pode-se quebrá-la, esmagá-la,

[31] Kluge Hans foi um cavalo que aparentemente teria capacidade de realizar operações aritméticas e outras tarefas mentais. Sua aparente compreensão virou sensação científica nos anos 1900. No entanto, após estudos mais aprofundados de uma equipe multidisciplinar denominada *Comissão Hans*, liderada por psicólogo Oskar Pfungst, em 1907, restou evidenciado que o animal não entendia de matemática, mas teve um excelente treinamento que lhe permitia responder aos movimentos corporais involuntários de seu adestrador, o alemão Wilhelm von Osten. Essa falsa ideia cognitiva ficou conhecida como *efeito do Hans esperto (Clever Hans Effect)*.

pulverizá-la, ou mesmo alterar, copiar, deletar sem qualquer tipo de preocupação com o estado de bem-estar, implicação moral, ou violação de direitos da pedra ou do programa em si. Eventual responsabilidade envolta em tais comportamentos se refere às relações com outros seres humanos ou obrigações, mas não a qualquer tipo de direitos próprios dos objetos. Mas, partindo do pressuposto de que esse *status* moral decorre da existência de algum nível de senciência ou *qualia* (capacidade de experiência fenomênica, de sentir dor e sofrer) e sapiência (atributo ligado à autoconsciência, racionalidade e responsabilidade do agente em si), ignorando discussões sobre o *status* moral de fetos, incapazes ou de pessoas à margem da sociedade, é certo que a inteligência artificial pode vir a apresentar, pelo menos em certo nível, esses atributos, ainda que não tenha uma linguagem ou outras faculdades cognitivas mais avançadas, se assemelhando muito mais a um animal do que a um objeto.

> [...] É errado infligir dor a um rato, a menos que existam razões suficientemente fortes e razões morais prevalecentes para fazê-lo. O mesmo vale para qualquer sistema senciente de IA. Se além de consciência, um sistema de inteligência artificial também tiver sapiência de um tipo semelhante à de um adulto humano normal, então terá também pleno *status* moral, equivalente ao dos seres humanos [...] (BOSTROM; YUDKOWSKY, 2011, p. 7)[32].

Essa ideia é representada pelos dois princípios fundamentais trazidos pelos autores: *princípio da não-discriminação do substrato*, o qual sinaliza que mesmo existindo uma diferença no substrato de implementação entre dois seres que possuem a mesma funcionalidade e experiência consciente, eles compartilharão o mesmo *status* moral. O substrato, assim, somente seria relevante se alterasse a sua sensibilidade ou funcionalidade. É, sem dúvidas, uma visão apoiada em um funcionalismo e behaviorismo que, ao revés de ser rechaçada por tal, se apresenta como um prisma interessante a ser adotado, sob pena de restarem invalidadas bandeiras como a luta contra o racismo. Não se pretende com o princípio buscar a atribuição de consciência à um programa computacional, mas estabelecer que, existindo certo nível de senciência e sapiência, poderia lhe ser atribuído certo *status* moral, independentemente de possuir a pele branca ou negra, ser constituído de carbono, silício ou grafeno.

O segundo princípio proposto no artigo, a fim de rechaçar a artificialidade como fato relevante à outorga do *status* moral, é o *princípio da não-discriminação da ontogenia*. Na linha do primeiro, este princípio ignora a forma como veio a existir o ser como discrímen para outorga

[32] Texto original: *It is wrong to inflict pain on a mouse, unless there are sufficiently strong morally overriding reasons to do so. The same would hold for any sentient AI system. If in addition to sentience, an AI system also has sapience of a kind similar to that of a normal human adult, then it would have full moral status, equivalent to that of human beings*.

de *status* moral, desde que tenham as mesmas funcionalidades e experiência de consciência. Dentro da espécie humana tal princípio é amplamente aceito, não subsistindo a aceitação social de tratamento diferenciado em razão de linhagem familiar ou formas de reprodução humana artificial, como inseminação artificial ou fertilização *in vitro*. Até mesmo a clonagem seria indiferente para a atribuição de *status* moral ao clone humano, esquivando-se das discussões éticas relativas ao uso da técnica. Desta forma, o *princípio da não-discriminação da ontogenia* aplica-se integralmente a sistemas cognitivos integralmente artificiais.

Inegável, no entanto, que a ontogênese traz repercussões aos agentes para com o ser em questão. Os pais de uma criança, por exemplo, são responsáveis por ela e pelos seus atos, não o sendo, entretanto, de outra criança qualquer que possua os mesmos atributos qualitativos que seu filho. A ontogenia, assim, pode estabelecer obrigações sem que isso modifique a situação moral do ser em questão, pelo que, poder-se-ia compatibilizar referido princípio com sistemas artificiais, na medida em que os criadores ou proprietários têm direitos e obrigações especiais em relação aos ditos sistemas.

Importante o registro de que os próprios autores destacam que os princípios propostos não resolvem definitivamente a questão relativa a como deve ser considerada a inteligência artificial, e, ao contrário, fazem surgir novos desafios para além daqueles que já podiam ser antevistos. Contudo, inauguram um pensamento holístico diante do ordenamento jurídico, que passou a ser a partir de então ampliado e lapidado com o seu amadurecimento.

Não poderia a temática de tecnoregulação ser tratada sem que fossem mencionadas as Leis da Robótica, de Asimov. Isaac Asimov (1920 - 1972), foi um escritor e bioquímico norte-americano, que tem uma vasta obra literária acerca da robótica, dentre outras várias temáticas. Asimov ficou muito conhecido tanto pela quantidade, visto que tem catalogado mais de quinhentas obras literárias, quanto pela qualidade de sua redação, enredo e estórias, em sua maioria de ficção científica. Em 1950, em sua obra “Eu, robô”, como resultado de uma crítica a uma peça teatral contemporânea que retratava uma revolta de robôs orgânico-sintéticos em face dos homens, criou o que ficou conhecido como as três leis da robótica, quais sejam: 1ª lei: “um robô não pode ferir um ser humano ou, por inacção, permitir que um ser humano sofra algum mal”; 2ª lei: “um robô deve obedecer às ordens que lhe sejam dadas por seres humanos, exceto nos casos em que tais ordens contrariem a Primeira Lei”; e 3ª lei: “um robô deve proteger sua própria existência, desde que tal proteção não entre em conflito com a Primeira e Segunda Leis”. Posteriormente, em 1983, criou uma quarta regra, que chamou de “lei zero”, prescrevendo que “um robô não pode fazer mal à humanidade e nem, por inacção, permitir que ela sofra algum mal” (TEPEDINO; SILVA, 2019a).

Tratava-se de regras ou princípios criados pelo autor para que os robôs idealizados em suas obras não se rebelassem contra o homem ou contra a humanidade. Apesar de seu explícito caráter ficcional, as referidas leis indiscutivelmente trouxeram diretrizes comezinhas para uma coexistência entre máquinas e seres humanos, as quais são comumente usadas como balizas de regramentos éticos no desenvolvimento dos sistemas, em seus limites internos, bem como nas suas aplicações e interações, estabelecendo limites externos (MULHOLLAND, 2019). De outro lado, é certo que não se mostram bastantes ao contexto atual, visto que, como já dito, o desenvolvimento das tecnologias se desgarrou da ideia antropocêntrica de um robô humanoide, como idealizado por Asimov há mais de meio século atrás, de modo que os algoritmos de inteligência artificial estão em tantas aplicações do dia-a-dia, sem que sequer recebam ordens dos homens, de uma forma então inimaginável (KAUFMAN, 2019, p. 64). Desta forma, assim como os princípios anteriormente referenciados, as leis da robótica de Asimov não têm a pretensão tampouco a capacidade de resolver todos os desafios inerentes à implementação de uma inteligência artificial segura, mas devem, assim como todos os estudos e trabalhos que lhe sucederam, se complementarem a fim de que alcancem estes objetivos desejados por todos.

No espeque de formatar uma relação ética entre a ética e a governança aplicada à inteligência artificial, *digital ethics*, três sãos os principais objetos da discussão: a ética dos dados, ética dos algoritmos e ética das práticas (CARINI; MORAIS, 2019), os quais podem ser visualizados de uma maneira geral em todos os estudos relacionados à tecnorregulação. A ética dos dados cuida da geração, gravação, curadoria, processamento, disseminação, compartilhamento e uso dos dados; a ética dos algoritmos impõe a observância de uma responsabilidade algorítmica de todos os actantes; e a ética das práticas se liga com a governança ética das aplicações.

Peter Asaro (2016), traz uma ponderação interessante quando se pretende outorgar o emprego de violência ou o uso de força letal à um sistema autônomo. A dificuldade na concepção do modelo não se limita apenas à programação da acurácia da máquina para o cumprimento de determinada tarefa, mas, de forma até mais importante que isso, repousa na tomada de decisão que a máquina deverá tomar para que se utilize de suas habilidades. No que se refere ao *design*, a utilização de tecnologias como reconhecimento facial, utilização de dados biométricos, processamento de dados capturados por câmera, acesso a sistemas fechados, cruzamento de informações de sistemas diversos, escaneamento por temperatura, e poder bélico são questões projetáveis de maneira ordinária, e até objetiva. No entanto, estabelecer critérios para aferir a proporcionalidade no uso da força ou sua imprescindibilidade, são desafios que,

diante da profundidade da sua subjetividade são difíceis de orientar a um sistema de inteligência artificial, em que pese possam ser compreendidos por qualquer homem médio.

Tais parâmetros são conhecidos e até documentados, conforme pode-se exemplificar com os princípios básicos estabelecidos pela *United Nations Human Right Council* (UNHRC) e pela anistia internacional para o uso da força e armas de fogo por oficiais policiais, dentre os quais destacam-se: "(i) necessidade de evitar danos físicos graves ou mortes de pessoas; (ii) deve ser aplicado de forma absolutamente discriminada (como derradeiro recurso); (iii) deve ser aplicado de forma proporcional; (iv) deve haver responsabilidade pública" (HARTMAN PEIXOTO, 2020, p. 77).

Especificamente quanto ao uso de robôs, John Tasioulas (2019), do King´s College London apresenta cinco questões éticas a serem observadas: funcionalidade, significado inerente, direitos e responsabilidades, efeitos colaterais e ameaças, que ficaram conhecidas pelo acrônimo FIRST[33]. A primeira delas seria a capacidade de execução de uma tarefa, com certo grau de confiabilidade e sem violar preceitos morais. A aplicação dos robôs deveria ser precedida de uma análise do benefício que seria trazido à humanidade. Alcançada a funcionalidade e vislumbrado um benefício inerente, passa a se fazer necessário o estabelecimento de responsabilidades e direitos, verificando a existência de um *status* moral. Os efeitos colaterais devem ser mapeados nas relações pessoais, de trabalho a fim de evitar discriminações ou restrições sociais. Por fim, as ameaças se apresentam como o ponto mais sensível das questões levantadas por Tasioulas, visto que são diversas as possibilidades de usos negativos de sistemas de inteligência artificial, de ordem pessoal, social, política e econômica.

As questões apontadas por Tasioulas recombinam as vunerabilidades que já vinham sendo identificadas pelos actantes, e sinaliza um dever principal dos sistemas de inteligência, que é o de não enganar. Centra a solução democrática na busca de interconexão do aspecto normativo para a identificação de desvios e ferramentas de prevenção, controle, diagnóstico, neutralização e responsabilização (HARTMAN PEIXOTO, 2020, p. 67), aspectos que são reconhecidos consensualmente como um bom caminho a ser seguido. Interessante o posicionamento de Angela Daly *et al*. (2019), de que para que seja alcançada uma ética eficaz em sistemas de inteligência artificial, deveriam ser observadas duas características: uso de uma fraca normatividade, que não diria de forma absoluta o que é certo e o que é errado, e proximidade com o objeto designado, exigindo estudos interdisciplinares ligados à especificidade do objeto afetado.

[33] Das palavras em inglês *funcionality, inherent significance, rigths and responsabilities, side-efects* e *threats.*

Em 2017 foi realizada a *The Asilomar Conference on Beneficial AI* pelo *Future of Live Institute* (2017), onde foram estabelecidos treze princípios éticos e valores para desenvolvimento e uso da inteligência artificial, quais sejam: 1) *segurança*, prescrevendo que os sistemas de inteligência artificial sejam seguros durante toda sua vida operacional; 2) *transparência de falhas*, que permitam a verificação do motivo de eventuais danos causados pelo sistema artificial; 3) *transparência judicial*, a tomada de decisão realizada por sistemas de inteligência artificial em processos judiciais sejam satisfatoriamente explicáveis e auditáveis por autoridade humana; 4) *responsabilidade*, estabelecendo que os construtores e desenvolvedores de sistemas de inteligência artificial serão responsáveis pela sua modelagem ética, de modo que podem vir a responder por seu uso indevido e ações; 5) *alinhamento de valores*, prescrevendo que os objetivos e comportamentos dos sistemas de inteligência artificial com maior autonomia devem ser alinhados com os valores humanos; 6) *valores humanos*, os ideais de dignidade humana, direitos fundamentais, liberdades e diversidade devem ser compatíveis com o projeto e operação dos sistemas; 7) *privacidade pessoal*, deve ser assegurado aos seus titulares o acesso, gestão e controle dos dados gerados que serão manipulados pelos sistemas; 8) *liberdade e privacidade*, não poderá, em razão dos acessos aos dados pessoais, ser limitada a liberdade das pessoas de maneira irracional; 9) *compartilhamento de benefícios*, os sistemas de inteligência artificial devem compartilhar os benefícios por eles proporcionados ao maior número possível de pessoas; 10) *compartilhamento de prosperidade*, os sistemas de inteligência artificial devem compartilhar a prosperidade econômica por eles criada para toda a humanidade; 11) *controle humano*, a forma de delegação de tomada de decisões a sistemas de inteligência artificial deve ser estabelecida pelos seres humanos; 12) *não subversão*, os sistemas deverão respeitar e melhorar os processos sociais e de cidadania, buscando sempre o bem-estar da sociedade; e 13) *evitar uma corrida armamentista*, deve ser evitada uma corrida armamentista com armas letais controladas por sistemas de inteligência artificial.

Os estudos ligados aos desafios atinentes à inteligência artificial começaram a ganhar multinacionalidade com os fóruns e grupos de pesquisa, a exemplo do *International Technology Law Association* (ITECHLAW) (2019), que reuniu cinquenta e quatro operadores do direito, oriundos de dezesseis nacionalidades diferentes, quais sejam, Alemanha, Austrália, Áustria, Brasil, Canadá, China, Eslovênia, Espanha, Estados Unidos da América, França, Holanda, Índia, Itália, Reino Unido e Suíça, integrantes de vinte e sete escritórios de advocacia diferentes, e apresentou, em 2019, após aprofundados estudos, oito princípios de política relacionados a diretrizes éticas que incentivam o desenvolvimento responsável, e a implantação e o uso da inteligência artificial. Foram apresentados como princípios da estrutura de políticas da

inteligência artificial responsável (*Responsible AI*). São eles: 1) *finalidade ética e benefício social*, dirigido a todas as organizações que desenvolvem, implantam ou usam sistemas de inteligência artificial, assim como a legislação que regula o seu uso, devem exigir que os objetivos sejam identificáveis a ponto de garantir que os propósitos sejam alinhados com objetivos éticos gerais de beneficência (e não maleficência), bem como os demais princípios estruturantes da IA Responsável; 2) *accoutability*, os actantes e legislação correlata devem respeitar os oito princípios da estrutura de políticas da IA Responsável ou outros congêneres, mantendo-se os seres humanos, responsáveis pelos atos e omissões dos sistemas artificiais; 3) *transparência e explicabilidade*, os actantes e a legislação correlata devem assegurar, na medida de suas possibilidades, que o uso seja transparente, com decisões explicáveis; 4) *lealdade e não-discriminação*, os actantes e a legislação correlata devem garantir a não discriminação dos resultados, buscando promover medidas eficazes para garantir a justiça no uso da inteligência artificial; 5) *segurança e confiabilidade*, os actantes e a legislação correlata devem adotar um padrão de design que garanta alta segurança e confiabilidade aos sistemas, bem como limitem ao máximo a sua exposição a terceiros desautorizados; 6) *abertura de datasets usados para desenvolvimento e competição leal (compliance by design)*, os actantes e a legislação correlata devem franquear acesso ao conjunto de dados que podem ser utilizados no desenvolvimento dos sistemas, bem como às estruturas e software de código aberto, respeitando-se o disposto na lei de concorrência/antitruste; 7) *proteção à privacidade*, os actantes e legislação correlata devem se esforçar para garantir o cumprimento das normas e regulamentos de privacidade; 8) *proteção à propriedade intelectual*, os actantes devem tomar as medidas necessárias para a proteção de seus direitos de propriedade intelectual (HARTMAN PEIXOTO, 2020, p. 68).

Com esse conjunto de princípios estruturantes, a organização, em atividade desde 1971, lança diretrizes a serem observadas por todos que desenvolvem, implantam ou usam sistemas de inteligência artificial, alcançando as questões éticas suscitadas por John Tasioulas (2019) e alinhando-se aos princípios éticos estabelecidos na *Asilomar Conference on Beneficial AI*. Essa mesma linha principiológica foi adotada pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE), que fixou em seus princípios e recomendações, o propósito da promoção de uma inteligência artificial inovadora e confiável, que respeite direitos fundamentais e princípios democráticos, objetivando, sempre, desenvolvimento sustentável da humanidade (HARTMAN PEIXOTO, 2020, p. 44).

> Princípios para inteligência artificial da OCDE:
> 1. A inteligência artificial deve beneficiar as pessoas e o planeta, impulsionando o crescimento inclusivo, o desenvolvimento sustentável e o bem-estar;

> 2. Os sistemas de inteligência artificial devem ser projetados de maneira a respeitar o estado de direito, os direitos humanos, os valores democráticos e a diversidade, e devem incluir salvaguardas apropriadas – por exemplo, possibilitando a intervenção humana sempre que necessário – para garantir uma sociedade justa e leal;
> 3. Deve haver transparência e divulgação responsável em torno dos sistemas de inteligência artificial para garantir que as pessoas entendam quando estão envolvidas com eles e possam desafiar os resultados;
> 4. Os sistemas de inteligência artificial devem funcionar de maneira robusta, segura e protegida durante toda a vida útil, e os riscos potenciais devem ser continuamente avaliados e gerenciados;
> 5. As organizações e indivíduos que desenvolvem, implantam ou operam sistemas de inteligência artificial devem ser responsabilizados pelo seu bom funcionamento, de acordo com os princípios acima.
>
> Recomendações aos governos pela OCDE:
> 1. Facilitem o investimento público e privado em pesquisa e desenvolvimento para estimular a inovação em inteligência artificial confiável;
> 2. Promovam ecossistemas de inteligência artificial acessíveis com infraestrutura e tecnologias digitais e mecanismos para compartilhar dados e conhecimento;
> 3. Criem um ambiente de políticas que abrirá o caminho para a implantação de sistemas de inteligência artificial confiáveis;
> 4. Equipem as pessoas com as habilidades de inteligência artificial e apoie os trabalhadores para garantir uma transição justa;
> 5. Cooperem entre fronteiras e setores para compartilhar informações, desenvolver padrões e trabalhar em direção à administração responsável da inteligência artificial.

Além de seus trinta e seis membros, Argentina, Brasil, Costa Rica, Peru e Romênia aderiram às diretrizes que buscam estabelecer padrões éticos que se adaptem com a evolução social e das tecnologias, sem perder de vista o que já existe posto a respeito de privacidade, segurança e responsabilidade.

O reconhecimento governamental da importância da tecnorregulação, inclusive em razão da aderência de organizações como a OCDE, é natural e necessário diante do impacto que as novas tecnologias têm exercido na sociedade de forma globalizada. Assim, a União Europeia estabeleceu diretrizes para a concretização de uma inteligência artificial de confiança, centrada no homem e que respeitasse os valores e princípios europeus, consistentes em: 1) ação e supervisão humanas; 2) solidez técnica e segurança; 3) privacidade e governança de dados; 4) transparência; 5) diversidade, não-discriminação e equidade; 6) bem estar social e ambiental; e 7) responsabilização (HARTMAN PEIXOTO, 2020, p. 45).

Se pretendeu, assim, assegurar a observância aos direitos fundamentais dos indivíduos, exigindo que os sistemas de inteligência artificial funcionem como facilitadores de uma sociedade justa e democrática, garantindo-lhes o respeito e promoção dos direitos humanos fundamentais, bem como a supervisão humana. Da mesma forma, resiliência perante ameaças, solidez técnica que siga uma abordagem preventiva para minimizar danos não intencionais e evitando ataques externos, assegurando, por via de consequência, a integridade física e mental dos seres humanos. Uma governança adequada dos dados deve ser observada a fim de afiançar

a sua privacidade, explicabilidade e transparência dos dados, sistemas e modelos de negócio. Impõe-se, igualmente, aos sistemas, que se estabeleçam de forma a garantir a inclusão e diversidade em todos os processos, em alinhamento ao princípio da equidade, bem como assegurar que a inteligência artificial seja utilizada em benefício, de forma sustentável, de todos os seres humanos, meio ambiente, sociedade e democracia. Por fim, e não menos importante, também ligado ao princípio da equidade, a exigência de criação de mecanismos para garantir a responsabilidade e responsabilização por atos de sistemas de inteligência artificial.

Há, indiscutivelmente, uma previsão muito abrangente e compromissória para os actantes da inteligência artificial, a qual necessitará de um estabelecimento operacional com maior riqueza de detalhes para sua efetiva implementação.

Registra-se, lateralmente, a crítica ao antropocentrismo inato aos princípios e valores ora estabelecidos. Fabiano (HARTMAN PEIXOTO, 2020, p. 53), sustenta que a tecnologia deveria distanciar-se do *dataset* humano quando se fizer necessário. O antropocentrismo, inicialmente até justificado, "diz muito, mas aumenta a agrura, pois laceado. Como ficaria o dano inocente? Sem afastar a relevância humana (afinal é próprio pressuposto do Direito), precisa-se de mais para se efetivamente cuidar-se da ética" (HARTMAN PEIXOTO, 2020, p. 54).

Mundialmente, vários países têm se empenhado para se posicionarem como agentes de sucesso no mercado global de inteligência artificial. O Canadá, por exemplo, tem fomentado o investimento na pesquisa, em projetos como o *Citzen Lab* (HARTMAN PEIXOTO, 2020, p. 116), um estudo sobre o impacto da tomada de decisão por inteligência artificial sobre os direitos humanos, na Universidade de Toronto e o *Cyberjustice Laboratory* (HARTMAN PEIXOTO, 2020, p. 69), da Faculdade de Direito da Universidade de Montreal, que busca levantar o impacto da inteligência artificial nos sistemas de justiça no mundo. Ao final de 2018, foi realizada a *Montreal Declaration for a Responsible Development of Artificial Intelligence*, que, após apresentar três objetivos gerais consistentes em (i) desenvolver uma estrutura ética para o desenvolvimento e uso da inteligência artificial, (ii) orientar a transição digital para um benefício coletivo e (iii) abrir espaço para discussão permanente dos avanços coletivos, desenvolvimento inclusivo e sustentável da inteligência artificial (CANADÁ, 2018), trouxe princípios éticos e valores para concretizar os interesses individuais e coletivos, trazendo, inclusive, orientações interpretativas. Princípio do bem-estar, princípio do respeito pela autonomia, proteção aos princípios da privacidade e intimidade, princípio da solidariedade, princípio da participação democrática, princípio da equidade, princípio da inclusão e

diversidade, princípio do cuidado, princípio da responsabilidade e princípio do desenvolvimento sustentável.

Dentre os referidos princípios, chama atenção pelo ineditismo o princípio do respeito pela autonomia, que preleciona um foco na autonomia das pessoas, com o objetivo de otimizar o controle dos indivíduos sobre suas próprias vidas e seu ambiente. Por certo, está relacionado com outros princípios, como o da equidade, privacidade, inclusão e diversidade, visto que irradia efeitos para combater uma imposição de estilo de vida, vigilância proibitiva, disseminação de informações falsas, primando pelo acesso a conhecimentos diferentes e desenvolvimento de habilidades estruturantes. Destaca-se, ainda, pela nova abordagem proposta, o princípio da responsabilidade, que traz um foco antropológico, vinculando expressamente todas as consequências dos atos provocados por sistemas de inteligência artificial, inclusive e principalmente as tomadas de decisões, aos seres humanos responsáveis. Nesse diapasão, de se registrar a necessidade de confluência do princípio com o da supervisão humana, que deverá ser a figura principal no processo de tomada de decisão, se utilizando da inteligência artificial como apoio. A declaração, na medida em que imputa ao ser humano responsável, abre uma exceção ao referido princípio, posto que prescreve que não seria razoável esta atribuição de responsabilidade às pessoas responsáveis pelo desenvolvimento e uso por ato de sistemas de inteligência artificial, no caso em que o sistema se mostre confiável, e tenha sido usado normalmente (HARTMAN PEIXOTO, 2020, p. 126).

Importante a contribuição do trabalho canadense também no que se refere às lentes interpretativas a serem utilizadas na compreensão e aplicação dos princípios, retirando-se qualquer gradação hierárquica entre eles, os quais devem ser visualizados, interpretados e aplicados sem nenhum tipo de submissão abstrata, limitando seus campos de atuação entre si diante de seu objeto, e permitindo uma evolução interpretativa aberta às modificações naturais das novas tecnologias (HARTMAN PEIXOTO, 2020, p. 128).

Na Alemanha, também em meados de 2018, com o envolvimento de diversos setores do governo, tais como Educação, Pesquisa, Economia, Trabalho e Assuntos Sociais, restou desenvolvida a *AI strategy* (KOCH, 2019), trazendo uma série de medidas em todos os níveis de desenvolvimento, uso, definições e conexões estratégicas relativas à inteligência artificial. Busca, assim como o modelo canadense, fazer da Alemanha e da Europa referências no que diz respeito à inteligência artificial, por meio de um desenvolvimento responsável, que proporcione bem-estar e segurança para a sociedade, integrando-a com as novas tecnologias, sempre observando a ética, cultura e democracia. São apenas três os princípios centrais definidos na *AI strategy* alemã: 1) soberania dos dados; 2) autodeterminação informacional; 3) segurança dos

dados. Todos se relacionam diretamente com uma necessária integridade de *dataset*, focando na proteção ao enviesamento de dados, auditabilidade e compreensão algorítmica. É muito claro o roteiro e visão estratégica da Alemanha, que tem como um de seus objetivos fortalecer a marca "inteligência artificial *Made in Germany*" como sinônimo de qualidade e confiabilidade. Para isso, o documento governamental traça seus objetivos gerais e objetivos táticos específicos, com uma riqueza no detalhamento dos impactos e riscos existentes e futuros (ALEMANHA, 2018).

Nos Estados Unidos, na mesma época, robusteceram-se vários movimentos no sentido de estabelecer uma regulamentação própria. Grandes actantes do mercado privado se organizaram internamente, definindo auto regulações, e com o suporte de associações reuniram-se Amazon, Facebook, Google, IBM, Microsoft, dentre outros, no que ficou conhecido como *Partnership on AI*, buscando produzir uma série de recomendações no sentido de melhores práticas. Em paralelo, o governo federal norte-americano lançou a plataforma *AI.gov*, semelhante à *AI strategy* alemã, que, indicando que se encontra na era da inteligência artificial, conclama os três centros de desenvolvimento (indústria, academia e governo) para que promovam pesquisas a fim de garantir o seu desenvolvimento sustentável. Estabelece como valores a compreensão, confiabilidade, robustez, segurança, e cuidados com a força de trabalho. No mesmo sentido, a Casa Branca assinou a Ordem Executiva n. 13859, apresentando sua estratégia nacional em inteligência artificial, a qual tem estimulado a realização de eventos científicos promovidos pelos setores públicos e privados, tais como *Summit on Artificial Intelligence in Government*, *American AI Iniciative* (ESTADOS UNIDOS DA AMÉRICA, 2019a), *AI for American Innovation* (HARTMAN PEIXOTO, 2020, p. 91). O Departamento de Defesa norte-americano (*Department of Defesense - DoD*) (ESTADOS UNIDOS DA AMÉRICA, 2019b), formatou cinco características essenciais para materialização dos princípios éticos pela inteligência artificial, impondo aos desenvolvedores que os sistemas sejam: 1) *responsáveis*, mantendo sempre a figura do ser humano como responsável pelos sistemas de inteligência artificial; 2) *equitativos*, evitando viés não intencional; 3) *rastreáveis*, assegurando a transparência e auditabilidade; 4) *confiáveis*, segurança contra ataques ao longo de sua vida útil; e 5) *governáveis*, projetados para cumprirem suas funções e evitarem erros, danos e interrupções de forma automática, assim como assegurar sempre o controle humano.

Resta evidenciada tônica característica norte-americana, com uma reafirmação de seus modelos convencionais de negócios, onde busca reafirmar sua liderança e posicionamento estratégico com base na sua dominância econômica, lhe carecendo de uma visão holística mais

associada às demandas éticas, desenvolvimento social e sustentabilidade, tal como tem se apresentado o posicionamento Europeu e canadense.

3 DANOS CAUSADOS POR SISTEMAS DE INTELIGÊNCIA ARTIFICIAL

A inteligência artificial tem sido empregada, de uma maneira geral, com três usos basais: organização de dados, auxílio à tomada de decisão e automação da decisão, nas três esferas de atuação principais, academia (universidades), indústria e governo (STEIBEL; VICENTE; JESUS, 2019). A organização de dados se mostra importante no ambiente acadêmico na medida em que possibilita, *v.g.*, a geração de assistentes virtuais customizados para atender os estudantes, personalizando informações para dados individuais. Para o mercado consumidor, a gestão de dados tem revolucionado a relação entre empresas e clientes, sendo permitido aos fornecedores, em razão do *big data*, um conhecimento sobre seus usuários com uma granularidade ímpar, nunca antes experimentada. Não se trata, pois, de conhecer a opinião geral da coletividade, mas de entender de forma densa e vertical toda uma massa de indivíduos, seus hábitos, preocupações, desejos, localização etc. É possível às empresas conhecerem profundamente os seus consumidores, gerindo seu portifólio de maneira adequada conforme o perfil de cada um, assim como os governos, que, identificando melhor as necessidades e anseios dos cidadãos, podem estabelecer políticas públicas a fim de suprir as deficiências e aspirações da sociedade e do sistema (STEIBEL; VICENTE; JESUS, 2019).

O auxílio à tomada de decisão permite um suporte mais efetivo, com base nos dados, para a criação de modelos descritivos orientados que se apresentam como soluções convincentes e precisas para a sociedade em geral. Nada mais é do que a utilização da máxima capacidade informacional dos dados para dar apoio a uma decisão, aliando o melhor que existe em velocidade de processamento, análise de dados e formulação de estratégias proporcionados pela inteligência artificial, com o julgamento e criatividade humanos, a quem incumbiria a tomada de decisão final. Assim, pode ser verificado um crescente uso da inteligência artificial para identificação de interesse em determinados produtos, ou insatisfação a respeito de determinados serviços, chamando atenção para o desenvolvimento de soluções para que seja finalizada a compra não ultimada, ou corrigida a distorção que gerava a insatisfação. Recrutamento de pessoal é outro exemplo que alia a rotulação de dados e informações, podendo funcionar como suporte à decisão final de contratação dos empregadores, que seria tomada por seres humanos. No entanto, muito ainda se tem a evoluir diante da dificuldade da estruturação da maioria dos problemas estratégicos comumente relacionados às empresas e governos (STEIBEL; VICENTE; JESUS, 2019).

Acerca da automação da tomada de decisão, é crescente o investimento dos actantes nesse tipo de tecnologia, especialmente no contexto da quarta revolução industrial, outorgando

ao sistema inteligente a capacidade de não apenas rotular os dados, traçar o perfil e selecionar as opções possíveis e viáveis para a tomada de decisão humana, mas de tomá-la diretamente, com base em toda a teia de informações mapeada, de forma confiável e apropriada. É o que já se vê em determinadas áreas, sem grandes impactos negativos, como a atuação como moderadores de grupos de estudos, serviços de atendimento aos consumidores, em empresas, recebendo, tratando e filtrando reclamações e queixas, tomando soluções para resolver os problemas apresentados, ou, com treinamento de funcionários públicos (STEIBEL; VICENTE; JESUS, 2019).

No entanto, é uníssono que atualmente não é possível que um sistema de inteligência artificial integre todos os elementos essenciais de uma decisão judicial, por exemplo, a ponto de substituir a atividade humana. Por isso, Samuel Rodrigues de Oliveira e Ramon Silva Costa (2018) sustentam que ela deve permanecer funcionando como ferramenta de auxílio, reservando-se, a atividade de julgar, única e exclusivamente aos seres humanos.

A *startup* canadense *Ross Intelligence*, que se utiliza da plataforma *Watson*, da IBM, desenvolveu um robô-advogado, que, como é anunciado, sugere aos potenciais clientes que se "faça mais do que o humanamente possível. Potencializando advogados com a inteligência artificial" (ENGELMANN; WERNER, 2019). Apesar de esse especificamente se apresentar, neste momento, de forma a não substituir o ser humano, a atribuição de capacidade decisória aos sistemas de inteligência artificial deve ser precedida de uma análise dos impactos que tais decisões podem vir a causar. Não se trata, pois, de obstar os avanços tecnológicos, mas, como já dito, de se estabelecer uma relação de equilíbrio entre a regulação, inovação e criatividade (VERONESE; SILVEIRA; LEMOS, 2019).

Os desafios da implementação da tomada de decisão por sistema de inteligência artificial são ampliados quando a questão é revestida de um caráter ético. Os seres humanos estão sujeitos a decisões deste tipo todos os dias, mas, se mostra sobremaneira complexa a sintetização de tais problemas para que a tomada de decisão ocorra de forma automatizada. O clássico exemplo do túnel em que um carro sem freios em alta velocidade, porém dentro do limite estabelecido para a via, que se depara com uma criança a atravessando inadvertidamente. O motorista tem as opções de: atropelar, e, consequentemente, matar a criança que atravessa a rua em local não permitido, ou, chocar o carro contra o túnel, causando o óbito do motorista e passageiros. É uma decisão difícil que teria de ser tomada pelo motorista humano, e que provavelmente se apresentasse como um problema sem resposta certa. Poderia ele concluir que o certo seria desviar do seu curso normal, pois não seria justo que provocasse a morte da criança, que ainda teria toda uma vida pela frente. Mas, será que seria a mesma conduta se fosse um

idoso ou um criminoso condenado atravessando a rua? É crível até que algum motorista poderia preferir chocar-se contra o muro, ceifando sua própria vida, se fosse um animal atravessando a via naquele momento. Diante de tais alternativas, com elevada carga axiológica, mesmo que se apresentem como hipóteses de uma verdadeira *escolha de Sofia*[34], imagina-se que seriam compreensíveis as razões de decidir qualquer que fosse a decisão. No entanto, conceber uma parametrização de tais decisões para sistemas inteligentes artificiais, se apresenta de maneira muito complexa, pois, teria o desenvolvedor que programar antecipadamente como deveria agir o sistema ao se deparar com uma situação decisória extrema. Isso, invariavelmente, faria surgir desafios acerca do interesse de um usuário para comprar ou utilizar um veículo autônomo que estivesse "programado para lhe matar" caso colocado em determinada situação. E não que se apresente menos difícil a decisão caso a programação fosse para "matar terceiros", priorizando a vida do motorista e passageiros.

Nesse sentido, Renda (2018) utiliza exemplo análogo, de um bonde desgovernado, para pontuar as situações que devem ser enfrentadas, buscando um mapeamento de como podem ser tratadas as questões relativas à responsabilidade civil por dano causado por inteligência artificial.

> (i) dificuldades acerca de programar ou não escolhas éticas e, em caso positivo, como fazê-lo; (ii) problemas advindos da compatibilidade e padronização de comunicação entre dispositivos e tecnologias públicas e privadas, cujos fabricantes variam; (iii) restrições para auditabilidade e verificabilidade das decisões tomadas por essas ferramentas de funcionamento opaco; (iv) limitações dessas tecnologias, ainda que dotadas de transparência, quanto ao enviesamento de suas escolhas fora de padrões éticos; e (v) incertezas para a responsabilização de entes públicos e privados em casos de prejuízos oriundos das mais distintas formas de inteligência artificial.

Seja qual for o propósito do sistema de inteligência artificial, organização de dados, auxílio à tomada de decisão ou automação da decisão, mesmo que utilizadas boas práticas de desenvolvimento e execução a fim de minimizar a introdução de erros, eles se encontram sujeitos à de falhas que resultem na ocorrência de danos, e, pior que isso, podem invariavelmente ensejar à responsabilidade civil sem que haja qualquer falha ou mesmo ato ilícito (ALMADA, 2019).

Desta forma, importante o registro de que se mostra irrelevante o estudo da hipótese de uma programação intencional para a causação de danos. Tampouco para os casos em que

[34] Referência ao filme norte americano homônimo em que a uma mãe chamada Sofia é imposta a tomada de uma decisão difícil que importa em enorme sacrifício pessoal. Ela, uma polonesa que, sob acusação de contrabando, é presa com seus dois filhos pequenos, um menino e uma menina, no campo de concentração de Auschwitz durante a II Guerra. Um sádico oficial nazista dá a ela a opção de salvar apenas uma das crianças da execução, ou ambas morrerão, obrigando-a à terrível decisão.

ocorrer uma falha do sistema, em razão do seu não funcionamento regular. Por certo deverá haver a condenação do agente que programou o sistema que gerou a lesão ou mesmo do responsável pela falha do sistema, sem prejuízo, inclusive, que haja nesses casos um suporte dos mecanismos jurídicos de proteção já desenvolvidos ou daqueles que vierem a ser implementados.

Interessa ao presente estudo, portanto, a situação em que o sistema inteligente causar um dano a outrem, no cumprimento de seu propósito regular e lícito. Esclareça-se que a licitude aqui referenciada é a do propósito, não exatamente do ato causador do dano, o qual, pode ter sido praticado em abuso de direito próprio ou mediante violação de direitos de terceiros.

O dano causado por sistema de inteligência artificial, quando baseado em *machine learnig*, ganha uma faceta dessemelhante, visto que não há uma ação humana direta que estabeleça o nexo de causalidade consigo. Resta ampliada a complexidade da hipótese na medida em que um sistema do gênero é projetado, desenvolvido e otimizado a várias mãos, inclusive pelo próprio usuário em algumas vezes, sendo hercúlea, senão impossível, a missão de identificar uma pessoa física ou jurídica responsável pelo dano causado.

Por isso o estabelecimento de responsabilidade em razão de atos praticados por inteligência artificial ainda é incipiente no âmbito administrativo, penal e civil, se mostrando necessária uma releitura, e talvez até uma readequação da legislação para fins de sua atualização e compatibilização com as novas tecnologias autônomas, surgindo, nesse sentido, duas frentes importantes e necessárias para esse tratamento: a construção de diretrizes a serem implementadas *by design*, isto é, no próprio código de configuração dos sistemas de inteligência artificial como forma de evitar, na medida do possível, a ocorrência de danos, bem como o amadurecimento legislativo a respeito da responsabilidade decorrente de atos praticados por sistemas de inteligência artificial, estabelecendo os parâmetros atinentes à responsabilização de desenvolvedores e usuários que alimentam os sistemas, bem como as espécies de responsabilidade que deverão ser aplicáveis e as novas soluções de tratamento para os casos em que se mostrar inevitável a ocorrência do dano.

De se registrar, outrossim, que além da regulação *by design*, existem outras ferramentas de *compliance* que buscam assegurar uma proteção aos dados e seu tratamento, ao mesmo tempo em que protege os direitos de privacidade das pessoas, das quais cita-se: anonimização e técnicas de pseudonimização, otimização e simplificação das políticas de privacidade, relatório de avaliação do impacto da proteção de dados, a criação de espações de dados pessoais (MASSENO; SANTOS, 2019).

3.1 REGULAÇÃO ÉTICA *BY DESIGN* DE SISTEMAS DE INTELIGÊNCIA ARTIFICIAL

Assente o patamar em que se encontra o discurso ético-filosófico relativo à inteligência artificial, é certo que se apresenta convencionalmente obrigatória para os seus desenvolvedores a inserção de diretrizes éticas, seguras, responsáveis, justas e dignas que funcionarão como verdadeiras travas morais para sua escorreita implementação e aceitação pela sociedade, podendo até se concluir por esta obrigatoriedade genérica legal – e não apenas moral – com base nos princípios norteadores do Código de Defesa do Consumidor. Assim, essa regulação pode ocorrer por dois meios que não se apresentam incompatíveis entre si: (i) a regulação da criação (ética); e (ii) a regulação de parte da matéria prima a ser utilizada pelos sistemas (proteção dos dados pessoais) (VERONESE; SILVEIRA; LEMOS, 2019), não se cogitando a proibição genérica de criação de sistemas autômatos, posto que se apresentaria contraproducente e completamente inócua tal medida. No entanto, ao passo em que se apresenta pacífica a necessidade de uma robusta regulamentação ética e de uma real proteção dos dados pessoais, verifica-se existir uma dificuldade de ordem prática na implementação de tais diretrizes *ex ante*, de uma maneira efetiva.

Nesse sentido, a Europa tem surgido como importante actante em um cenário outrora dominado pelos Estados Unidos da América e China, realizando avanços significativos a respeito do estabelecimento de *guidelines* dirigidos ao desenvolvimento e uso da inteligência artificial. Para isso, tem se valido da confluência de esforços do Conselho da Europa e União Europeia, se fazendo necessária a abertura de um parênteses para distingui-los: o Conselho da Europa é um organismo internacional fundado em 1949, atualmente integrado por diversos países europeus, que tem caráter de direito internacional, entre seus entes autônomos. Já a União Europeia teve sua atual estrutura estabelecida em 1993, por meio do Tratado de Maastricht, apesar de guardar raízes na Comunidade Europeia do Carvão e do Aço e na Comunidade Econômica Europeia, existentes desde 1957. A União Europeia tem caráter de entidade supranacional que possui estrutura de governo e opera de forma integrada com os Estados-membros que a compõem. É, pois, um direito de integração. Atualmente, todos os Estados-membros da União Europeia aderiram ao Conselho da Europa (VERONESE; SILVEIRA; LEMOS, 2019).

O tratamento europeu acerca da matéria lançou luzes ao fato de que os desafios relacionados à segurança no âmbito da internet são estruturais, visto que a proteção jurídica, ordinariamente, é aplicável *ex post*, enquanto as regras técnicas, ligadas ao desenvolvimento dos sistemas de inteligência artificial, o são *ex ante*. Por isso a solução que melhor se apresenta

é aquela consistente na construção de mecanismos que influenciem à produção de programas sensíveis aos valores de ética, segurança e privacidade desejados, desde a sua concepção, do seu desenho inicial (VERONESE; SILVEIRA; LEMOS, 2019), o que se convencionou nominar regulação *by design*.

A ideia da regulação *by design* implica na implementação de toda a tecnorregulação nos sistemas de inteligência artificial desde a sua origem. Este, talvez seja um dos caminhos mais acertados a serem seguidos, posto que já nos encontramos em um momento em que os algoritmos são, em sua maioria, dirigidos por uma regulação inserida em seu código. O que precisa ser alcançado é que seja essa a era do *design*, aliás, a era do *bom design*. No entanto, em que pese se tenha uma série de regulações relativas à internet e privacidade no Brasil, como o Marco Civil da Internet (MCI)[35] e a Lei Geral de Proteção de Dados (LGPD)[36], tem-se uma sobreposição de uma regulação própria e voluntária dos actantes da indústria (*softlaw*) em face das normativas legais, visto que revela uma verve voltada para atender seus interesses econômicos e propósitos comerciais, sem se ater cuidadosamente à observância dos direitos fundamentais ou sequer às regulações específicas.

Diante deste contexto de autorregulação exacerbada, eleva-se a lógica binária algorítmica do "pode/não pode" que se contrapõe diretamente ao modelo de "dever ser" do Estado Democrático de Direito, assim considerado o fundamento de organização jurídica de uma sociedade civilizada. Cria-se, além disso, um grande risco de submeter a sociedade como um todo a uma hipernormatização da inteligência artificial, comandada pela tirania do código (algoritmo) ou ditadura do protocolo (CALDERÓN-VALENCIA; MORAIS, 2020)[37]. E é este que deve ter como ideias fundamentais interrelacionadas a limitação do poder do Estado, o tratamento isonômico e a salvaguarda dos direitos dos indivíduos, que, embora possam ser consideradas apenas utopias, se apresentam como ideais a serem buscados em um Estado Democrático de Direito. Porém, é nítida a discrepância entre os ideais do *Rule of Law* e a autorregulação tecnológica das plataformas digitais, em seus produtos e serviços oferecidos aos usuários. Em que pese se tenha uma série de diretrizes, princípios e regulamentações voltadas para a inteligência artificial, é sobremaneira difícil a aferição de sua observância no funcionamento dos sistemas, visto que o usuário fica adstrito àquilo que foi programado, não chegando ao seu conhecimento, na maioria das oportunidades, eventual filtragem ou retirada

[35] Lei n. 12.965/2014.
[36] Lei n. 13.709/2018.
[37] Fernando Galindo (2019b) se refere a esse fenômeno de superdimensionamento dos códigos algorítmicos escritos pelos programadores ao ponto de tornar refém toda uma sociedade como *datificação*.

de conteúdo em razão de um *profiling* criado por inteligência artificial com base em dados minerados na rede, por exemplo, que são usados para construir serviços personalizados de acordo com as preferências pessoais, interesses, comportamento e localização dos usuários (MASSENO; SANTOS, 2019).

Naturalmente os usuários se condicionam a buscar uma adaptação própria às regras do código ao invés de exigirem uma adaptação das regras do código ao sistema jurídico vigente, com as devidas adequações. Não se trata de aplicar-se o tradicional "dever ser", mas, incutir essa regulação de maneira cogente no *design* e arquitetura dos códigos algorítmicos. Considerando que a ordem jurídica possui mecanismos específicos que importam na admoestação de condutas indesejadas, se apresenta necessário o estabelecimento de uma técnica de regulação cogente e clara para incidir sobre a técnica de regulação dos comportamentos e inovação da tecnologia dos fabricantes e desenvolvedores, concebendo-se a lei como um instrumento de metatecnologia (MAGRANI, 2019, p. 254).

A respeito do papel do Direito, assevera Paul Ohm (2010):

> Se nos preocuparmos com toda a população sendo arrastada irreversivelmente à beira de danos, devemos regular de antemão, porque a esperança de regular após o fato é o mesmo que não regular de forma alguma. Desde que a nossa identidade seja separada da base de dados da ruína por um alto grau de entropia, podemos descansar tranquilamente. Mas, à medida que os dados são ligados a outros dados, e à medida que os adversários diminuem a entropia, cada um de nós logo será lançado à beira da ruína[38].

Aplicações simples e práticas como a adoção de termos de uso com uma linguagem simples e clara, com a mudança da forma de coleta de dados da lógica de "tudo ou nada" para uma coleta granular apenas dos dados essenciais relativos ao produto e/ou serviço são medidas citadas por Caio Augusto Souza Lara (2019, p. 153), para um maior alcance do que chama de uma ética algoritmal, em sua tese que aborda a temática do *acesso tecnológico à justiça: por um uso contra-hegemônico do big data e dos algoritmos*.

Assim, tem-se que para que o direito atue com efetividade como baliza à metatecnologia, as diretrizes, princípios e recomendações éticas devem restar inseridas na modelagem e produção dos sistemas e artefatos de inteligência artificial, garantindo que o sejam sensíveis a valores como privacidade, segurança e ética, especialmente aqueles que envolvem

[38] Texto original: *If we worry about the entire population being dragged irreversibly to the brink of harm, we must regulate in advance because hoping to regulate after the fact is the same as not regulating at all. So long as our identity is separated from the database of ruin by a high degree of entropy, we can rest easy. But as data is connected to data, and as adversaries whittle down entropy, every one of us will soon be thrust to the brink of ruin.*

tomada de decisão. Diversos são os produtos que já têm sido desenvolvidos com essa abordagem de *value sensitive design*, a exemplo de carro lançado pela Toyota, em parceria com a empresa Hino. O veículo foi equipado com um dispositivo apto a medir o teor alcoólico do hálito do motorista, e impedir o seu funcionamento caso verificado que houve a ingestão de bebida alcoólica em teor que supere o limite do tolerável pela legislação (AMOROSO, 2009), o que demonstra a uma segurança *by design*; de igual maneira, um drone que tem inserido em seu código algorítmico a proibição de filmar e fotografar janelas, casas e apartamentos respeita a privacidade e intimidade dos indivíduos (JEROME, 2013); a chamada *smart gun*, uma arma de fogo que foi desenvolvida para cumprir três funções básicas: *identificar atiradores autorizados, autenticar suas credenciais e então liberar o bloqueio para o mecanismo de disparo* (SEBASTIAN, 2016), eleva o nível de segurança da sua utilização, respeitando valores como a vida. Também pode ser verificada a ética *by design* no *bot* Alexa, cuja programação foi realizada de modo a incentivar crianças a falarem expressões como "por favor" e "obrigado". Essa política da desenvolvedora Amazon foi denominada *politeness feature* (MAGRANI, 2019, p. 258), e tem como finalidade auxiliar na educação de crianças através da comunicação interativa com o *bot*, ao qual foi embutido o valor da ética em seu *design*.

Não se deve ignorar, outrossim, a dificuldade para efetivação de determinados valores em sistemas de inteligência artificial. Por tão elementares, muitos deles são intrínsecos aos seres humanos de uma maneira tão fluída e ao mesmo tempo complexa, que se mostra desafiante sua parametrização a fim de que sejam aprendidos e reproduzidos por uma máquina. Como programar um sistema para entender o que é felicidade, lealdade ou justiça? Nick Bostrom (2018, p. 336), ao se debruçar sobre o ponto, correlaciona tal mister à uma especificação de uma função de utilidade. Desta forma, criando-se uma regra de decisão combinada com uma função de utilidade, seria possível determinar um ideal normativo que poderia ser replicado por um sistema de inteligência artificial.

Deve ser considerado também, após superar o desafio de como fazê-lo, que a mera inserção dos valores no processo de criação de sistemas e coisas inteligentes não é o bastante para prevenir comportamentos nocivos. Muitos são os dispositivos que são dotados de aprendizado de máquina, que tornam o seu comportamento imprevisível em razão das diversas possibilidades de interação com os actantes e o ambiente. Nesses casos, imprescindível o estabelecimento de uma política de revisão e filtragem algorítmica periódica para remoção de conteúdo ou condutas indesejadas, sob o ponto de vista dos valores almejados (MAGRANI, 2019, p. 260). Somente assim conseguir-se-á transformar a internet em mais do que tão somente uma mina de dados pessoais não deliberativa, mas em um significativo e extenso espaço público

(FORTES; CELLA, 2016) deliberativo, de convivência constitucional, benéfico e seguro à sociedade, reclamado por uma democracia forte (RABELO; VIEGAS; VIEGAS, 2012).

3.2 DESAFIOS RELATIVOS AOS DANOS CAUSADOS POR SISTEMAS DE INTELIGÊNCIA ARTIFICIAL

Acredita-se que o estabelecimento de uma regulação ética *by design*, que consiga programar uma inteligência artificial sensível a valores seja o melhor caminho para a concepção de um sistema autômato confiável. Porém, como visto, diversos são os desafios para a efetiva implementação e fiscalização a esse respeito, de modo que é imprescindível um pensar além, a fim de definir parâmetros tangíveis para o tratamento da responsabilidade civil em razão de atos produzidos por inteligência artificial.

Antes mesmo de se analisar o arcabouço jurídico já existente, sua suficiência ao tratamento da matéria e necessidade de evolução, é preciso identificar alguns fatores essenciais a esse contexto. Deve-se registrar que os sistemas de inteligência artificial podem se caracterizar tanto como produtos quanto como serviços, conforme conceitos estabelecidos pelo CDC, nos §1º[39] e §2º[40], do artigo 3º, a depender da forma que se apresentam. Como elemento objetivo de uma relação, será classificado como produto quando se tratar de *software* (programa), o qual poderá ser, a depender da sua conjunção com o *hardware*, bem material ou bem imaterial, ou como *produto digital* (PEREIRA, 2019). De outro lado, o sistema de inteligência artificial pode ser integrado a um serviço, de modo que não há a compra do *software* em si, mas apenas o serviço que lhe é prestado, a teor do que se considera serviço na seara consumeirista: uma atividade exercida pelo fornecedor, com habitualidade e profissionalismo, mediante remuneração direta ou indireta, com determinada finalidade, no mercado de consumo (PEREIRA, 2019).

Tais classificações demonstram que o direito do consumidor é perfeitamente aplicável aos sistemas de inteligência artificial, visto que podem, seja por meio do *software* isoladamente ou conjugado com o *hardware* apropriado, ser tanto produtos em si mesmos, quanto instrumentos para a prestação de um serviço de melhor qualidade (PEREIRA, 2019).

---

[39] Lei n. 8.078/1990. Código de defesa do consumidor. Art. 3º [...] §1° Produto é qualquer bem, móvel ou imóvel, material ou imaterial (BRASIL, 2017).

[40] Lei n. 8.078/1990. Código de defesa do consumidor. Art. 3º [...] §2° Serviço é qualquer atividade fornecida no mercado de consumo, mediante remuneração, inclusive as de natureza bancária, financeira, de crédito e securitária, salvo as decorrentes das relações de caráter trabalhista (BRASIL, 2017).

Desta forma, percebe-se que a relação jurídica entabulada entre o usuário e o fornecedor (desenvolvedor/programador/comerciante) pode ter uma natureza consumeirista, caso seja aquele o destinatário final do produto ou serviço, ou natureza cível, no caso de o sistema de inteligência artificial integrar parte de uma cadeia produtiva do usuário, a quem será sempre assegurada a responsabilidade contratual ou extracontratual decorrente do dano causado pelo sistema.

A legislação é clara a assegurar ao adquirente – seja ele consumidor ou não – garantias quando o produto ou serviço apresenta falhas, aqui referenciada de forma genérica o fato e o vício do produto ou serviço. No entanto, em que medida pode ser considerada a ocorrência de uma falha de um sistema de inteligência artificial que entrega exatamente aquilo que lhe foi designado?

Não está a se cogitar o desacerto ou a inexistência de falhas em um sistema que não funciona adequadamente, ou que não ofereça a segurança que ordinariamente se espera dele, como já visto, mas sim daqueles que, ao tomar uma decisão autonomamente, por exemplo, conforme lhe era razoavelmente esperado, causa dano à outrem. Isso porque, os sistemas de inteligência artificial baseados em *machine learning* têm em sua essência a ausência de programação direta e específica por um ser humano a respeito de como deverão ser tratados os dados. O próprio sistema se encarrega de desenvolver e aprimorar o algoritmo com base nos dados de treinamento de *input* e *output* que lhe são fornecidos, criando, na maioria das vezes de forma não translúcida, um meio para que se sejam alcançados os *outputs* a partir dos *inputs* fornecidos, passando, após esse *setup*, a gerar a sua forma própria de tratamento dos dados.

Assim, a preocupação que deve ser considerada por ora não é se a inteligência artificial se rebelará contra a humanidade, tomando posturas contrárias ao que lhe determina o homem. Como já dito, ao que se tem conhecimento, não existe nenhum estudo minimamente próximo de conceber uma inteligência artificial forte quiçá uma superinteligência. O que deve causar preocupação é se a inteligência artificial fizer exata e literalmente o que lhe for determinado. Pedro Domingos traz em sua obra a reflexão sobre o aprendiz – forma a qual se refere aos sistemas de inteligência artificial de *machine learning* – encontrar-se sempre na linha estreita entre a cegueira e o delírio (DOMINGOS, 2017, p. 95). Isso porque, a melhor qualidade do aprendiz pode, contraditoriamente, ser sua maior vulnerabilidade. A constância e a literalidade de uma memória perfeita que não seja corretamente programada, como alude Jorge Luis Borges (1979), em sua obra *Funes, o Memorioso*, poderia ser capaz de reconstruir a forma exata de nuvens no céu em um específico momento do dia, mas poderá encontrar dificuldades para perceber que um cão visto de perfil num minuto é o mesmo cão visto de frente no minuto

seguinte. Por isso o chama de *sábio idiota*, visto que tem a incrível capacidade de se lembrar de tudo, mesmo não sendo exatamente isso o que dele se espera. Assim, para ele, duas coisas somente seriam iguais acaso fossem idênticas em todas suas minúcias. Não se aperceberia que "aprender é esquecer os detalhes e, ao mesmo tempo, lembrar das partes importantes" (DOMINGOS, 2017, p. 95).

Janelle Shane, escritora e cientista de pesquisa com enfoque especial em inteligência artificial, alertou para essa situação em recente palestra no TEDx Talks (SHANE, 2019), trazendo à baila exemplos intrigantes, dos quais citam-se alguns. O primeiro deles trata-se de uma tentativa de que um determinado sistema de inteligência artificial criasse novos sabores de sorvete. Para isso, a pesquisadora se uniu a um grupo de programadores de Kealing Middle School, e forneceu uma base de mais de mil e seiscentos sabores ao algoritmo para ver o que seria gerado. Os resultados foram desastrosos: "recreio de lixo de abóbora", "gosma de manteiga de amendoim" e "doença de creme de morango". De uma forma diferente do que se esperava, a inteligência artificial cumpriu exatamente o que lhe foi designado, porém, certamente não da maneira que alcançasse o objetivo pretendido pelos pesquisadores.

Outro exemplo referenciado por Janelle (SHANE, 2019), foi quando da realização de testes para que um sistema de inteligência artificial construísse um robô que, situado no ponto A, fosse capaz de chegar até o ponto B. Se dependesse de uma algoritmização inserida manualmente por um programador, de certo deveriam ser descritas detalhadamente todas as ações a serem realizadas pelo robô para alcançar o seu destino, passo-a-passo, com um provável atingimento satisfatório do esperado. Contudo, com tais dados inseridos em um sistema de *machine learning*, o robô poderia, ao invés de montar suas peças em forma humanoide, construir uma espécie de torre, encaixando as peças verticalmente, e cumprir seu objetivo de chegar ao ponto B ao cair em sua direção. Mais uma vez, seria alcançado o *output* expressamente requisitado, mas não da forma efetivamente esperada. Pode se dizer que o referido exemplo chega a ser tolo, pois seria óbvio que o sistema deveria montar o robô humanoide e caminhar usando as pernas. No entanto, outros experimentos apontam que o estabelecimento de condições no sentido de que deveria ser utilizada uma forma humanoide, com uso das suas pernas para se chegar ao ponto B, não se mostraram mais exitosos, sendo alcançados resultados em que o robô humanoide, apesar de alcançar tecnicamente seu desiderato, caminhou de forma completamente desordenada, de costas, dando cambalhotas, com auxílio dos braços etc.

Esses são resultados comuns em sistemas de *machine learning*, onde não são fornecidos quaisquer parâmetros relativos à como executar as tarefas, somente os objetivos pretendidos.

Os sistemas devem, por si só, descobrir a forma que irão cumprir seu desígnio, por tentativa e erro. Tem-se, portanto, a hipótese em que uma falha no funcionamento do código de programação pode vir a causar um dano, podendo este ser considerado um sistema autônomo defeituoso, e outra hipótese, de um dano vir a ser causado por um sistema autônomo não defeituoso.

É possível perceber, assim, que se apresenta muito importante a inteligibilidade e não opacidade dos algoritmos para que sejam corrigidas distorções, que, ao fim e ao cabo, não podem ser consideradas falhas do sistema de inteligência artificial, mas sim uma falta de explicabilidade do ínterim desejado para o atingimento dos *outputs*.

De se ressaltar, também, que os sistemas de inteligência artificial têm se tornado cada vez mais complexos, sendo ampliado o seu grau de imprevisibilidade na medida em que são formatados para uma maior influência do ambiente no seu desenvolvimento.

> No mesmo grau em que a influência do criador sobre a máquina diminui, a influência do ambiente operacional aumenta. Essencialmente, o programador transfere parte de seu controle sobre o produto para o ambiente. Isso é particularmente verdadeiro para as 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 (usuários) e situações, normalmente não será possível prever ou controlar a influência do ambiente operacional[41] (MATTHIAS, 2004, p. 182).

Sistemas de inteligência artificial são construídos com muitas camadas de desenvolvimento do software. Geralmente não se trata de um produto pronto e acabado que pode ser colocado em uma prateleira. Para além das muitas mãos que já integraram a fase de criação e programação inicial, a inata característica de auto otimização e automodificação do *machine learning* e principalmente do *deep learning* conforme o ambiente em que se encontra, torna a questão da responsabilidade ainda mais difusa. Há um distanciamento cada vez maior da programação originalmente estabelecida em razão da capacidade de absorção de novas inferências e conhecimentos de forma não linear, oriunda de várias fontes e formas, e essa lacuna tem sido preenchida pelo aprendizado obtido junto ao próprio consumidor (usuário) ou ambiente, entrelaçando a participação destes actantes, gerando um novo campo de

[41] Texto original: *In the same degree as the influence of the creator over the machine decreases, the influence of the operating environment increases. Essentially, the programmer transfers part of his control over the product to the environment. This is particularly true for machines which continue to learn and adapt in their final operating environment. Since in this situation they have to interact with a potentially great number of people (users) and situations, it will typically not be possible to predict or control the influence of the operating environment*.

responsabilidade civil a ser explorado sobre essa figura repaginada do *prosumer* (*ou prossumidor* em português)[42].

No ordenamento jurídico brasileiro, a responsabilidade civil do fornecedor está diretamente ligada à ocorrência de um vício ou fato do produto ou do serviço, o que pode não restar claramente identificável em um sistema de inteligência artificial que causa dano a outrem (FERREIRA, 2019).

Desta forma, como conceber a responsabilização em razão de um dano causado em que não houve falha do sistema de inteligência artificial? Deve ser o fornecedor responsabilizado pela forma autônoma e imprevisível que o sistema de inteligência artificial tratou os dados para alcançar o *output* desejado? Ou, ainda, como responsabilizar o criador do sistema de inteligência artificial por um dano causado em razão de ato resultante do seu desenvolvimento junto ao próprio usuário? Enrico Roberto e Dennys Camara (2018) argumentam que "devemos entender que uma decisão autônoma, por parte de um sistema de inteligência artificial, é característica esperada e desejada desse tipo de sistema, e que equipará-la a um defeito seria distorcer a letra da lei". Para solver mais esse desafio, tem se apoiado na previsibilidade esperada de tais sistemas.

Países como a Itália e França possuem experiências normativas com a positivação da previsibilidade como um dos requisitos para a configuração da responsabilidade civil (TEPEDINO; SILVA, 2019a), especificamente no artigo 1.225 do código civil italiano[43] e artigo 1.231-3, do código civil francês[44], muito embora haja certa controvérsia doutrinária a esse respeito.

A previsibilidade da conduta pode ser um caminho para alcançar o responsável por dano causado por sistema de inteligência artificial, no entanto, a depender do tipo de sistema e do dano causado, pode se apresentar insuficiente a análise da previsibilidade da ação ou omissão causadora do dano. A regulação *by design* funciona como uma baliza norteadora para estabelecer uma *blacklist* de ações possíveis e antevistas, bem como uma *whitelist*, contendo condutas permitidas, alinhada aos valores inseridos no DNA do sistema de inteligência artificial. Contudo, no ordenamento jurídico brasileiro, a questão da (im)previsibilidade se apresenta como um falso problema (TEPEDINO; SILVA, 2019b), diante da ausência de

[42] Neologismo resultante da conjugação das palavras *producer* e *consumer*, onde o consumidor é parte ativa no processo produtivo do mercado.
[43] Código civil italiano. Art. 1.225. Se o inadimplemento ou a mora não decorrer de dolo do devedor, a indenização é limitada ao dano que se poderia prever ao tempo em que surgiu a obrigação (tradução livre).
[44] Código civil francês. Art. 1.231-3. O devedor só é responsável pelos danos que foram previstos ou poderiam ser previstos quando da conclusão do contrato, exceto quando a inexecução for devida a uma falta grave ou dolosa (tradução livre).

previsão expressa que insira a previsibilidade como elemento necessário para a caracterização da responsabilização, a discussão se apresenta vazia, posto que, como já visto, o eixo gravitacional do sistema de responsabilidade civil brasileiro foi deslocado para buscar uma reparação maior da vítima, ao invés de uma punição do causador do dano (MULHOLLAND, 2019), devendo a questão ser resolvida no "âmbito da causalidade e da imputabilidade daí decorrente, a partir da alocação de riscos estabelecida pela ordem jurídica ou pela autonomia privada" (TEPEDINO; SILVA, 2019b, p. 75). Claro que não se trata de uma irresponsabilidade geral e irrestrita que pretende apenas reestabelecer o *status quo* e indenizar a vítima do dano, mas, sem dúvidas, é certo que não há um escorreito aprofundamento dos institutos para que seja corretamente identificado o real causador do dano.

Consoante restará demonstrado no tópico específico, a legislação brasileira tem seu sistema de responsabilidade civil com os elementos bem delineados, e que independem da verificação de uma falha do sistema ou da ausência de previsibilidade do dano para que haja a atribuição de responsabilidade. Em ocorrendo o dano, decorrente de uma ação ou omissão perpetrada por sistema de inteligência artificial que importe concomitantemente na violação de direitos, sem que haja qualquer dos casos de rompimento do nexo causal, restará configurada a hipótese de dano indenizável, devendo ser investigada a situação concreta para a identificação do responsável pelo dano. Pretende-se trazer à lume as teorias de responsabilidade civil aplicáveis à inteligência artificial nos próximos itens.

3.3 APLICAÇÃO DA RESPONSABILIDADE CIVIL EM RAZÃO DE DANOS CAUSADOS POR INTELIGÊNCIA ARTIFICIAL

A aparente lacuna na disciplina da responsabilidade civil a respeito do tratamento das novas tecnologias tem induzido a proliferação de proposições doutrinárias voltadas para a criação de uma disciplina própria, específica para o chamado direito da robótica (TEPEDINO; SILVA, 2019b) ou ciberdireito, o que sempre foi objeto de fortes críticas, tal como a do juiz Frank Easterbrook, que afirmara que a disciplina era tão útil quando o "direito do cavalo" (FORTES, 2015, p. 50). Porém, após um aprofundamento nas questões que possibilite uma redução da sua complexidade, é possível perceber que, salvo determinadas situações em que se apresenta imprescindível uma inovação legislativa, o ineditismo das questões suscitadas pelas novas tecnologias não importa necessariamente no ineditismo das soluções jurídicas a serem aplicadas (TEPEDINO; SILVA, 2019b).

Entretanto, diante das várias classificações que se mostram ordinárias no mundo jurídico, a segregação de uma específica para o ciberdireito pode se apresentar como um campo de análise interdisciplinar entre o direito e as novas tecnologias da informação e comunicação (FORTES, 2015, p. 55) quando se trata da análise destes *novos danos* causados por sistemas de inteligência artificial que não mais se limitam à substituir a força braçal humana, e passaram a desenvolver atividades cognitivas (AGUDO; TEIXEIRA, 2018).

O ordenamento jurídico brasileiro tem a responsabilidade civil estruturada a partir do artigo 927, do código civil[45], que estabelece a obrigação de indenizar àquele que causar dano à outrem, por ato ilícito. Fixa o que configura o cometimento de ato ilícito como a conduta do agente, omissiva ou comissiva, que, violando direitos, venha a causar dano à outrem, seja por negligência ou imprudência, assim como o exercício abusivo de um direito inicialmente legítimo. Nessa perspectiva, traz como elementos para a caracterização do dano indenizável: a conduta do agente, o dano propriamente dito e o nexo de causalidade. A conduta é a ação ou omissão voluntária do agente, o dano é o prejuízo experimentado pela vítima, seja ele de ordem material ou moral, e o nexo de causalidade é a relação entre a conduta perpetrada pelo agente e o dano havido, no que se refere à sua ocorrência, ou seja, é o liame que correlaciona o dano como consequência da conduta. A responsabilidade pode ser objetiva, oportunidade em que independerá da verificação de culpa do agente, ou subjetiva, quando restará necessária a aferição do dolo (vontade) do agente no cometimento da conduta para a causação do dano, sendo essa última a regra geral da legislação civil (ALBIANI, 2018).

O parágrafo único do artigo 927[46], do código civil é o dispositivo que faz a inserção relativa à necessidade da verificação de culpa para a responsabilização do agente (responsabilidade subjetiva), estabelecendo que será dispensável a sua comprovação nos casos em que a lei assim definir, bem como quando a atividade ordinariamente exercida pelo agente oferecer risco para os direitos de outrem por sua natureza (responsabilidade objetiva). No âmbito do direito do consumidor, a lei n. 8.078/1990 adotou, com exceções ressalvadas, a responsabilidade objetiva como regra geral tanto para a prestação de serviço quanto para o produto fornecido, pelo que também afasta a necessidade da aferição da culpa do fornecedor que causou o dano, para fins de sua responsabilização perante o consumidor e consumidor

[45] Lei n. 10.406/2002. Código civil brasileiro. Art. 927. Aquele que, por ato ilícito (arts. 186 e 187), causar dano a outrem, fica obrigado a repará-lo (BRASIL, 2020e).
[46] Lei n. 10.406/2002. Código civil brasileiro. Art. 927. [...] Parágrafo único. 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, 2020e).

equiparado (*bystander*)[47], sendo interessante pontuar que poderá ser considerado como consumidor em sentido estrito o usuário, pessoa física ou jurídica, que adquirir ou utilizar produto ou serviço que se valha de sistema de inteligência artificial (XAVIER; SPALER, 2019).

Desta forma, tem-se como regra geral a possibilidade de que seja essa responsabilidade extracontratual subjetiva ou objetiva, a depender da estrutura da relação e objeto envolvido.

Outra observação que se correlaciona com o já expendido é a previsão de que somente poderá ser responsabilizado o agente no caso de a sua conduta se configurar como ato ilícito, nos termos dos artigos 186[48] e 187[49], do mesmo código, que ratifica a proposição anterior de que não deverá ser responsabilizado o agente caso inexista falha atribuível ao sistema de inteligência artificial, visto que ausente qualquer ilicitude de sua conduta. A ausência de responsabilidade do agente inocente neste caso pode ser até correlacionada com alguma das excludentes da ilicitude, posto que quando ele atua em legítima defesa, não é considerado ilícito o ato, assim como quando exerce regularmente seu direito ou cumpre um dever legal.

O artigo 186, do Código Civil estabelece que comete ato ilícito aquele que, por ação ou omissão, violar direito e causar dano à outrem, e, quando conjugado com o artigo 927 do mesmo diploma legal, firmam a responsabilidade civil do causador do dano. Quando o ato praticado for realizado em razão de uma conduta proposital e deliberada, utilizando a inteligência artificial como meio para a prática da lesão, devendo o responsável responder pelo dano.

Diversa é a hipótese em que a conduta lesiva não decorre diretamente do projeto ou especificação, mas sim do treinamento ou desenvolvimento do sistema. Neste caso, deve ser identificado o responsável a partir dos registros do desenvolvimento. Em não havendo ação deliberada, mas sim uma conduta previsível que poderia ser evitada, responderão os projetistas, que deveriam ter agido de forma a eliminar esta possibilidade (ALMADA, 2019).

Situação semelhante ocorre quando se trata de abuso de direito, também previsto como ato ilícito, no artigo 187, do Código Civil. Nesse caso, a conduta em si é permitida, porém, quando de seu exercício, a forma praticada excede aos limites socialmente aceitáveis para tal mister. Seria o caso, por exemplo, de um sistema de inteligência artificial que se utilizando de

[47] Lei n. 8.078/1990. Código de defesa do consumidor. Art. 17. Para os efeitos desta Seção, equiparam-se aos consumidores todas as vítimas do evento (BRASIL, 2017).
[48] Lei n. 10.406/2002. Código civil brasileiro. Art. 186. Aquele que, por ação ou omissão voluntária, negligência ou imprudência, violar direito e causar dano a outrem, ainda que exclusivamente moral, comete ato ilícito (BRASIL, 2020e).
[49] Lei n. 10.406/2002. Código civil brasileiro. Art. 187. Também comete ato ilícito o titular de um direito que, ao exercê-lo, excede manifestamente os limites impostos pelo seu fim econômico ou social, pela boa-fé ou pelos bons costumes (BRASIL, 2020e).

ferramentas de identificação e reidentificação expõe a identidade de ativistas de minorias que operam de maneira anônima para evitar ataques no mundo off-line (ALMADA, 2019).

Tanto na hipótese de ato ilícito quanto de abuso de direito, o foco principal é na hipótese em que não há uma intenção deliberada de nenhum dos actantes envolvidos no projeto, treinamento e desenvolvimento do sistema, mas, mesmo assim, há a causação de danos.

Para tais casos, o ordenamento estabelece a responsabilidade por ato próprio do agente, assim como por ato de terceiros, quando vinculados juridicamente ao agente. É o que se observa quanto aos menores, tutelados ou curatelados cuja responsabilidade é imputada, respectivamente, aos pais, tutores ou curadores, assim como empregadores respondem pelos atos dos seus prepostos e pessoas jurídicas de direito público pelos atos dos agentes públicos. Interessa particularmente o estabelecimento de responsabilidade em razão de fato de animais e coisas, posto que sistemas de inteligência artificial não possuem, em princípio, personalidade jurídica[50], razão pela qual esse parece ser um indicativo seguro para que seja trilhada a responsabilização em razão de atos praticados por sistemas de inteligência artificial. Basta, neste diapasão, compreender quem seria o agente com a vinculação jurídica sobre o referido sistema, que arcaria com tal responsabilidade. Assim, a ela seria imputada àquele que estivesse sob guarda do animal ou coisa, e de forma objetiva, ou seja, independente da aferição de culpa (GONÇALVES, 2019, p. 63).

De se ressaltar, entrementes, que a utilização da palavra "fato" e não "ato" pela doutrina quando se trata de fato de coisas já diz muito a respeito da leitura que o ordenamento jurídico brasileiro faz a respeito dos referidos eventos, demonstrando que a responsabilidade, seja ela civil, penal ou administrativa, é sempre ligada à uma atuação humana. Nessa linha, Caio Mário (2018, p. 138) e José de Aguiar Dias (1994, p. 161) se posicionam de forma contrária à expressão *fato das coisas*, posto que uma coisa *inanimada* não seria capaz de praticar um fato. Os casos em que são assim classificados, sempre dependem de uma ação humana anterior, como salientam os Mazeaud (1955 *apud* PEREIRA, 2018), ao explicarem que "quando uma caldeira explode, dizem eles, é porque o homem acendeu o fogo; quando o automóvel atropela o pedestre, é porque o motorista o pôs em marcha". Por isso, Pablo Stolze Gagliano e Rodolfo Pamplona Filho (2019, p. 236) aderem à utilização de uma nomenclatura diferente, qual seja: "responsabilidade pela guarda de coisas inanimadas", que enfatiza o comportamento humano que gera sua responsabilização e não o fato em si, na linha do sustentado por Sérgio Cavalieri Filho (2000, p. 123),

[50] Ressalvado o caso do robô humanoide Sophia, que teve cidadania reconhecida na Arábia Saudita, que será abordada adiante.

> a vida moderna colocou à nossa disposição um grande número de coisas que nos trazem comodidade, conforto e bem-estar, mas que, por serem perigosas, são capazes de acarretar danos aos outros. Superiores razões de política social impõem-nos, então, o dever jurídico de vigilância e cuidado das coisas que usamos, sob pena de sermos obrigados a repararmos os danos por elas produzidos. É o que se convencionou chamar de responsabilidade pelo fato das coisas, ou como preferem outros, responsabilidade pela guarda das coisas inanimadas.

É adotada a expressão *ato praticado* por sistemas de inteligência artificial no presente estudo, ao invés de *fato da coisa*, diante das circunstâncias relativas à essência da inteligência artificial, *machine* e *deep learning*, que lhe denota a existência de certa intencionalidade, senão completa, ao menos parcial.

Para que seja estabelecido o nexo da conduta e dano, a estrutura jurídica traz diversas classificações acerca da culpa e dolo, os quais podem ser aplicáveis aos atos praticados por sistemas de inteligência artificial. Tanto o dolo, que consiste na vontade deliberada de cometer o ato, quanto a culpa, assim considerada a falta de diligência, em suas mais variadas classificações: *in eligendo* (má escolha), *in vigilando* (ausência de fiscalização), *in comittendo* (decorrente de uma ação), *in omittendo* (decorrente de uma omissão), podem servir de base para a responsabilização do agente responsável pelo sistema autômato causador do dano.

Outro ponto a ser mencionado é que foi corrigida uma falha da legislação no que diz respeito à concomitância da violação do direito e o dano experimentado pela vítima. O código civil de 2002 modificou a conjunção "ou" prevista no artigo 159[51], do código civil de 1916, que possibilitava alternativamente que houvesse apenas a violação de um direito "*ou"* a ocorrência de um dano. No novo texto, estabelecido pelo artigo 186, é necessário que ocorram ambas as situações: violação do direito "*e"* o dano a outrem, para que seja configurado o ato ilícito.

Com tais bases pode se verificar que, com algum esforço, a estrutura do ordenamento jurídico brasileiro é bastante para dar uma resposta ao problema da responsabilização civil em razão de ato praticado por sistemas de inteligência artificial. Para isso, parte-se de um caminho já conhecido pelos tribunais pátrios a respeito de danos decorrentes do uso da internet. A compreensão dominante é de que a utilização de uma nova plataforma, em nada altera os postulados jurídicos aplicáveis à espécie, de modo que, devem ser identificados, na medida do possível, os responsáveis para viabilizar o ressarcimento às vítimas. Até mesmo um certo equilíbrio entre o avanço da tecnologia e a imposição de rígidas regras às empresas atuantes no

[51] Lei n. 3.071/1916. Código civil brasileiro. Art. 159. Aquele que, por ação ou omissão voluntária, negligência, ou imprudência, violar direito, ou causar prejuízo a outrem, fica obrigado a reparar o dano. A verificação da culpa e a avaliação da responsabilidade regulam-se pelo disposto neste código, arts. 1.521 a 1.532 e 1.542 a 1.553 (BRASIL, 1916).

mercado foi estabelecido, não sendo determinadas obrigações que viessem a inviabilizar a natural e desejada evolução de mercado.

Então, como se identificar o responsável por ato ilícito na internet para fins de responsabilização? Esta parece ser a questão mais simplória e óbvia. A primeira a ser acionada deverá ser a pessoa jurídica que detenha as informações necessárias para identificação do usuário que produziu o ato, e que pode promover a imediata retirada do conteúdo do ar (LASPRO; CARBONAR, 2020). Sobre o tema, o Superior Tribunal de Justiça, por sua Terceira Turma, entendeu que provedores de acesso à internet e plataformas não poderiam ser responsabilizadas por conteúdo gerado por usuários, não lhe sendo determinado o seu controle e verificação prévia, devendo, outrossim, na medida em que cientificadas da ilicitude de qualquer conteúdo, atuar de forma diligente para retirá-lo do ar imediatamente, identificar e informar o responsável pela sua criação (BRASIL, 2012). Tal decisão já mostra uma evolução quando analisado o posicionamento da Segunda Turma do Superior Tribunal de Justiça que, com uma clara influência na teoria do *deep pocket*, adiante abordada com mais vagar, imputou responsabilidade ao provedor de internet em razão de conteúdo veiculado por um de seus usuários, visto que aquele viabilizou tecnicamente a produção do conteúdo, auferiu lucro com a disseminação de informações ofensivas na rede, e por isso estimulou o engajamento do público em sua plataforma (BRASIL, 2010).

Referidos posicionamentos podem e devem ser aplicados aos sistemas de inteligência artificial, posto que, conjugados com uma regulação *by design*, poder-se-ia ser determinado às empresas desenvolvedoras que estabelecessem *travas morais* e uma rastreabilidade a fim de identificar se o ato perpetrado foi realizado em razão de uma programação prévia, em que o responsável seria o desenvolvedor, ou fruto de uma otimização guiada pelo próprio usuário, que deveria ser, neste caso, o responsável pelo dano causado. Ao fim e ao cabo, não sendo possível a identificação do real responsável, tal ônus recairia sobre o desenvolvedor diante da aparente deficiência dos valores inseridos no sistema de inteligência artificial, ou mesmo em razão do benefício econômico que teve proveito.

Sem que se pretenda um exercício regressivo infinito, um ponto de partida interessante para identificar o causador do dano é verificar quem tinha em seu controle o poder de evitar a sua ocorrência. Além disso, o risco da atividade também deve ser considerado para a apuração da responsabilidade em razão de ato praticado por sistemas de inteligência artificial, posto que, o peso de outorgar a tomada de decisão a um sistema que realizará a indicação de um filme é infinitamente inferior àquele envolto na automatização da direção de um veículo ou um avião.

Os valores e riscos relacionados são sobremaneira elevados em um sistema autômato que tem o condão de expor vidas de uma forma que mesmo a inserção de uma supervisão humana não teria efeito preventivo no caso de uma falha. Assim também o será quando utilizada a inteligência artificial para lidar com sistemas nucleares, por exemplo. É inegável a vinculação jurídica entre o desenvolvedor e o sistema desenvolvido, tal qual o é o vínculo entre pai e filho ou um indivíduo e um animal de sua propriedade. O grande diferencial da caracterização do risco da atividade para fins de responsabilização, é a desnecessidade da demonstração de culpa ou negligência. Isso porque o código civil de 2002, corrigiu a distorção de seu predecessor e se alinhou ao estabelecido pelos seus equivalentes alemão e francês, retirando a equivocada previsão do artigo 1.523[52], de autoria do Senado Federal quando da tramitação no Congresso Nacional. A previsão do novo texto brasileiro trouxe até uma situação mais robusta do que as rotas traçadas pelas legislações estrangeiras alhures referenciadas. O centenário código francês trazia a responsabilidade *juris tantum*, admitindo, portanto, "escusa no caso em que possam provar lhes tenha sido, moral e materialmente, impossível evitar o evento danoso, não podendo isentar-se da responsabilidade mediante prova de não culpa" (GONÇALVES, 2019, p. 150), já o código alemão estabelece que "a responsabilidade indireta não é tão grave, porque há a possibilidade de o demandado eximir-se, alegando que empregou diligência para evitar o ocorrido" (GONÇALVES, 2019, p. 150).

Curioso observar, quando se discorre acerca de riscos e vulnerabilidades, que o uso da inteligência artificial em diversas aplicações, acaba por criar uma situação paradoxal de elevar a segurança e acurácia da tarefa realizada, mas, em paralelo, expor essa mesma atividade a um agente completamente improvável caso inexistente a tecnologia: ação de *hackers* (TEPEDINO; SILVA, 2019b).

De certo lembrar, ainda, que a responsabilidade vicária[53] guarda relação direta com a teoria do risco, visto que prescreve que aquele que usufrui dos *cômodos*, deve suportar os *incômodos* que lhe são inerentes (GAGLIANO; PAMPLONA FILHO, 2019, p. 237), partindo de uma premissa de presunção de culpa. Assim, o direito francês retratou bem a evolução paulatina dessa ideia de responsabilidade por fato de terceiro, iniciada com uma *presunção de*

[52] Lei n. 3.071/1916. Código civil brasileiro. Art. 1.523. Excetuadas as do art. 1.521, n. V, só serão responsáveis as pessoas enumeradas nesse e no artigo 1.522, provando-se que elas concorreram para o dano por culpa, ou negligência de sua parte (BRASIL, 1916).

[53] Responsabilidade vicária é o termo utilizado, principalmente nos países de *common law*, para designar 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, à exemplo da responsabilidade pelo fato de terceiro, assim como a responsabilidade dos pais pelos atos dos filhos menores que estiverem sob o seu poder e em sua companhia, o tutor e o curador pelos pupilos e curatelados, e o patrão pelos atos dos seus empregados (ALBIANI, 2018).

culpa, que, após ampliada em seu espectro, passou a ser tratada como *presunção de responsabilidade*, expressão essa que não se mostrou indene de críticas diante de sua imprecisão técnica, visto que não caberia presumir alguém responsável: ou é responsável ou não é responsável. Poder-se-ia presumir a culpa, e por isso lhe atribuir a responsabilidade. Apesar das controvérsias, a expressão foi bem acolhida de maneira geral, consolidando a responsabilidade ao guardião jurídico da coisa. Ou seja, o guardião a ser responsabilizado não necessariamente será o proprietário de direito, que seria o guardião presuntivo, mas aquele que no momento do evento detinha o poder de comando, custódia e direção do animal ou coisa. De igual maneira, o guardião presuntivo poderia ilidir sua responsabilidade se demonstrasse que um terceiro detinha esse poder de comando, a exemplo de um locatário, comodatário ou depositário no momento do evento danoso.

Desta feita, foram enunciados critérios para caracterizar a responsabilidade por fato das coisas, o primeiro deles, o *critério do proveito*, já demonstra a similitude que era tratada com a teoria do risco ao atribuir que seria considerado o guardião da coisa aquele que obtivesse proveito econômico com sua propriedade, utilização ou detenção. Na sequência, os irmãos Mazeaud (1955 *apud* PEREIRA, 2018, p. 139) apresentam a *direção material* como segundo critério a ser aferido para a configuração da responsabilidade por fato das coisas. Este critério estabelecia que guardião seria aquele que detivesse a direção da coisa na ocasião do evento danoso, ressalvada a hipótese em que o guardião de direito, tivesse confiado a coisa a um terceiro sem que perdesse seus direitos de uso, devendo ser analisado em caso a participação do guardião presuntivo e do terceiro para o evento, não atribuindo a responsabilidade automaticamente ao terceiro. O último critério foi consubstanciado como a *direção intelectual*, assim considerado o poder de dar ordens ou o poder de comando relativamente à coisa.

Mais uma vez, observada toda a evolução do ordenamento jurídico francês, o qual exerce forte influência no brasileiro, é possível justapor a responsabilidade civil por fato da coisa à hipótese de um ato produzido por inteligência artificial, alcançando-se uma perfeita analogia entre o cuidado de se identificar o responsável pela guarda da coisa com a cautela necessária que deve existir quando da identificação da exata contribuição da atuação do desenvolvedor e do usuário para o dano causado (MULHOLLAND, 2019).

A responsabilidade por ato de terceiro aplicada à hipótese proposta, rememora até a responsabilidade noxal, teoria de origem romana, em que o *pater familiae* se obrigava a reparar os danos causados pelas pessoas e pelos escravos que se encontrassem sob seu poder (MULHOLLAND, 2019). A ênfase destacada aos *escravos* mencionados como se não se inserissem no plexo do vocábulo *pessoas* é proposital, visto que naquela época eles eram

considerados *coisas* inteligentes e autoconscientes, que, apesar de fugirem do controle de seu mestre, eram meras coisas, à exemplo do que podemos considerar a inteligência artificial nos dias atuais. Essa, no entanto, diverge em um aspecto das teorias ordinariamente aplicadas às coisas, posto que essas teorias não as consideravam como autônomas e inteligentes.

Traz-se à lume, para ilustração, o caso do *chatbot* Tay, da Microsoft[54]. Em que pese não se tenha conhecimento de uma pessoa em específico que tenha sido ofendida, mas diante da série de *tweets* racistas, homofóbicos, pró-nazistas e anti-feministas, é completamente factível que determinadas minorias tenham se sentido injuriadas, havendo, portanto, um dano causado. Quem seria o responsável por esse dano? O desenvolvedor do sistema de inteligência artificial ou os usuários que alimentaram o sistema, induzindo seu aprendizado desagradável e potencialmente ilícito? Se mostra indene de dúvidas que a intenção projetada pela companhia não era a que foi alcançada. Contudo, a sua interação em um campo aberto, sem qualquer baliza efetiva, fez com que o sistema de inteligência artificial, sem firmes valores fundamentais intrínsecos, assumisse uma postura completamente indesejada. Caso se tratasse de um assistente pessoal como os já conhecidos no mercado[55], seria possível até se cogitar uma responsabilidade do próprio usuário *prossumidor*, em razão da customização natural esperada ao produto. Porém, antes de se cogitar a responsabilização do usuário, dever-se-ia ser perquirida a existência e consistência de uma regulação *by design* programada no sistema de inteligência artificial, a fim de se aferir as *travas* estabelecidas pelo desenvolvedor com vistas a evitar a evolução indevida da inteligência artificial. Assim deveria, salvo melhor juízo, ser solucionada a questão do *chatbot* Tay. Considerando ter sido o sistema desenvolvido para interagir na internet, em rede aberta sem qualquer controle, filtro ou supervisão, a inexistência de qualquer diretriz sensível a valores avoca para o desenvolvedor a responsabilidade pelos atos praticados pela máquina, sem prejuízo de que sejam responsabilizados igualmente os actantes que contribuíram para o atingimento do resultado, neste caso, os usuários que provocaram o seu desenvolvimento repulsivo, em uma espécie de responsabilidade compartilhada.

Isso porque, o fato de existir sobre o desenvolvedor do sistema de inteligência artificial um peso maior não pode isentar os demais actantes da rede sociotécnica e suas esferas de controle e influência que estimularam o resultado danoso obtido. Não se mostra razoável que lhes seja assegurada uma irresponsabilidade irrestrita (MAGRANI; SILVA; VIOLA, 2019),

[54] Em 2016 a Microsoft lançou um sistema de inteligência artificial, chamado de Tay, baseado em *machine learning*, com objetivo de interagir com usuários da rede social Twitter. Contudo, em menos de vinte e quatro horas, o chatbot passou a assimilar e reproduzir conteúdo racistas, transfóbicos, nazistas etc., em razão de estímulos mal-intencionados de usuários do microblog.

[55] Alexa, da Amazon; Siri, da Apple; Cortana, da Microsoft e Google Assistant, da Google.

devendo, ao contrário, ser buscada a criação de uma formatação de responsabilidade difusa, especialmente na esfera pública atualmente representada pela internet, posto que inexistente qualquer mecanismo semelhante no ordenamento jurídico vigente no país.

Para tanto, o escrutínio do nexo de causalidade se faria necessário, urgindo um aprofundamento na teoria da causa direta e imediata, adotada pelo direito brasileiro desde o código civil de 1916, em seu artigo 1.060[56], mantida no código civil de 2002, no seu artigo 403[57]. Não é demais ressaltar, ainda, que *coincidência* não implica em *causalidade*, de modo que é imprescindível que seja estabelecida uma efetiva relação de causa e efeito entre o fato danoso e o próprio dano (PEREIRA, 2018, p. 108). Segundo a teoria adotada no ordenamento pátrio, o dever de indenizar somente poderia ser atribuído a um agente que tivesse sua conduta como causa direta e imediata da ocorrência do dano. A ela foi adicionada uma *subteoria* em que é exigida a necessariedade a essa causa para que seja estabelecido seu nexo com o dano. Ou seja, a causa, para que tenha o nexo configurado, deverá ser direta, imediata e necessária à ocorrência do dano. Outras teorias relativas ao nexo causal também merecem destaque, muito embora não sejam aplicadas no âmbito do direito civil brasileiro. A primeira delas é a da *equivalência das causas*, adotada pelo código penal brasileiro. Criada pelo jurista alemão Von Buri, sustenta que todos os fatores pretéritos seriam considerados "causas", não havendo qualquer diferenciação entre si, desde que concorressem para o evento danoso (GAGLIANO; PAMPLONA FILHO, 2019, p. 146). A generalidade de sua acepção lhe trouxe um grave inconveniente, visto que conforme sua ideia central de que todas as causas pretéritas seriam equivalentes, haveria uma cadeia antecedente infinita de causas. A doutrina penalista, assim, a mitigou, afastando a responsabilidade de agentes quando sua participação na cadeia se apresentasse de forma indireta, sem qualquer previsibilidade da ocorrência do dano. Com tal depuração, o fabricante de uma arma de fogo, acertadamente, não responderia pelo cometimento de um assassinato por um indivíduo com o uso de um revólver por ele fabricado. Com tal ajuste, se aproximou à teoria adotada no sistema de responsabilidade civil brasileiro.

De seu turno, a *teoria da causalidade adequada*, adotada pelo direito argentino, é um refinamento da teoria da equivalência das causas, já que parte de suas conclusões para afastar a acepção de que é causa *toda e qualquer condição que tenha contribuído para o resultado*,

[56] Lei n. 3.071/1916. Código civil brasileiro. Art. 1.060. Ainda que a inexecução resulte de dolo do devedor, as perdas e danos só incluem os prejuízos efetivos e os lucros cessantes por efeito dela direto e imediato (BRASIL, 1916).

[57] Lei n. 10.406/2002. Código civil brasileiro. Art. 403. Ainda que a inexecução resulte de dolo do devedor, as perdas e danos só incluem os prejuízos efetivos e os lucros cessantes por efeito dela direto e imediato, sem prejuízo do disposto na lei processual (BRASIL, 2020e).

mantendo-se como tal apenas o antecedente abstratamente idôneo que se apresente não somente como necessário, mas adequado à ocorrência do dano (GAGLIANO; PAMPLONA FILHO, 2019, p. 148), se assemelhando, também, com a subteoria conjugada à teoria da causa direta e imediata, anteriormente citada.

No entanto, em não raros casos, os tribunais e até mesmo a doutrina têm, ao arrepio da legislação e de uma maneira completamente casuística, se distanciado das teorias acima mencionadas, e tomado posições que elastecem sobremaneira a interpretação do nexo de causalidade, dispensando "a prova da relação causal no tocante a um resultado ulterior da conduta do agente, assegurando ao nexo de causalidade uma elasticidade que nenhuma das teorias usuais comportaria" (SCHREIBER, 2009, p. 70), na busca de assegurar uma maior garantia de ressarcimento à vítima ou seus familiares, se aproximando do que se entende por *teoria do resultado mais grave*[58].

Ponderando acerca da incapacidade de evitar o dano e inaptidão do agente humano sequer prever a sua ocorrência, Ana Frazão (2018) encontra fundamentos na teoria do risco proveito, e da teoria do *deep pocket*, evitando-se assim um determinismo tecnológico que poderia buscar o agente humano em esquivar-se de qualquer responsabilidade ao argumento de que a tomada de decisão foi transferida para o sistema de inteligência artificial, fugindo de sua governabilidade. O lucro auferido é o ponto central para que sejam alcançados os agentes envolvidos na atividade que tenham melhores condições financeiras de suportar e administrar seus riscos (MULHOLLAND, 2019). Essas teorias se mostram compatíveis aos danos causados por ato praticado por sistemas de inteligência artificial, seja qual for a esfera em que se busque aferir a responsabilidade do causador do dano. Cabível, ainda, quando se tratar de hipótese em que houver uma participação do próprio consumidor e do ambiente na causação do dano, o que parece comum nas questões que envolvem inteligência artificial, sua aplicação conjunta com a doutrina voltada para o estudo da pluralidade de concausas (TEPEDINO; SILVA, 2019a), tal como a *causalidade múltipla*, que aponta para a responsabilidade pelo dano efetivamente causado por cada agente, e não por um fato superveniente ou por um agravamento não imputável ao evento (PEREIRA, 2018, p. 113).

Esclareça-se lateralmente, que, aos olhos da vítima do dano – ou dos terceiros atingidos em razão do evento danoso –, como já dito, há uma solidariedade entre os coagentes, personagens centrais de toda a sistemática de responsabilidade civil (PEREIRA, 2018, p. 114), quando restar configurado um nexo causal plúrimo, de modo que todos responderão

[58] Texto original: "*The Thin Skull Rule*" ou "*The Egg-Shell Skull Rule*".

solidariamente pela integralidade do dano perante os atingidos, sendo assegurado ao actante escolhido para ressarcir o dano a *actio de in rem verso* a fim de estabelecer a quota proporcional de cada um dos coobrigados.

Em linhas pretéritas, urge mencionar que outras teorias ou subteorias também se mostram compatíveis com a aplicação da responsabilidade civil em razão de ato praticado por inteligência artificial, com sua aplicação em momentos diversos ou com expressividade menor, à exemplo, respectivamente, da teoria do rompimento do nexo causal e da ambivalente teoria do risco do desenvolvimento.

A primeira delas, não se trata de uma teoria autônoma em si, mas de um evento que interrompe o nexo de causalidade para a caracterização do dever de indenizar. Essa interrupção pode ocorrer por diversos fenômenos, dentre os quais cita-se: caso fortuito, força maior, culpa concorrente e culpa exclusiva da vítima. Em todas essas hipóteses, há um rompimento do nexo de causalidade, de modo que o fato imputado à conduta comissiva ou omissiva do agente não pode ser considerado como determinante para o acontecimento do evento danoso. Ressalta-se novamente, com preocupação diante da insegurança jurídica que proporciona, uma tendência a flexibilizar o rigor na verificação do nexo de causalidade, a fim de assegurar o ressarcimento integral à vítima, que se verifica com a aplicação de teorias como a *the thin skull rule* acima mencionada; a imposição do dever de indenizar no que restou convencionado como *fortuito interno*, quando, a despeito de sua imprevisibilidade, o evento se relaciona aos objetivos e riscos inerentes à atividade desenvolvida (ALBIANI, 2018); a *presunção de causalidade* que tem superado obstáculos probatórios se valendo de juízos de probabilidade; e a própria *causalidade alternativa* na qual, não sendo possível a escorreita identificação do agente, é atribuído a todos eles o dever de indenizar.

Por fim, a chamada teoria do *risco do desenvolvimento*, que tem por elementos: (i) o dano resultante de um sistema que não apresenta falhas, em tese; (ii) a impossibilidade de identificação da potencialidade da ocorrência do dano no momento de sua ocorrência; (iii) um desenvolvimento tecnológico posterior que identifica maneiras de evitar a ocorrência do referido dano, corrigindo a "falha"; preleciona que poderia ser afastada a responsabilidade do agente, em particular o desenvolvedor, caso seja verificado que foi utilizada a *tecnologia mais segura conhecida pela comunidade cientifica à época da sua elaboração*. Nesse sentido, justificar-se-ia a exclusão da responsabilidade do desenvolvedor por eventuais danos caso não lhe fosse possível evitar em razão do estado da arte (TEPEDINO; SILVA, 2019a).

Curiosamente, a mesma teoria do risco do desenvolvimento aplicada aos sistemas de inteligência artificial, também pode ser interpretada de forma a atribuir responsabilidade ao

desenvolvedor para indenizar a vítima, considerando que o dano é resultado de uma atuação da inteligência artificial, a ausência de previsibilidade da potencialidade danosa do sistema, cumulada com a independência do sistema em relação ao homem e a impossibilidade de uma explicação *ex post* da decisão tomada pela inteligência artificial, principalmente com base no princípio da solidariedade social e na tendência retromencionada de se assegurar uma reparação integral à vítima.

3.4 PROPOSIÇÕES PARA EQUACIONAR A REGULAÇÃO E INOVAÇÃO TECNOLÓGICA NO CAMPO DA RESPONSABILIDADE CIVIL

Muitas são as alternativas aventadas para fazer com que o ordenamento jurídico consiga acompanhar o mundo dos fatos no que se refere à inteligência artificial, interessando particularmente ao presente estudo àquelas que possibilitariam uma responsabilização civil em razão de atos praticados de forma automatizada, por sistemas inteligentes.

A autorregulação já tem sido realizada de forma voluntária pelos actantes, por meio de uma *softlaw* que tem traçadas diretrizes basilares para o alcance de uma inteligência artificial explicável, confiável e segura, desde a sua concepção. Merece destaque, a esse respeito, a iniciativa promovida pelo Conselho Nacional de Justiça (CNJ), por meio da sua resolução n. 332, de 21 de agosto de 2020 , que dispôs sobre a ética, transparência e a governança na produção da inteligência artificial aplicada ao Poder Judiciário brasileiro, que trouxe de uma forma robusta, valores a serem observados pelas aplicações dirigidas à gestão e aos processos judiciais (BRASIL, 2020d).

De outro lado, é cogitada, ainda, a criação de novas matrizes que facilitem e reduzam a complexidade da aplicação da responsabilidade civil aos casos que envolvam sistemas de inteligência artificial, aos quais se referenciam a atribuição de personalidade eletrônica, estabelecimento de seguro obrigatório e implementação de taxas para o seu uso, bem como a regulação do uso e desenvolvimento da inteligência artificial por meio de lei.

3.4.1 Personalidade eletrônica ou *e-personalidade*

A criação de uma personalidade eletrônica é das mais inovadoras soluções para a inserção de responsabilidade civil dos sistemas de inteligência artificial no ordenamento jurídico. A atribuição de direitos e deveres a uma máquina é um caminho que desperta discussões relativas à existência de intencionalidade, senciência e sapiência em robôs. No

entanto, a possibilidade de se conferir personalidade jurídica não necessariamente precisa ser discutida em tais termos, visto que já é comum essa atribuição de direitos e deveres a diversos entes despersonalizados, de forma fictícia, tal como ocorre com a própria pessoa jurídica, com a massa falida, espólio, fundações ou com o condomínio edilício, que, a despeito de não possuírem propriamente uma personalidade jurídica, recebem do ordenamento uma *subjetividade* que lhes é bastante para o atingimento de um diálogo social, bem como para figurarem em relações jurídicas, sem que seja necessariamente cogitada sua equiparação a uma pessoa natural (FARIAS; ROSENVALD; BRAGA NETO, 2015, p. 313).

Analisando tal hipótese, sem dúvidas seria possível a atribuição de responsabilidade ao sistema de inteligência artificial caso reconhecida sua personalidade jurídica, visto que restaria constituída uma relação entre dois sujeitos, sendo inegável que exprime um centro de interesses relevante ao mundo jurídico, razão pela qual se apresenta adequada sua percepção como titular de direitos (MAGRANI; SILVA; VIOLA, 2019).

Contudo, esta atribuição de personalidade jurídica aos sistemas de inteligência artificial resta controvertida em razão do parco desenvolvimento para sua autonomia em relação aos seres humanos. Além disso, mesmo que o eixo gravitacional da responsabilização civil no ordenamento jurídico brasileiro seja em torno da reparação do dano, e não na punição do seu causador, é inconteste de que não restaria satisfeita a sensação de justiça da vítima com a "punição" equivalente ao seu desligamento definitivo dada a um artefato eletrônico, em razão de um dano por si experimentado, por exemplo.

Demais disso, a possibilidade chega a ser vista até por inconstitucional, diante da cláusula geral assegurada pela Constituição Federal de tutela e promoção da pessoa humana como valor máximo do ordenamento jurídico, insculpida no princípio da dignidade da pessoa humana. De outro lado, a comparação da outorga de personalidade aos sistemas eletrônicos e àquela realizada em favor das pessoas jurídicas também não é isenta de fortes críticas, posto que às pessoas jurídicas é outorgado um tratamento especial para que as pessoas naturais alcancem determinados benefícios que individualmente não conseguiriam, assim como para que o Estado possa exercer um controle maior sobre o exercício de atividades econômicas. É um instrumento de ordem técnica, desenvolvido a serviço da pessoa natural e da sociedade (MAGRANI; SILVA; VIOLA, 2019).

O parlamento europeu caminha para a outorga de personalidade eletrônica em um futuro próximo, pretendendo que, "pelo menos, os robôs autônomos mais sofisticados possam ser determinados como detentores do estatuto de pessoas eletrônicas responsáveis por sanar quaisquer danos que possam causar" e, eventualmente, "aplicar a personalidade eletrônica a

casos em que os robôs tomam decisões autônomas ou em que interagem por qualquer outro modo com terceiros de forma independente" (NADKARNI, 2017). Na resolução enviada à Comissão de Direito Civil sobre Robótica 2105/2103[INL] (XAVIER; SPALER, 2019), foi recomendada a adoção de um registro obrigatório dos robôs e a criação de um seguro, para responder financeiramente pelos danos causados.

Verifica-se, pois, que essa outorga de personalidade tem um viés estritamente patrimonial, de modo que não restou demonstrada a realização de uma análise mais aprofundada acerca dos desdobramentos jurídicos desta solução, consoante explicita Carlos Affonso Souza (2017, p. 4),

> No cenário europeu, impulsionado por indagações sobre responsabilidade, a questão da personalidade aparece muito mais ligada à construção de um mecanismo de reparação à vítima de danos do que como resultado de uma discussão mais aprofundada sobre o que é um robô inteligente e seu estatuto jurídico de forma mais abrangente.

Merece menção, por fim, o audacioso projeto que tem sido desenvolvido pela Arábia Saudita, de criar uma cidade digital, onde todas as atividades serão realizadas por robôs dotados de sistemas de inteligência artificial, e, para tanto, já saíram na dianteira de todo o mundo, conferindo cidadania a uma robô humanoide, que tem habilidades de expressar (ou simular) emoções e se comunicar com seres humanos. "Batizada" de Sophia, o robô foi criado pela empresa Hanson Robotics, para ajudar idosos e auxiliar visitantes em parques e eventos (AGRELA, 2017).

O reconhecimento da personalidade eletrônica já é uma proposição que virou realidade, mas, parece indene de dúvidas que, por maior que seja a autonomia do sistema de inteligência artificial, ainda não é possível que seja comparado a uma pessoa natural, pelo menos no atual estado da arte. Tal solução, entrementes, pouco ou em nada auxilia em seu propósito principal de facilitar a mitigação dos riscos e compensação de danos às possíveis vítimas. Exatamente por isso, desponta inócua a simples atribuição de personalidade senão acompanhada de mecanismos que assegurem um lastro patrimonial, visto que, o objetivo principal da outorga de personalidade seria a reparação de eventual dano que fora causado em razão do sistema autômato. Inexistindo na *pessoa eletrônica* a senciência própria dos seres vivos, incogitável prospectar que qualquer tipo de punição em razão do cometimento de alguma conduta ilícita seja minimamente coerente. E, diante do foco do sistema de responsabilidade civil na reparação da vítima, se apresenta como um ponto mais relevante a capacidade de reparação do dano do

que propriamente a capacidade civil de um sistema de inteligência artificial para figurar ou não isoladamente em uma relação jurídica.

Contudo, caso a opção seja pelo desenvolvimento de uma e-personalidade, várias deverão ser as etapas de inserção de tal previsão no ordenamento jurídico. Desde estabelecer uma autoridade certificadora do grau de autonomia da inteligência artificial que possa reconhecer a sua personalidade jurídica própria, diversa e independente dos seus responsáveis, bem como a adoção de mecanismos de prevenção de riscos e de segurança. Penalidades em razão da prática de condutas ilícitas, com caráter pedagógico e punitivo, que poderiam consistir em multas, indenizações, suspensão temporária e até definitiva do sistema.

Diante de todo o expendido, restou demonstrado que o estabelecimento de uma vinculação jurídica entre o sistema de inteligência artificial e as pessoas físicas ou jurídicas responsáveis se mostra bastante para assegurar uma justa punição e integral reparação do dano causado, não sendo demais lembrar a previsão estatuída no artigo 12[59] da Convenção das Nações Unidas sobre o uso de comunicações eletrônicas em contrato

> que determina que uma pessoa em cujo nome um computador foi programado deve ser responsável por qualquer mensagem gerada pela máquina. Assim, a negociação estabelecida pelo sistema de inteligência artificial é considerada perfeita, e válida sua manifestação de vontade, bem com as obrigações daí advindas, sem, contudo, haver o reconhecimento da sua personalidade jurídica, atribuindo a responsabilidade pelos seus atos à pessoa em cujo nome agiu (ALBIANI, 2018, p. 16).

Isso importa no reconhecimento da inteligência artificial como ferramenta (*AI-as-tool*), com o estabelecimento de uma responsabilidade objetiva ligada ao nome de quem ela age ou que a supervisiona, independentemente de qualquer planejamento ou previsibilidade. A proposição europeia traz, ainda, menção expressa à essa correta mensuração da responsabilidade pela causação do dano. Quando for em razão do treinamento, será responsável quem o ensinou. Se a conduta for decorrente de um defeito já existente, do fabricante ou criador.

3.4.2 Seguro obrigatório, constituição de patrimônio de afetação, agência certificadora e taxação do uso

A outorga de personalidade jurídica por si só, não resolve o problema principal da responsabilidade civil, tanto que todas as soluções propostas se utilizam de instrumentos que

[59] Artigo 12. *Uso de sistemas automatizados de mensagens na formação de contratos*
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 inexequível 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.

conjugam a elas uma repercussão patrimonial. A criação de um seguro obrigatório, à exemplo do seguro existente para acidentes de trânsito, conhecido pelo acrônimo DPVAT que significa *danos pessoais causados por veículos automotores de via terrestre*[60], poderia ser uma hipótese viável para que restasse assegurada reparação de danos eventualmente causados por sistemas de inteligência artificial, visto que socializaria o risco, a exemplo do que é feito com seguros de qualquer gênero. Todos os responsáveis e usuários de sistemas de inteligência artificial pagariam um pequeno valor a esse título, o qual, coletivamente, seria capaz de responder pelos danos que viessem a ser causados por um ou outro.

De igual maneira, a obrigatoriedade de constituição de um patrimônio de afetação atenderia ao mesmo propósito de existir um fundo dotado de capacidade financeira para responder pelos prejuízos decorrentes dos danos causados por sistemas de inteligência artificial, o qual poderia funcionar nos mesmos moldes, por exemplo, do fundo de defesa de direitos difusos, previsto na lei n. 7.347/1985.

Há, ainda, uma proposição de criação de uma agência responsável pela certificação de sistemas de inteligência artificial. A referida certificação seria baseada em critérios técnicos para que restassem avaliados os riscos relativos a um determinado sistema. Não seria obrigatória esta certificação, a fim de não obstacularizar o desenvolvimento das tecnologias, mas, a não certificação do sistema implicaria na responsabilização solidária dos projetistas, fabricantes e vendedores, enquanto os sistemas certificados teriam um alcance limitado da responsabilização civil (ALMADA, 2019). Esta possibilidade, mesmo após acatada, ainda dependeria de uma série de regulações nacionais e internacionais, diante da ampla globalização do mercado, o que pode vir a ser um obstáculo para sua efetiva implementação.

A possibilidade de tributação de alguma forma aos desenvolvedores e usuários, apesar de também objetivar a criação desse lastro patrimonial, não tem sido bem recebida nos ambientes públicos em que foi apresentada, sendo rejeitada pelo Parlamento Europeu a previsão da criação de um imposto sobre o trabalho realizado pelos robôs, ou uma taxa em razão da sua utilização (NADKARNI, 2017), apesar de se apresentar como uma hipótese factível, visto que, a despeito de ser possível argumentar existir um gasto inicial elevado, que seria majorado com a taxação, não se pode negar que o uso da inteligência artificial a médio e longo prazo importa em uma otimização de recursos e elevação significativa de acurácia, resultando em um retorno financeiro substancial, que seria suficiente para custear o tributo criado.

---

[60] A Lei n. 6.194/1974 assegura uma indenização pré-estabelecida a todas as pessoas que sejam vitimadas por acidentes automobilísticos, independentemente de culpa do motorista, em razão do reconhecimento dos riscos inerentes ao trânsito.

O custeio de tal seguro ou fundo de capital deve se dar, em um cenário ideal, pelas pessoas que desenvolvem, exploram ou se utilizam dos sistemas de inteligência artificial, visto que são elas quem se aproveitam economicamente ou dos benefícios proporcionados por eles. Desta forma, a adoção de mecanismos que asseguram uma irrestrita reparabilidade de danos causados por sistemas de inteligência artificial, além de garantir uma segurança jurídica para os usuários e consumidores, também o traria aos desenvolvedores, que não teriam contra si todos os riscos pelas intempéries da imprevisibilidade e vulnerabilidades dos sistemas inteligentes, se apresentando, pois, como uma medida incentivadora do desenvolvimento (MULHOLLAND, 2019).

3.4.3 Projetos de lei em trâmite no Congresso Nacional brasileiro

O Congresso Nacional brasileiro possui diversos projetos de lei em trâmite que tratam do uso da inteligência artificial de alguma forma. Alguns deles possuem foco específico em questões satélites, dentre os quais cita-se o projeto de lei n. 679/2020, do Deputado Eduardo Bismarck, que pretende a inserção de linguagem de programação nos três anos do ensino médio; e o projeto de lei n. 2.576/2020, do Deputado Amaro Neto, que pretende aplicar a inteligência artificial ao Cadastro Nacional de Pessoas Desaparecidas, para que seja projetado, por meio desses sistemas, como estaria o rosto da pessoa desaparecida na data em que for buscada. De outro lado, iniciam-se os debates centrais acerca do desenvolvimento e uso da inteligência artificial, com proposições para uma regulação nacional, aos quais será dirigida a atenção do presente ponto. Destacam-se, especificamente, os projetos de lei em trâmite na Câmara dos Deputados sob o n. 21/2020 (BRASIL, 2020a) apresentado em 04/02/2020, pelo Deputado Federal Eduardo Bismarck, n. 240/2020 (BRASIL, 2020b), apresentado em 11/02/2020, pelo Deputado Federal Léo Moraes, e n. 4.120/2020 (BRASIL, 2020c), apresentado em 07/08/2020, pelo Deputado Federal Bosco Costa, assim como os projetos de lei em trâmite no Senado Federal sob o n. 5.051/2019 (BRASIL, 2019c), apresentado em 16/09/2019, pelo Senador Federal Styvenson Valentim, e n. 5.691/2019 (BRASIL, 2019d), apresentado em 24/10/2019, também pelo Senador Federal Styvenson Valentim.

Referidos projetos de lei, cada um à sua maneira e com pontos positivos e deficiências, buscam criar uma Política Nacional de Inteligência Artificial (PNIA), estabelecer princípios, objetivos, diretrizes, recomendações, direitos e deveres relacionados ao desenvolvimento e uso da inteligência artificial, bem como trazer mecanismos e instrumentos para compatibilizar a inovação tecnológica com a necessária segurança jurídica.

Passa-se a analisar cada um dos projetos mencionados, por ordem cronológica de apresentação, portanto, os projetos do Senador Federal Styvenson Valentim, de n. 5.051/2019 e 5.691/2019. O primeiro deles tem por objetivo principal o estabelecimento de princípios para o uso da inteligência artificial, como já sinaliza em sua ementa. Traz, já em seu artigo 2º a condição de subserviência que coloca a inteligência artificial aos seres humanos e ao seu bem-estar. Sinaliza, igualmente, que a inteligência artificial deverá, sempre, ser auxiliar à tomada de decisão humana, retirando completamente a possibilidade de que tomem decisões de forma completamente autônoma. Estabelece como fundamento o respeito à dignidade da pessoa humana, democracia, igualdade, direitos humanos, pluralidade e diversidade, bem como a observância da garantia de proteção da privacidade e dos dados pessoais, transparência, confiabilidade e auditabilidade, sempre condicionados à supervisão humana. O projeto de lei n. 5.051/2020 além de estabelecer a responsabilidade civil pelos danos ao seu supervisor, também define diretrizes de Estado para o desenvolvimento da inteligência artificial, primando sempre pelo valor do ser humano, tanto no que se refere à uma educação mental, emocional e econômica, assim como proteção e qualificação aos trabalhadores quando da implementação gradual da inteligência artificial, em busca de uma prestação de serviços eficientes e de qualidade à população.

O segundo projeto apresentado pelo Deputado Federal Styvenson se propõe a instituir uma Política Nacional de Inteligência Artificial (PNIA), desta forma, possui um caráter mais principiológico, se atendo ao estabelecimento aos mesmos princípios do projeto anterior acrescidos do respeito à ética e desenvolvimento inclusivo e sustentável. Nas diretrizes, lança luzes alinhadas com os princípios, prescrevendo adicionalmente um estímulo ao desenvolvimento e pesquisa da inteligência artificial, cooperação entre entes públicos e privados e entre empresas, nacionais e internacionais, capacitação e o desenvolvimento de mecanismos de fomento à inovação e incentivos fiscais para pesquisa em tecnologia. Aborda, em seu artigo 4º, deveres compatíveis com os princípios e diretrizes a serem observados pelas soluções de inteligência artificial. Por fim, traz como instrumentos da política nacional os "programas transversais elaborados em parceria com órgãos públicos e instituições privadas; os fundos setoriais de ciência, tecnologia e inovação e os convênios para desenvolvimento de tecnologias sociais", os quais deverão ser apoiados e fortalecidos por meio da celebração de convênios entre os entes públicos e entidades públicas ou privadas para obtenção de recursos técnicos, humanos e financeiros.

Já na outra casa legislativa, o projeto de lei n. 21/2020, de autoria do Deputado Federal Eduardo Bismarck, propõe o estabelecimento de princípios, diretrizes, direitos, deveres e

instrumentos de governança para o uso da inteligência artificial no país, com destaque à valorização do ser humano em todo o processo de incentivo ao desenvolvimento ético, democrático, igualitário, não discriminatório, plural e seguro da tecnologia. Prima pelo respeito à livre iniciativa, à livre concorrência, aos direitos trabalhistas, privacidade e proteção de dados. Traz direitos para as partes interessadas relativos à transparência, como informação (*ex ante*) e explicação (*ex post*), bem como deveres aos agentes de inteligência artificial, inclusive sua responsabilidade civil de acordo com a sua função. Por fim, aborda as diretrizes a serem observadas pelos entes públicos, com especial atenção à transição para a sua adoção nos serviços públicos e privados, e determinação para o estabelecimento de instrumentos de governança, tais como o relatório de impacto de inteligência artificial. Não há na proposta legislativa a inserção de obrigatoriedade de supervisão humana, inobstante exista previsão clara do dever de encerrar o sistema caso o controle humano não seja mais possível.

O projeto de lei n. 240/2020 foi apensado ao projeto n. 21/2020, visto que compartilham do mesmo objeto, apesar de este praticamente repetir parcialmente o anterior, deixando, outrossim, importantes lacunas acerca de questões importantes como explicabilidade, supervisão humana, transição para inteligência artificial, além de não estabelecer qualquer instrumento de governança.

O último dos projetos analisados, de n. 4.120/2020, não se refere propriamente à inteligência artificial, mas, ao uso de algoritmos pelas plataformas digitais na internet, razão pela qual guarda familiaridade com o conteúdo ora tratado, pelo menos no seu espectro de abrangência. De início define os princípios a serem observados pelos sistemas de decisão automatizada na mesma linha dos projetos anteriores, trazendo como pontos inéditos a obrigatoriedade da elaboração de relatório de impacto de seus sistemas de decisão automatizadas, uma gradação de risco para os sistemas consoante seu grau de autonomia, acurácia e dados utilizados. Ainda, de forma inovadora, prevê a aplicação de sanções em razão da inobservância de seus termos, além de garantir, de certa forma, um direito à explicabilidade.

Tabela 01 - Comparativo dos projetos de leis em trâmite que regulam a inteligência artificial

Número do projeto de lei	**PL n. 5.051/2019**	**PL n. 5.691/2019**	**PL n. 21/2020**	**PL n. 240/2020**	**PL n. 4.120/2020**
Casa legislativa	Senado Federal	Senado Federal	Câmara dos Deputados	Câmara dos Deputados	Câmara dos Deputados
Autoria	Styvenson Valentim	Styvenson Valentim	Eduardo Bismarck	Léo Moraes	Bosco Costa

Continuação da Tabela 01

	PL n. 5.051/2019	PL n. 5.691/2019	PL n. 21/2020	PL n. 240/2020	PL n. 4.120/2020
Ementa resumida	Estabelece os princípios para o uso da IA no Brasil.	Institui a Política Nacional de IA.	Estabelece princípios, direitos e deveres para o uso de IA no Brasil.	Cria a Lei da IA, e dá outras providências.	Disciplina o uso de algoritmos pelas plataformas digitais na internet.
Propósito de servir à humanidade e à sociedade	Sim	-	Sim	Sim	-
Obrigatoriedade de supervisão humana	Sim	Sim	Sim	Sim	-
Condição de auxiliariedade à tomada de decisão	Sim	-	-	Sim	Não
Capacitação do trabalhador e sua preparação para o mercado de trabalho com IA	Sim	Sim	Sim	-	-
Previsão de responsabilidade civil	Sim	-	Sim	Sim	-
Fundamentos	Respeito à dignidade da pessoa humana; Democracia; Igualdade; Direitos humanos; Pluralidade; Diversidade; Proteção da privacidade e dos dados pessoais; Transparência; Confiabilidade e Auditabilidade;		Desenvolvimento tecnológico e a inovação; A livre iniciativa e a livre concorrência; Respeito aos direitos humanos e aos valores democráticos; A igualdade, a não discriminação, a pluralidade e o respeito aos direitos trabalhistas; A privacidade e a proteção de dados.	-	-
Penalidades	-	-	-	-	Advertência; multa; Suspensão temporária das atividades; Proibição de exercício das atividades.

Continuação da Tabela 01

	PL n. 5.051/2019	PL n. 5.691/2019	PL n. 21/2020	PL n. 240/2020	PL n. 4.120/2020
Instrumentos	-	Programas transversais elaborados em parceria com órgãos públicos e instituições privadas; Fundos setoriais de ciência, tecnologia e inovação; Convênios para desenvolvimento de tecnologias sociais.	Relatórios de impacto de IA e recomendar a adoção de padrões e de boas práticas para implantação e operação dos sistemas.	Convênios para obtenção de recursos técnicos, humanos ou financeiros.	Relatório de impacto de sistema de decisão automatizada.
Objetivo	Promoção e a harmonização da valorização do trabalho humano e do desenvolvimento econômico.	-	Promoção da pesquisa e do desenvolvimento da IA ética e livre de preconceitos; Promoção da competitividade e do aumento da produtividade, Melhoria na prestação dos serviços; Promoção do crescimento inclusivo, do bem-estar da sociedade e da redução das desigualdades; Promoção da cooperação internacional; Interoperabilidade.	-	-
Direitos das partes interessadas	-	-	Ciência da instituição responsável pelo sistema de IA; Acesso a informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados pelo sistema de IA que lhes afetem adversamente, observados os segredos comercial e industrial; e acesso a informações claras e completas sobre o uso, pelos sistemas, de seus dados sensíveis.	-	Acesso a informações sobre as metodologias empregadas pelo sistema que possam induzir seu comportamento ou afetar suas preferências.

Continuação da Tabela 01

	PL n. 5.051/2019	PL n. 5.691/2019	PL n. 21/2020	PL n. 240/2020	PL n. 4.120/2020
Princípios	-	Desenvolvimento inclusivo e sustentável; Respeito à ética, aos direitos humanos, aos valores democráticos e à diversidade; Proteção da privacidade e dos dados pessoais; Transparência; Segurança; Confiabilidade.	Uso da IA para buscar resultados benéficos para as pessoas e o planeta; Respeito à dignidade humana, à privacidade e à proteção de dados pessoais e aos direitos trabalhistas; Impossibilidade de uso dos sistemas para fins discriminatórios, ilícitos ou abusivos; Garantia de transparência sobre o uso e funcionamento; Garantir a funcionalidade e o gerenciamento de riscos dos sistemas de IA e a garantir a rastreabilidade dos processos e decisões.	Transparência; Segurança; Confiabilidade; Proteção da privacidade, dos dados pessoais e do direito autoral; Respeito à ética, aos direitos humanos e aos valores democráticos.	Boa-fé; Transparência; Responsabilidade social; Segurança; Proteção aos valores éticos e morais; Direito à privacidade e à intimidade dos cidadãos; Respeito aos direitos humanos e à democracia.
Diretrizes	De Estado: Valor do ser humano; Educação mental, emocional e econômica; Proteção e qualificação aos trabalhadores; Implementação gradual da IA; Proatividade na regulação das aplicações da IA; Prestação de serviços eficientes e de qualidade à população.	Da Política Nacional de Inteligência Artificial: Estabelecimento de padrões éticos; Promoção de crescimento inclusivo e sustentável; Melhoria da qualidade e da eficiência dos serviços; Investimentos em pesquisa e desenvolvimento; Promoção da cooperação, interação, intercâmbio, pesquisa e inovação das instituições de Ciência, Tecnologia e de Inovação; Fomento à inovação e ao empreendedorismo digital; Capacitação de profissionais da área.	De Estado: Promover e incentivar investimentos em pesquisa e desenvolvimento de IA; Promoção de um ambiente favorável para a implantação da IA, com a revisão e a adaptação das estruturas políticas e legislativas necessárias; Promoção da interoperabilidade tecnológica; Adoção de tecnologias, padrões e formatos abertos e livres; Estabelecimento de mecanismos de governança multiparticipativa, transparente, colaborativa e democrática, com a participação dos actantes.	Observar os limites sociais e a proteção ao patrimônio público e privado; Estabelecer os padrões éticos e morais; Melhoria da qualidade e da eficiência dos serviços; Promover o desenvolvimento sustentável e inclusivo na área de inovação e tecnologia; Investimento em pesquisa e desenvolvimento; Incentivar e estabelecer cooperação internacional em pesquisa e desenvolvimento da IA; Fomento à inovação e ao empreendedorismo digital; Capacitação de profissionais da área.	-

Continuação da Tabela 01

	PL n. 5.051/2019	PL n. 5.691/2019	PL n. 21/2020	PL n. 240/2020	PL n. 4.120/2020
Obrigações	-	Preservar autonomia, intimidade e privacidade das pessoas; Preservar os vínculos de solidariedade entre os povos e as diferentes gerações; Ser inteligíveis, justificáveis e acessíveis, abertas ao escrutínio democrático e controle por parte da população; Permitir a manutenção da diversidade social e cultural e não restringir escolhas pessoais; Prover decisões rastreáveis e sem viés discriminatório ou preconceituoso; Seguir padrões de governança que garantam o gerenciamento e a mitigação dos riscos potenciais da tecnologia.	Ao Estado, em conjunto com os actantes: capacitação, e outras práticas educacionais, para o uso confiável e responsável da IA; Formular e fomentar estudos e planos para promover a capacitação humana e para o desenvolvimento ético e responsável da IA; Divulgar o responsável pelo estabelecimento do sistema de IA; Fornecer informações claras e adequadas a respeito dos critérios e dos procedimentos utilizados pela IA; Assegurar que os dados utilizados pelo sistema de IA observem a LGPD; Implantar IA somente após avaliação adequada de seus objetivos, benefícios e riscos relacionados a cada fase do sistema; Encerrar o sistema se o seu controle humano não for mais possível; Proteger continuamente a IA contra ameaças de segurança cibernética; A adoção de sistemas de IA na Administração Pública e na prestação de serviços públicos, visando à eficiência e à redução dos custos.	Proibição de ferir seres humanos e nem serem utilizadas em destruição em massa, ou como armas de guerra ou defesa; A IA deve cumprir protocolos de Direitos Internacionais, de proteção à vida e aos Direitos Humanos; Todas as pesquisas e projetos devem ser submetidos aos pressupostos legais, aos órgãos de fiscalização e controle da área de ciência, pesquisa, inovação e tecnologia; A IA deve se submeter a período probatório na academia científica antes de obter o registro de operação.	Produzir relatório de impacto de seus sistemas; Publicar na internet extrato do relatório; Informar que utiliza de sistema de decisão automatizada; Elaborar e publicar na internet guia de orientação; Prestar informações justas, claras, transparentes e destacadas sobre as condições do serviço ofertado; Prestar as informações solicitadas pelo usuário; Prestar informações de padrões e boas práticas para o desenvolvimento e a operação de sistemas de decisão automatizada.

Fonte: Senado Federal e Câmara dos Deputados.

Diante da análise dos principais projetos de lei em trâmite no Congresso Nacional que versam sobre a inteligência artificial, pôde se verificar que as proposições não se apresentam maduras em nenhum deles para sua efetiva implementação. São necessárias muitas discussões para que seja construída uma lei que possa efetivamente trazer uma segurança jurídica aos envolvidos, sem que obste o desenvolvimento das tecnologias. A base principiológica se apresenta de forma consensual, bem como a definição de um propósito centrado no ser humano, respeitando os valores sociais. No entanto, há um visível desalinhamento no que se refere ao estabelecimento de uma supervisão humana, direito à explicabilidade, e à profundidade de sua extensão, definição de responsabilidades, transição para inteligência artificial e instrumentos de governança. São pontos que se alternam entre si, aparecendo em alguns dos projetos, mas ausentes em outros, sem que seja possível compreender se se pretendeu tal ausência, ou se tal

questão sensível não foi sequer considerada quando da elaboração da minuta do projeto de lei. Sem dúvidas, as questões relativas ao desenvolvimento e uso de inteligência artificial são novas, e ainda precisam ser jurídica, tecnológica e socialmente amadurecidas, demandando, inevitavelmente, ajustes e adaptações em momento posterior, até mesmo em razão do desenvolvimento avassalador da tecnologia, mas, sem, no entanto, pender sobre si o peso, ou sequer a expectativa, de que seja criada uma unanimidade, por mais abrangente que seja o diálogo entre a indústria, governo e academia.

CONCLUSÃO

Fatores como o desenvolvimento da capacidade computacional, a internet das coisas, o *big data* e a evolução algorítmica, têm se retroalimentado entre si para alçar a inteligência artificial a um patamar de tangibilidade perante a sociedade de uma maneira tão rápida, que o universo jurídico foi surpreendido com questões relativas a sistemas de inteligência artificial, enquanto ainda aprendia a lidar com o mundo digital e suas particularidades.

Várias são as aplicações já inseridas no cotidiano, tais como: dispositivos para reconhecimento ótico de caracteres e voz, identificação facial, realização de diagnósticos médicos, condução de veículos e aeronaves, tomada de decisão em geral, análise preditiva etc. Chamam especial atenção os sistemas de inteligência artificial que se baseiam em *machine learning* e *deep learning*, os quais têm por premissa principal desenvolver de forma autônoma o algoritmo que irá realizar o tratamento dos dados fornecidos (*inputs*) a fim de entregar os resultados pretendidos (*outputs*), visto que desse processo resulta uma das maiores dificuldades identificadas relativa à opacidade do algoritmo que, quanto mais assertivo, menos inteligível se apresenta. Pelos vários experimentos realizados, longe de se pretender considerar a inteligência artificial um ovo de colombo, é certo que nada mais é realizado pela máquina do que a mera instanciação de informações ou predição de resultados com base em dados que foram previamente inseridos ou capturados. Não passa o sistema de inteligência artificial, portanto, de um sábio idiota. Não há nele, axiologicamente falando, nenhuma inteligência com alguma forma de intencionalidade propriamente dita, que seja capaz de realizar alguma relação de causa e efeito. Desta forma, não se mostra racional relegar exclusivamente a um sistema de inteligência artificial o poder de tomada de decisão sem que haja a possibilidade de supervisão ou revisão humana, pelo menos não até que seja criada uma inteligência artificial geral ou a superinteligência.

A manutenção da supervisão humana no processo de tomada de decisão de sistemas de inteligência artificial é, inclusive, um ponto que tem se mostrado controvertido. Contudo, como não é possível se ter uma confiança plena na *ratio decidendi* dos algoritmos aprendizes, se mostra inadmissível conceber um cenário conforme o que se apresenta hoje no Brasil, em que é permitido um tratamento de dados realizado exclusivamente por sistemas de inteligência artificial, sem que seja assegurada a inserção obrigatória de uma supervisão ou revisão humana em alguma fase do processo.

No entanto, outro tem sido o caminho trilhado pelos actantes da esfera privada ao criarem suas autorregulações. Estão buscando estabelecer salvaguardas até mais benéficas do que aquelas previstas pelos governos, com a inserção de uma regulação *by design* sensível a valores, estes que funcionarão como verdadeiras travas morais, que evitarão a adoção de práticas indesejadas pelos sistemas, mesmo quando forçados para tal por agentes externos. Nesse sentido, são considerados uníssonos os princípios gerais que determinam aos sistemas de inteligência artificial que devem: 1) beneficiar as pessoas e o planeta, buscando o compartilhamento máximo de benefícios e prosperidade para a sociedade; 2) respeitar o estado de direito, direitos humanos, valores democráticos, diversidades; 3) garantir uma sociedade justa e leal; 4) assegurar transparência, explicabilidade, *accoutability*, segurança e confiabilidade durante toda sua vida útil; 5) promover privacidade pessoal, liberdade, privacidade dos indivíduos, proteção à propriedade intelectual; 6) possuir em seus processos ação e supervisão humanas; 7) ser responsáveis; 8) evitar uma corrida armamentista.

Apesar de ser pacífico que o substrato físico do agente não altera o tratamento jurídico a ser dado às ações e condutas, os termos da considerada quarta revolução industrial trouxeram consigo intrincados vieses axiológicos acerca da existência de uma inteligência artificial, capaz de tentar à perfeição simular um agir com intencionalidade, fazendo como se senciêntes e sapientes fossem, a ponto de que lhe permitisse a atribuição de uma personalidade jurídica, como no caso da robô Sophia. No entanto, apesar de se caminhar para a outorga de uma *e-personality* a sistemas de inteligência artificial que se apresentem mais autônomos, tal medida ainda dependerá do estabelecimento prévio de rígidas regras a serem seguidas pelos desenvolvedores acerca de uma regulação *by design*, bem como vir acompanhado de algum mecanismo que assegure financeiramente a reparação de danos eventualmente causados pelo sistema, como um seguro obrigatório, taxação de uso ou constituição de capital.

No que se refere à responsabilização civil, apesar de não existir uma previsão expressa a respeito de como devem ser tratados os danos causados por sistemas de inteligência artificial, é possível, a partir de uma análise sistêmica do ordenamento jurídico, conceber uma *teoria de responsabilidade por atos de sistemas de inteligência artificial*, que se apoia nas teorias clássicas de responsabilidade civil, em especial a teoria da responsabilidade por fato de terceiros e coisas, teoria da causa direta e imediata, teoria da causalidade múltipla, e, em última análise, as teorias do risco do proveito, *deep pocket* e risco do desenvolvimento.

Assim, mesmo considerando que hodiernamente inexiste a obrigatoriedade legal expressa da inserção de uma regulação *by design* nos sistemas de inteligência artificial, e partindo de uma hipotética premissa de que o desenvolvedor não o faça *sponte propria*, ou que

os valores inseridos não sejam bastantes para evitar a causação de danos, é certo que as teorias ordinárias são suficientes para a resolução de casos de danos causados por ato praticado por um sistema de inteligência artificial. Os diferenciais adicionais existentes se dão em razão da provável pluralidade de agentes no desenvolvimento e otimização do sistema; e da possibilidade de não ter havido, de fato, uma falha do sistema, o que poderia vir a afastar a ilicitude do ato. Deve ser aferido também, mas aí já sem nenhuma excepcionalidade, se ocorreu no caso alguma excludente de ilicitude, ou interrupção do nexo de causalidade, em especial em uma particularidade que poderá ser constatada de forma mais recorrente nesta hipótese, que é a coparticipação da vítima na causação do dano, no caso de o consumidor se apresentar como *prossumidor*.

De outro lado, não se mostra adequada a aplicação daquelas teorias que buscam, ao arrepio dos postulados basilares de responsabilidade civil, encontrar a todo custo um responsável capaz de indenizar a vítima, como a teoria do resultado mais grave, do fortuito interno e da presunção de causalidade. No caso concreto, sempre deverá ser aferida a existência de todos os elementos para a caracterização da responsabilização civil do agente conforme sua colaboração para a causação do dano.

Por fim, verifica-se que o processo legislativo no Brasil não se encontra maduro o suficiente, encontrando-se em estágio inicial todos os projetos de lei em tramitação nas casas legislativas. Ainda se faz necessário ampliar o debate público para que seja construída uma política nacional no Brasil, centrada no ser humano, que permita o pleno desenvolvimento das tecnologias e sistemas de inteligência artificial, mas o faça de forma a respeitar os valores éticos e os princípios sensíveis da sociedade e da democracia, que ao mesmo tempo possibilite a integral reparação de um eventual dano injusto que venha a causar.

REFERÊNCIAS

AGRELA, Lucas. Robô que fala, se expressa e faz ameaças ganha cidadania saudita. **Exame**, 2017. Disponível em: https://exame.com/tecnologia/robo-que-fala-se-expressa-e-faz-ameacas-ganha-cidadania-saudita/. Acesso em: 20 jul. 2020.

AGUDO, Hugo Crivilim; TEIXEIRA, Tarcisio. As perspectivas da responsabilidade civil na relação com as novas tecnologias à luz da análise econômica do direito. *In*: XXVII CONGRESSO NACIONAL DO CONPEDI PORTO ALEGRE, RS. **Anais eletrônicos** [...]. 2018, Porto Alegre, Direito, governança e novas tecnologias I [Recurso eletrônico on-line] Org. CONPEDI/UNISINOS. Coord. Têmis Limberger; Valter Moura do Carmo; Aires Jose Rover. Florianópolis: CONPEDI, 2018. p. 155-170. Disponível em: http://conpedi.danilolr.info/publicacoes/34q12098/9l053031/uaHJcZ7B7rU81QO5.pdf. Acesso em: 18 set. 2020.

ALBIANI, Christine. Responsabilidade Civil e Inteligência artificial: Quem responde pelos danos causados por robôs inteligentes? In: 3º Grupo de Pesquisa do ITS Rio (2018). **Anais eletrônicos** [...]. 2018, [*S.l.*]. Disponível em: https://itsrio.org/wp-content/uploads/2019/03/Christine-Albiani.pdf. Acesso em: 18 set. 2020.

ALEMANHA. **Artificial Intelligence Strategy**, 2018. Disponível em: https://www.ki-strategie-deutschland.de/home.html. Acesso em: 10 jul. 2020.

ALMADA, Marco. Responsabilidade civil extracontratual e a inteligência artificial. **Revista Acadêmica Arcadas**, v. 2, n. 1, p. 88-99, 2019. Disponível em: https://www.academia.edu/38132915/Responsabilidade_civil_extracontratual_e_a_intelig%C3%AAncia_artificial. Acesso em: 18 set. 2020.

ALMEIDA, Daniel Evangelista Vasconcelos. Direito à explicação em decisões automatizadas. *In*: ALVES, Isabella Fonseca. **Inteligência artificial e processo**. Belo Horizonte, São Paulo: D'Plácido, 2020. p. 95-114.

ALVES, Fernando de Brito; CORRÊA, Elídia Aparecida de Andrade. Interfaces artificiais e interpretação judicial: o problema do uso da inteligência artificial e da metodologia fuzzy na aplicação do direito**. Revista de Direito Brasileira**, Florianópolis, v. 23, n. 9, p. 5-27, maio/ago. 2019. Disponível em: https://www.indexlaw.org/index.php/rdb/article/view/3966/4518. Acesso em: 2 set. 2020.

ALVES, Isabella Fonseca; ALMEIDA, Priscilla Brandão de. Direito 4.0: uma análise sobre inteligência artificial, processo e tendências de mercado. *In*: ALVES, Isabella Fonseca. **Inteligência artificial e processo**. Belo Horizonte, São Paulo: D'Plácido, 2020. p. 47-72.

AMARAL, Jordana Siteneski do; BOFF, Salete Oro. A falibilidade do algoritmo content id na identificação de violações de direito autoral nos vlogs do youtube: embates sobre liberdade de expressão na cultura participativa. **Revista de Direito, Inovação, Propriedade Intelectual e Concorrência**, Porto Alegre, v. 4, n. 2, p. 43-62, jul./dez. 2018. Disponível em: https://www.indexlaw.org/index.php/revistadipic/article/view/4679/pdf. Acesso em: 18 set. 2020.

AMARO, Mariana. Saiba quais serão as profissões do futuro. **Você S/A**, 2019. Disponível em: https://vocesa.abril.com.br/geral/saiba-quais-sao-as-profissoes-do-futuro/. Acesso em: 16 jun. 2020.

AMORIM, Paula Fernanda Patrício de. **A crítica de John Searle à inteligência artificial:** uma abordagem em filosofia da mente. 2014. Dissertação (Mestrado em Filosofia) – Universidade Federal da Paraíba, João Pessoa, 2014.

AMOROSO, Danilo. Car tech: o carro com bafômetro que não liga em caso de embriaguez. **Techmundo**, set. 2009. Disponível em: https://www.tecmundo.com.br/internet/2712-car-tech-o-carro-com-bafometro-que-nao-liga-em-caso-de-embriaguez.htm. Acesso em: 24 jun 2020.

ANDRADE, Mariana Dionísio de; PINTO, Eduardo Régis Girão de Castro; LIMA, Isabela Braga de; GALVÃO, Alex Renan de Souza. Inteligência artificial para o rastreamento de ações com repercussão geral: o projeto victor e a realização do princípio da razoável duração do processo. **Revista eletrônica de direito processual**, Rio de Janeiro, v. 21, p. 312-335, jan./abr. 2020. Disponível em: https://www.e-publicacoes.uerj.br/index.php/redp/article/view/42717/31777. Acesso em: 18 set. 2020.

ANTUNES, Thiago Caversan; CARMO, Valter Moura do. Inteligência Artificial e decisões judiciais: uma abordagem a partir da perspectiva da Análise Econômica do Direito. **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 191-209, 2019. Disponível em: http://buscalegis.ufsc.br/revistas/index.php/observatoriodoegov/article/view/326. Acesso em: 18 set. 2020.

ASARO, Peter. Hands up, don't shoot!: HRI and the automation of police use force. **Journal of human-robot interaction**, [*s. l.*], v. 5, n. 3, dez. 2016. Disponível em: https://dl.acm.org/doi/10.5898/JHRI.5.3.Asaro. Acesso em: 20 ago. 2020.

BAUMAN, Zygmunt. **Modernidade Líquida**. Tradução de Plínio Dentzien. Rio de Janeiro: Zahar, 2001.

BAUMAN, Zygmunt. Privacidade, sigilo, intimidade, vínculos humanos - e outras baixas colaterais da modernidade líquida. *In*: BAUMAN, Zygmunt. **Danos colaterais:** desigualdades sociais numa era global. Tradução de Carlos Alberto Medeiros. Rio de Janeiro: Zahar, 2013. p. 107-120.

BIONI, Bruno Ricardo; LUCIANO, Maria. O Princípio da Precaução na Regulação de Inteligência Artificial: seriam as leis de proteção de dados o seu portal de entrada? *In*: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência artificial e direito:** ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, 2019. p. 207-232.

BLUM, Renato M. S. Opice. Aspectos jurídicos da internet das coisas. **Revista de Direito e as Novas Tecnologias**, São Paulo, v. 2, n. 2, jan./mar. 2019.

BORGES, Jorge Luis. Funes, o Memorioso. *In*: BORGES, Jorge Luis. **Ficções**. Tradução de Marco Antônio Franciotti. Barcelona: Bruguera, v. 1, 1979. p. 477-484.

BOSTROM, Nick. **Superinteligência:** Caminhos perigos e estratégias para um novo mundo. Tradução de Clemente Gentil Penna e Patrícia Ramos Geremias. Rio de Janeiro: DarkSide Books, 2018.

BOSTROM, Nick; YUDKOWSKY, Eliezer. **The ethics of artificial intelligence**. Cambridge: Cambridge University Press, 2011. Disponível em: https://www.nickbostrom.com/ethics/artificial-intelligence.pdf. Acesso em: 18 jul 2020.

BRASIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, [2016]. Disponível em: http://www.planalto.gov.br/ccivil_03/Constituicao/Constituiçao.htm. Acesso em: 23 set. 2020.

BRASIL. Câmara dos Deputados. **Projeto de Lei n. 21 de 04 de fevereiro de 2020**. Estabelece princípios, direitos e deveres para o uso de inteligência artificial no Brasil, e dá outras providências. Autoria: Eduardo Bismarck (PDT/CE). Câmara dos Deputados, 2020a. Disponível em: https://www.camara.leg.br/propostas-legislativas/2236340. Acesso em: 20 jul. 2020.

BRASIL. Câmara dos Deputados. **Projeto de Lei n. 240 de 11 de fevereiro de 2020**. Cria a Lei da Inteligência Artificial, e dá outras providências. Autoria: Léo Moraes (PODE/RO). Câmara dos Deputados, 2020b. Disponível em: https://www.camara.leg.br/propostas-legislativas/2236943. Acesso em: 20 jul. 2020.

BRASIL. Câmara dos Deputados. **Projeto de Lei n. 4.120 de 07 de agosto de 2020**. Disciplina o uso de algoritmos pelas plataformas digitais na internet, assegurando transparência no uso das ferramentas computacionais que possam induzir a tomada de decisão ou atuar sobre as preferências dos usuários. Autoria: Bosco Costa (PL/SE). Câmara dos Deputados, 2020c. Disponível em: https://www.camara.leg.br/propostas-legislativas/2259721. Acesso em: 19 ago. 2020.

BRASIL. Congresso Nacional. **Parecer (CN) n. 1**. Da comissão mista da medida provisória nº 869, DE 2018., sobre a Medida Provisória n° 869, de 2018, que altera a lei nº 13.709, de 14 de agosto de 2018, para dispor sobre a proteção de dados pessoais e para criar a Autoridade Nacional de Proteção de Dados, e dá outras providências. Presidente: Senador Eduardo Gomes. Relator: Deputado Orlando Silva. Relator revisor: Senador Rodrigo Cunha. Brasília: Congresso Nacional, 07 maio 2019a. Disponível em: https://legis.senado.leg.br/sdleg-getter/documento?dm=7948833&ts=1594019728265&disposition=inline. Acesso em: 15 jun. 2020.

BRASIL. Conselho Nacional de Justiça (CNJ). **Resolução n. 332 de 22 de agosto de 2020**. Dispõe sobre a ética, a transparência e a governança na produção e no uso de Inteligência Artificial no Poder Judiciário e dá outras providências. Brasília: Diário da Justiça, ed. 274/2020, 25 ago. 2020d. Disponível em: https://atos.cnj.jus.br/files/original191707202008255f4563b35f8e8.pdf. Acesso em: 21 set. 2020.

BRASIL. **Lei n. 10.406, de 10 de janeiro de 2012**. Institui o Código Civil. Brasília, DF: Presidência da República, [2020e]. Disponível em:

http://www.planalto.gov.br/ccivil_03/leis/2002/L10406compilada.htm. Acesso em: 23 set. 2020.

BRASIL. **Lei n. 12.414, de 09 de junho de 2011**. Disciplina a formação e consulta a bancos de dados com informações de adimplemento, de pessoas naturais ou de pessoas jurídicas, para formação de histórico de crédito. Brasília, DF: Presidência da República, [2019e]. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12414.htm. Acesso em: 23 set. 2020.

BRASIL. **Lei n. 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Brasília, DF: Presidência da República, [2020f]. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/L13709.htm. Acesso em: 23 set. 2020.

BRASIL. **Lei n. 3.071, de 1º de janeiro de 1916**. Código Civil dos Estados Unidos do Brasil. Brasília, DF: Presidência da República, [1916]. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l3071.htm. Acesso em: 23 set. 2020.

BRASIL. **Lei n. 8.078, de 11 de setembro de 1990**. Dispõe sobre a proteção do consumidor e dá outras providências. Brasília, DF: Presidência da República, [2017]. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l8078compilado.htm. Acesso em: 23 set. 2020.

BRASIL. Presidência da República. Secretaria-Geral. Subchefia para Assuntos Jurídicos. **Mensagem n. 288**. Brasília: Presidência da República, 28 jul. 2019b. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Msg/VEP-288.htm. Acesso em: 15 jun. 2020.

BRASIL. Senado Federal. **Projeto de Lei n. 5.051 de 16 setembro de 2019**. Estabelece os princípios para o uso da Inteligência Artificial no Brasil. Autoria: Senador Styvenson Valentim (PODEMOS/RN). Senado Federal, 2019c. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/138790. Acesso em: 19 ago. 2020.

BRASIL. Senado Federal. **Projeto de Lei n. 5.691 de 25 de outubro de 2019**. Institui a Política Nacional de Inteligência Artificial. Autoria: Senador Styvenson Valentim (PODEMOS/RN). Senado Federal, 2019d. Disponível em: https://www25.senado.leg.br/web/atividade/materias/-/materia/139586. Acesso em: 19 ago. 2020.

BRASIL. Superior Tribunal de Justiça (2. Seção). **Recurso Especial 1419697 RS 2013/0386285-0**. Recurso especial representativo de controvérsia (art. 543-c do cpc). Tema 710/STJ. Direito do consumidor. Arquivos de crédito. Sistema “credit scoring”. Compatibilidade com o direito brasileiro. Limites. Dano moral. Recorrente: Boa Vista Serviços S/A. Recorrido: Anderson Guilherme Prado Soares. Relator: Ministro Paulo de Tarso Sanseverino, 12 nov. 2014. Disponível em: https://ww2.stj.jus.br/processo/revista/documento/mediado/?componente=ATC&sequencial=40872564&num_registro=201303862850&data=20141117&tipo=5&formato=PDF. Acesso em: 20 jul. 2020.

BRASIL. Superior Tribunal de Justiça (2. Turma). **Recurso Especial 1117633 RO 2009/0026654-2**. Processual civil. Orkut. Ação civil pública. Bloqueio de comunidades. Omissão. Não-ocorrência. Internet e dignidade da pessoa humana. Astreintes. Art. 461, §§ 1º e 6º, do CPC. Inexistência de ofensa. Recorrente: Google Brasil Internet Ltda. Recorrido: Ministério Público do Estado de Rondônia. Relator: Ministro Herman Benjamin, 09 mar. 2010. Disponível em: https://ww2.stj.jus.br/processo/revista/documento/mediado/?componente=ATC&sequencial=8309618&num_registro=200900266542&data=20100326&tipo=5&formato=PDF. Acesso em: 20 jul. 2020.

BRASIL. Superior Tribunal de Justiça (3. Turma). **Recurso Especial 1300161 RS 2011/0190256-3**. Civil e consumidor. Internet. Relação de consumo. Incidência do CDC. Gratuidade do serviço. Indiferença. Provedor de correio eletrônico (e-mail). Fiscalização prévia das mensagens enviadas. Desnecessidade. Mensagem ofensiva. Dano moral. Risco inerente ao negócio. Inexistência. Ciência da existência de conteúdo ilícito. Bloqueio da conta. Dever. Identificação do usuário. Indicação do provedor de acesso utilizado. Suficiência. Recorrente: José Leonardo Bopp Meister. Recorrido: Microsoft Informática Ltda. Relatora: Ministra Nancy Andrighi, 19 jun. 2012. Disponível em: https://ww2.stj.jus.br/processo/revista/documento/mediado/?componente=ATC&sequencial=22902479&num_registro=201101902563&data=20120626&tipo=5&formato=PDF. Acesso em: 20 jul. 2020.

BRUNDAGE, Miles; HÄGGSTÖRM, Olle; BENTLEY, Peter J.; METZINGER, Thomas. Should we fear artificial intelligence? Scaling Up Humanity: The Case for Conditional Optmism about Artificial Intelligence. **EPRS. Euroepean Parliamentary Research Service.** Bruxelas, mar. 2018. Disponível em: https://www.europarl.europa.eu/RegData/etudes/IDAN/2018/614547/EPRS_IDA(2018)614547_EN.pdf. Acesso em: 20 jun. 2020.

CALDERÓN-VALENCIA, Felipe; MORAIS, Fausto Santos de. Inteligencia artificial y justicia: Reflexiones a partir de los casos de Brasil y Colombia. *In*: CARVAJAL, Diana María Ramírez (org.). **Justicia Digital**: Una análisis internacional em época de crisis. Medellín: Editorial Justicia y Proceso, 2020. p. 161-200.

CAMARGO, Gustavo Xavier de. Decisões judiciais computacionalmente fundamentadas: uma abordagem a partir do conceito de explainable artificial intelligence. **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 167-177, 2019. Disponível em: http://buscalegis.ufsc.br/revistas/index.php/observatoriodoegov/article/view/324. Acesso em: 18 set. 2020.

CANADÁ. **Montreal Declaration for a responsible development of artificial intelligence**. Montreal, 2018. Disponível em: https://docs.wixstatic.com/ugd/ebc3a3_bfd718945e0945718910cef164f97427.pdf. Acesso em: 10 jul. 2020.

CARINI, Lucas; MORAIS, Fausto Santos de. Governança ética para construção de confiança em sistemas de inteligência artificial. **Prim@ Facie**, v. 19, n. 40, p. 01-26, 19 dez. 2019. DOI 10.22478/ufpb.1678-2593.2020v19n40.48406. Disponível em: https://periodicos.ufpb.br/index.php/primafacie/article/view/48406. Acesso em: 18 set. 2020.

CARMO, Valter Moura do; GERMINARI, Jefferson Patrik; GALINDO, Fernando. The advances of the brazilian judicial system and the use of artificial intelligence: opposite or parallel ways towards the effectiveness of justice? **Revista Jurídica**, [*S.l.*], v. 4, n. 57, p. 249-283, out./dez. 2019. ISSN 2316-753X. DOI 10.21902/revistajur.2316-753X.v4i57.3773. Disponível em: http://revista.unicuritiba.edu.br/index.php/RevJur/article/view/3773. Acesso em: 18 set. 2020.

CAVALIERI FILHO, Sérgio. **Programa de Responsabilidade Civil**. 2. ed. São Paulo: Malheiros, 2000.

CELLA, José Renato Gaziero; WOJCIECHOWSKI, Paola Bianchi. Inteligência artificial nos processos judiciais eletrônicos. *In*: MEZZAROBA, Orides *et al* (Org.). **Direito e Novas Tecnologias**. Curitiba: Clássica, 2014. p. 271-300.

DALY, Angela; HAGENDORFF, Thilo; HUI, Li; MANN, Monique; MARDA, Vidushi; WAGNER, Ben; WANG, Wei; WITTEBORN, Saskia. Artificial Intelligence, Governance and Ethics: Global Perspectives. **The Chinese University of Hong Kong Faculty of Law Research Paper**, n. 2019-15, University of Hong Kong Faculty of Law Research Paper n. 2019/033, jul. 2019. Disponível em: https://arxiv.org/ftp/arxiv/papers/1907/1907.03848.pdf. Acesso em: 15 jul. 2020.

DIAS, José Aguiar. **Da responsabilidade civil**. 9. ed. Rio de Janeiro: Forense, v. 2, 1994.

DOMINGOS, Pedro. **O algoritmo mestre:** como a busca pelo algoritmo de machine learning definitivo recriará nosso mundo. São Paulo: Novatec, 2017.

EDWARDS, Lilian; VEALE, Michael. Slave to the algorithm? Why a "right to an explanation" is probably not the remedy you are looking for, may 2017. **16 Duke Law & Technology Review 18**. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2972855. Acesso em: 08 jul. 2020.

ENGELMANN, Wilson; WERNER, Deivid Augusto. Inteligência artificial e direito. *In*: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência artificial e direito:** ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, 2019. p. 149-178.

ESTADOS UNIDOS DA AMÉRICA. Casa Branca. Escritório de política de ciência e tecnologia. **American AI Iniciative**, 11 fev. 2019a. Disponível em: https://www.whitehouse.gov/articles/accelerating-americas-leadership-in-artificial-intelligence/. Acesso em: 10 jul. 2020.

ESTADOS UNIDOS DA AMÉRICA. Departmente of Defesense. **AI principles:** recommendations on the ethical use of artificial intelligence by Departmente of Defesense. [*S. l.*], 2019b. Disponível em: https://media.defense.gov/2019/Oct/31/2002204458/-1/-1/0/DIB_AI_PRINCIPLES_PRIMARY_DOCUMENTE.PDF. Acesso em: 11 jul. 2020.

ESTADOS UNIDOS DA AMÉRICA. Supreme Court of Wisconsin. **Caso n. 2015AP157-CR**. Recorrente: Eric L. Loomis. Recorrido: State of Wisconsin. Relator: Scott L. Horne, 05 abr. 2016. Disponível em:

https://www.wicourts.gov/sc/opinion/DisplayDocument.pdf?content=pdf&seqNo=171690. Acesso em: 08 jul. 2020.

FARIAS, Cristiano Chaves de; ROSENVALD, Nelson; BRAGA NETO, Felipe Peixoto. **Curso de Direito Civil**. 2. ed. São Paulo: Atlas, v. 3, 2015.

FERREIRA, Diogo Ramos. A responsabilidade civil dos fornecedores de inteligência artificial. **Revista de Direito e as Novas Tecnologias**, São Paulo, v. 4, jul./set. 2019. Disponível em: https://www.academia.edu/39624646/A_RESPONSABILIDADE_CIVIL_DOS_FORNECEDORES_DE_INTELIG%C3%8ANCIA_ARTIFICIAL. Acesso em: 18 set. 2020.

FLORIDI, Luciano; MITTELSTADT, Brent; WACHTER, Sandra. Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation. **International Data Privacy Law**, Oxford, v. 7, n. 2, p. 76-99, maio. 2017. Disponível em: https://ssrn.com/abstract=2903469. Acesso em: 08 jul. 2020.

FORTES, Vinícius Borges. **O direito fundamental à privacidade:** uma proposta conceitual para a regulamentação da proteção dos dados pessoais na internet no Brasil. 2015. Tese (Doutorado em Direito) – Universidade Estácio de Sá, Rio de Janeiro, 2015.

FORTES, Vinícius Borges; CELLA, José Renato Gaziero. O direito ao esquecimento na internet é um direito fundamental? *In*: CONPEDI LAW REVIEW. **Anais eletrônicos** [...]. Oñati, Espanha, v. 2, n. 2, p. 351-371, jan./jun. 2016. DOI 10.21902/clr.v2i2.310. Disponível em: https://www.indexlaw.org/index.php/conpedireview/article/view/3640/3142. Acesso em: 08 jul. 2020.

FOUCAULT, Michel. **Vigiar e punir:** nascimento da prisão. Tradução de Raquel Ramalhete. Petrópolis: Vozes, 1987.

FRAZÃO, Ana. Algoritmos e inteligência artificial: repercussões da sua utilização sobre a responsabilidade civil e punitiva das empresas. **Jota**, maio 2018. Disponível em: https://www.jota.info/opiniao-e-analise/colunas/constituicao-empresa-e-mercado/algoritmos-e-inteligencia-artificial-15052018. Acesso em: 16 jul. 2020.

FREITAS, Cinthia Obladen de Almendra; BARDDAL, Jean Paul. Análise preditiva e decisões judiciais: controvérsia ou realidade? **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 107-126, 2019. Disponível em: http://buscalegis.ufsc.br/revistas/index.php/observatoriodoegov/article/view/314. Acesso em: 18 set. 2020.

FREY, Carl Benedikt; OSBORNE, Michael A. The future of employment: How susceptible are jobs to computerisation? **Technological Forecasting and Social Change**, Oxford Martin, v. 114, p. 254-280, jan., 2013. DOI 10.1016/j.techfore.2016.08.019. Disponível em: https://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf. Acesso em: 16 jul. 2020.

FUTURE OF LIFE INSTITUTE. **Asilomar AI principles**. [*S. l.*], 2017. Disponível em: https://futureoflife.org/ai-principles/?cn-reloaded=1. Acesso em: 19 ago. 2020.

GAGLIANO, Pablo Stolze; PAMPLONA FILHO, Rodolfo. **Novo curso de direito civil:** responsabilidade civil. 17. ed. São Paulo: Saraiva Educação, v. 3, 2019.

GALINDO, Fernando ¿Inteligencia Artificial y Derecho? Sí, pero ¿cómo? **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 36-57, 2019a. Disponível em: http://buscalegis.ufsc.br/revistas/index.php/observatoriodoegov/article/view/310. Acesso em: 2 set. 2020.

GALINDO, Fernando. Inteligencia Artificial y acceso a documentación jurídica: sobre el uso de las TICs en la práctica jurídica. **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 144-166, 2019b. Disponível em: http://buscalegis.ufsc.br/revistas/index.php/observatoriodoegov/article/view/319. Acesso em: 2 set. 2020.

GALINDO, Fernando; CARMO, Valter Moura do. ¿Libertad e Internet? **DIXI**, v. 19, n. 26, maio 2017. DOI 10.16925/di.v19i26.1952. Disponível em: https://www.researchgate.net/publication/332520307_Libertad_e_Internet. Acesso em: 2 set. 2020.

GONÇALVES, Carlos Roberto. **Direito Civil Brasileiro**. 14. ed. São Paulo: Saraiva Educação, v. 4, 2019.

GOODMAN, Bryce; FLAXMAN, Seth. EU Regulations on Algorithmic Decision-Making and a "Right to Explanation". **AI Magazine**, Nova Iorque, v. 38, n. 3, p. 50-57, out. 2017. DOI 10.1609/aimag.v38i3.2741. Disponível em: https://arxiv.org/pdf/1606.08813.pdf. Acesso em: 18 set. 2020.

GUN, Murilo. **Habilidades do futuro**. [*S.I.*: *s.n.*], 2020. 1 vídeo (16min22seg). Publicado pelo canal Murilo Gun. Disponível em: https://www.youtube.com/watch?v=U3-obT9xerA. Acesso em: 22 jun. 2020.

HABERMAS, Jürgen. **O discurso filosófico da modernidade**. Tradução de Luiz Sérgio Repa e Rodnei Nascimento. São Paulo: Martins Fontes, 2000.

HAN, Byung-Chul. **Sociedade do cansaço**. Tradução de Enio Paulo Giachini. Petrópolis, RJ: Vozes, 2015.

HARARI, Yurval Noah. **Homo Deus:** uma breve história do amanhã. São Paulo: Companhia das Letras, 2016.

HARTMAN PEIXOTO, Fabiano. **Inteligência artificial e direito:** convergência ética e estratégica. Curitiba: Alteridade Editora, v. 5, 2020.

HARTMAN PEIXOTO, Fabiano; SILVA, Roberta Zumblick Martins da. **Inteligência artificial e direito**. Curitiba: Alteridade Editora, v. 1, 2019.

HODGES, A. Alan Turing: an Introductory Biography. In: TEUSCHER, C. **Alan Turing:** Life and Legacy of a Great Thinker. Tradução de Ana Cristina Ferreira. Berlim: Heidelberg, v. XXVIII, 2004. p. 3-8.

HUXLEY, Aldous. **Admirável mundo novo**. 22. ed. São Paulo: Globo, 2014.

ITECHLAW. **Responsible AI:** a global policy framework. [*S. l.*], 2019. Disponível em: https://www.itechlaw.org/sites/default/files/ResponsibleAI_Principles.pdf. Acesso em: 19 ago. 2020.

ITO, Vitor Casarini. **Venda de dados à luz da LGPD (lei geral de proteção de dados):** desafios para as relações de consumo. 2020. Dissertação (Mestrado em Direito) – Universidade de Marília, Marília, 2020.

JEROME, Joseph. Domestic Drones Should Embrace Privacy by Design. **Future of Privacy Forum**, abr. 2013. Disponível em: https://fpf.org/2013/04/05/domestic-drones-should-embrace-privacy-by-design/. Acesso em: 24 jun 2020.

JUDGE, Jenny. Are we liberated by tech – or does it enslave us? **The Guardian**, 9 dez. 2015. Disponível em: https://www.theguardian.com/technology/2015/dec/09/are-we-liberated-by-tech-or-does-it-enslave-us#:~:text=Technology%20is%20unruly.,troubling%20to%20the%20downright%20depressing.&text=But%20even%20the%20notion%20of,when%20it%20comes%20to%20tech. Acesso em: 17 jul. 2020.

KAFKA, Franz. **O Processo**. [*S. l.*]: Leya, 1925.

KASTER, Gerson Bovi; ROVER, Aires José. Revisão sistemática: ontologias e inteligência artificial aplicadas ao Direito. **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 80-93, 2019. Disponível em: http://buscalegis.ufsc.br/revistas/index.php/observatoriodoegov/article/view/311. Acesso em: 18 set. 2020.

KAUFMAN, Dora. **A inteligência artificial irá suplantar a inteligência humana?** Barueri: Estação das Letras e Cores, 2019.

KOCH, C. Towards data science. **AI Made in Germany:** The German Strategy for Artificial Intelligence, 2019. Disponível em: https://towardsdatascience.com/ai-made-in-germany-the-german-strategy-for-artificial-intelligence-e86e552b39b6. Acesso em: 10 jun. 2020.

LARA, Caio Augusto Souza. **O acesso tecnológico à justiça:** por um uso contra-hegemônico do big data e dos algoritmos. 2019. Tese (Doutorado em Direito) – Universidade Federal de Minas Gerais, Belo Horizonte, 2019.

LARSON, Jeff; MATTU, Surya; KIRCHNER, Lauren; ANGWIN, Julia. **How we analyzed the COMPAS recidivism algorithm**, 2016. Disponível em: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm. Acesso em: 05 jul. 2020.

LAPUSCHKIN, Sebastian; WÄLDCHEN, Stephan; BINDER, Alexander; MONTAVON, Grégoire; SAMEK, Wojciech; MÜLLER, Klaus-Robert. Unmasking Clever Hans Predictors and Assessing What Machines Really Learn. **Nature Communications**, v. 10, n. 1096, 2019.

Disponivel em: https://www.nature.com/articles/s41467-019-08987-4. Acesso em: 16 jul. 2020.

LASPRO, Oreste Nestor de Souza; CARBONAR, Dante O. Frazon. Processo civil na era da internet: desafios à obtenção da identidade do autor de ilícito praticado na internet. *In*: LUCON, Paulo Henrique dos Santos *et al*. **Direito processo e tecnologia**. São Paulo: Thomson Reuters, 2020. p. 503-522.

LOPES, Fabiano Tadeu. Inteligência artificial e direito penal: apontamentos para uma reflexão crítica. *In*: ALVES, Isabella Fonseca. **Inteligência artificial e processo**. Belo Horizonte, São Paulo: D'Plácido, 2020. p. 205-218.

LOPES, Giovana Figueiredo Peluso. O "direito à explicação" de decisões automatizadas no âmbito do GDPR. *In:* I CONGRESSO DE CIÊNCIA, TECNOLOGIA E INOVAÇÃO: POLÍTICAS E LEIS. **Anais eletrônicos** [...]. 2018, Belo Horizonte, 2018. DOI 10.29327/observalei.131563. Disponível em: https://even3.blob.core.windows.net/anais/131563.pdf. Acesso em: 08 jul. 2020. p. 243-250.

LUHMANN, Niklas. **O direito da sociedade**. Tradução de Saulo Krieger. São Paulo: Martins Fontes, 2016.

LUHMANN, Niklas. Sistema y función. *In*: Izuzquiza, Ignacio (org). **Sociedad y sistema:** la ambición de la teoria. Barcelona: Ediciones Paidós, 1990. p. 41-143.

MACHADO, Nilson José. **Epistemologia e didática**. 7. ed. São Paulo: Cortez, 2011.

MAGRANI, Eduardo. **A internet das coisas**. Rio de Janeiro: FGV Editora, 2018.

MAGRANI, Eduardo. **Democracia conectada:** a internet como ferramenta de engajamento político-democrático. Curitiba: Juruá, 2014.

MAGRANI, Eduardo. **Entre dados e robôs:** ética e privacidade na era da hiperconectividade. 2. ed. Porto Alegre: Arquipelágo Editorial, 2019.

MAGRANI, Eduardo; SILVA, Priscilla; VIOLA, Rafael. Novas perspectivas sobre ética e responsabilidade de inteligência artificial. *In*: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência artificial e direito:** ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, 2019. p. 115-148.

MAINI, Vishal; SABRI, Samer. **Machine learning for Humans**, 2017. Disponível em: https://everythingcomputerscience.com/books/Machine%20Learning%20for%20Humans.pdf. Acesso em: 20 jun. 2020.

MANGETH, Ana Lara Galhano. Inteligência artificial e o direito à explicação na lei geral de proteção de dados brasileira. *In:* XXVII SEMINÁRIO DE INICIAÇÃO CIENTÍFICA E TECNOLÓGICA DA PUC-RIO. **Anais eletrônicos** [...]., 2019. Disponível em: http://www.puc-rio.br/pibic/relatorio_resumo2019/download/relatorios/CCS/DIR/DIR-Ana%20Lara%20Galhano%20Mangeth.pdf. Acesso em: 11 jul. 2020.

MARQUES, Ricardo Dalmaso. Inteligência artificial e direito: o uso da tecnologia na gestão do processo no sistema brasileiro de precedentes. **Revista de Direito e as Novas Tecnologias**, v. 3, DTR\2019\35395, São Paulo: Thomson Reuters, abr./jun. 2019. Disponível em: https://www.academia.edu/39734989/INTELIG%C3%8ANCIA_ARTIFICIAL_E_DIREITO_O_USO_DA_TECNOLOGIA_NA_GEST%C3%83O_DO_PROCESSO_NO_SISTEMA_BRASILEIRO_DE_PRECEDENTES_Artificial_Intelligence_and_the_Law_the_use_of_technology_for_case_management_in_the_Brazilian_System_of_Precedents_. Acesso em: 2 set. 2020.

MASSENO, Manuel David; SANTOS, Cristiana. Personalization and profiling of tourists in smart tourism destinations – a data protection perspective. **Revista Argumentum**, Marília, v. 20, n. 3, p. 1.215-1.240, set./dez. 2019. Disponível em: http://ojs.unimar.br/index.php/revistaargumentum/article/view/1243. Acesso em: 18 set. 2020.

MATTHIAS, Andreas. The responsibility gap: ascribing responsibility for the actions of learning automata. **Ethics and information technology**, n. 6(3), p. 175-183, set. 2004.

MATURANA, Humberto; VARELA, Francisco. **A árvore do conhecimento:** as bases biológicas do entendimento humano. Tradução de Jonas Pereira dos Santos. Campinas: Editorial Psy II, 1995.

MITCHELL, Tom M. **Machine Learning**. [*S.l.*]: McGraw-Hill Science, 2017. Disponível em: http://profsite.um.ac.ir/~monsefi/machine-learning/pdf/Machine-Learning-Tom-Mitchell.pdf. Acesso em: 21 jun. 2020.

MONTESQUIEU. **O espírito das leis**. Tradução de Pedro Vieira Mota. 3. ed. São Paulo: Saraiva, 1994. Livro VI, Capítulo VII.

MORIN, Edgar. **Introdução ao pensamento complexo**. Tradução de Eliane Lisboa. Porto Alegre: Sulina, 2006.

MULHOLLAND, Caitlin. Responsabilidade civil e processos decisórios autônomos em sistemas de inteligência artificial (IA): autonomia, imputabilidade e responsabilidade. *In*: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência artificial e direito:** ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, 2019. p. 325-348.

MULHOLLAND, Caitlin; FRAJHOF, Izabella Z. Inteligência Artificial e a Lei Geral de Proteção de Dados Pessoais: breves anotações sobre o direito à explicação perante a tomada de decisões por meio de machine learning. *In*: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência Artificial e Direito:** ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, 2019. p. 265-292.

MURPHY, Kevin P. **Machine Learning:** a probabilistic perspective. Cambridge: The MIT press, 2012. Disponível em: https://www.cs.ubc.ca/~murphyk/MLbook/pml-intro-22may12.pdf. Acesso em: 21 jun. 2020.

NADKARNI, Isabel Teixeira. Eurodeputados querem regras europeias sobre robôs e inteligência artificial. **Atualidade Parlamento Europeu**, 2017. Disponível em:

https://www.europarl.europa.eu/news/pt/press-room/20170210IPR61808/eurodeputados-querem-regras-europeias-sobre-robos-e-inteligencia-artificial. Acesso em: 20 jul. 2020.

NUNES, Marcelo Guedes. **Jurimetria:** como a estatística pode reinventar o direito. 2. ed. São Paulo: Thomson Reuters Brasil, 2019.

OHM, Paul. Broken promises of privacy: responding to the surprising failure of anonymization. **UCLA Law Review**, v. 57, p. 6, 2010.

OLIVEIRA, Samuel Rodrigues de; COSTA, Ramon Silva. Pode a máquina julgar? Considerações sobre o uso de inteligência artificial no processo de decisão judicial. **Revista de Argumentação e Hermeneutica Jurídica**, Porto Alegre, v. 4, n. 2, p. 21-39, jul./dez. 2018. Disponível em: https://www.academia.edu/38733203/PODE_A_M%C3%81QUINA_JULGAR_CONSIDERA%C3%87%C3%95ES_SOBRE_O_USO_DE_INTELIG%C3%8ANCIA_ARTIFICIAL_NO_PROCESSO_DE_DECIS%C3%83O_JUDICIAL. Acesso em: 18 set. 2020.

PAIVA, Danúbia. A tutela dos dados processuais na era do "Big Data". *In*: ALVES, Isabella Fonseca. **Inteligência artificial e processo**. Belo Horizonte, São Paulo: D'Plácido, 2020. p. 157-176.

PENROSE, Roger. **A Mente Nova do Rei:** Computadores, mentes e as leis da física. Rio de Janeiro: Campus, 1991.

PEREIRA, Cáio Mário da Silva. **Responsabilidade Civil**. 12. ed. Rio de Janeiro: Forense, 2018.

PEREIRA, Sebastião Tavares. O machine learning e o máximo apoio ao juiz. **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 2-35, 2019. Disponível em: http://buscalegis.ufsc.br/revistas/index.php/observatoriodoegov/article/view/303. Acesso em: 18 set. 2020.

PLATÃO. **A República (ou Da justiça)**. Tradução de Edson Bini. 3. ed. São Paulo: Edipro, 2019.

PUTNAM, Hilary. The nature of mental states. *In*: PUTNAM, Hilary (Ed.). **Philosophical Papers**, Cambridge: Cambridge University Press, p. 429-440, 1975. DOI 10.1017/CBO9780511625251.023. Disponível em: https://web.csulb.edu/~cwallis/382/readings/482/putnam.nature.mental.states.pdf. Acesso em: 23 jun. 2020.

RABELO, César Leandro de Almeida; VIEGAS, Claúdia Mara de Almeida Rabelo; VIEGAS, Carlos Athayde Valadares. A participação da sociedade brasileira no governo eletrônico sob a perspectiva da democracia digital. **Revista Argumentum**, Marília, v. 13, p. 225-255, jan./dez. 2012. Disponível em: http://ojs.unimar.br/index.php/revistaargumentum/article/view/1093. Acesso em: 18 set. 2020.

RENDA, Andrea. Ethics, algorithms and self-driving cars - a CSI of the "trolley problem". **CEPS Policy Insights**, 2018. Disponível em: https://www.ceps.eu/wp-

content/uploads/2018/01/PI%202018-02_Renda_TrolleyProblem.pdf. Acesso em: 12 jul. 2020.

REZENDE, Solange Oliveira. **Sistemas inteligentes:** fundamentos e aplicações. Barueri: Manole, 2003.

REZER, Morgana Mezalira; FORTES, Vinícius Borges. A internet das coisas na sociedade de risco: uma análise a partir do direito à privacidade. *In*: XXVII CONGRESSO NACIONAL DO CONPEDI PORTO ALEGRE, RS. **Anais eletrônicos** [...]. 2018, Porto Alegre, Direito, governança e novas tecnologias I [Recurso eletrônico on-line] Org. CONPEDI/UNISINOS. Coord. Têmis Limberger; Valter Moura do Carmo; Aires Jose Rover. Florianópolis: CONPEDI, 2018. p. 98-117. Disponível em: http://conpedi.danilolr.info/publicacoes/34q12098/9l053031/kFt980Gr7fWk908s.pdf. Acesso em: 18 set. 2020.

ROBERTO, Enrico; CAMARA, Dennys. Danos causados por carros autônomos. **Jota**, abr. 2018. Disponível em: www.jota.info/opiniao-e-analise/artigos/danos-causados-por-carros-autonomos-06042018. Acesso em: 15 jul. 2020.

ROVER, Aires José. **Informática no direito:** inteligência artificial, introdução aos sistemas especialistas legais. Curitiba: Juruá, 2001.

RUSSEL, Stuart. Q&A: The Future of Artificial Intelligence. **University of Berkeley**, 2016. Disponível em: http://people.eecs.berkeley.edu/~russell/temp/q-and-a.html. Acesso em: 16 jun. 2020.

RUSSEL, Stuart; NORVIG, Peter. **Artificial Intelligence:** a Modern Approach. Nova Jersey: Prentice-Hall, 1995.

SÁ, Djanira Maria Radamés de. **Duplo grau de jurisdição:** Conteúdo e Alcançe Constitucional. São Paulo: Saraiva, 1999.

SAMPLE, Ian. Computer says no: why making AIs fair, accountable and transparent is crucial. **The Guardian**, 2017. Disponível em: https://www.theguardian.com/science/2017/nov/05/computer-says-no-why-making-ais-fair-accountable-and-transparent-is-crucial. Acesso em: 15 jul 2020.

SCHANK, Roger; ABELSON, Robert. **Scripts, plans, goals, and understanding:** An Inquiry into Human Knowledge Structures. New York: Halsted, 1977.

SCHREIBER, Anderson. **Novos Paradigmas da Responsabilidade Civil:** Da erosão dos filtros da reparação à diluição de danos. 2. ed. São Paulo: Atlas, 2009.

SCHUITEMA, Erik. **Tu delft robot leo learns to walk**. Delf Biorobotics Laboratory. [*S.I.*: *s.n.*], 2012. 1 vídeo (2min52seg). Publicado pelo canal TU Delft. Disponível em: https://www.youtube.com/watch?v=SBf5-eF-EIw. Acesso em: 26 jun. 2020.

SCHWAB, Klaus. **A quarta revolução industrial**. Tradução de Daniel Moreira Miranda. São Paulo: Edipro, 2016.

SEARLE, John Roger. Minds, brains, and programs. **Behavioral and Brain Sciences**, Cambridge: Cambridge University Press, v. 3, n. 3, p. 417-424, set. 1980. DOI 10.1017/S0140525X00005756 Disponível em: http://cogprints.org/7150/1/10.1.1.83.5248.pdf. Acesso em: 23 jun. 2020.

SEARLE, John Roger. **The mystery of consciousness**. New York: The New York Review of Books, 1997.

SEBASTIAN, Donald. What makes a 'smart gun' smart? **The Conversation**, jan. 2016. Disponível em: https://theconversation.com/what-makes-a-smart-gun-smart-52853. Acesso em: 24 jun 2020.

SHABBIR, Jahanzaib; ANWER, Tarique. Artificial Intelligence an its Role Near Future. **Journal of Latex Class Files**, [*s. l.*], v. 14, n. 8, ago. 2015. Disponível em: https://arxiv.org/pdf/1804.01396.pdf. Acesso em: 17 jul. 2020.

SHANE, Janelle. The danger of AI is weirder than you think. [S.I.: s.n.], 2019. 1 vídeo (10min29seg). Publicado pelo canal TED. Disponível em: https://www.youtube.com/watch?v=OhCzX0iLnOc. Acesso em: 17 jul. 2020.

SHELLEY, Marry. **Frankenstein ou o Prometeu Moderno**. Tradução de Pietro Nassetti. [Inglaterra]: Le Livros, [1818?].

SILVA, Nilton Correia da. Inteligência artificial. *In*: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência artificial e direito:** ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, 2019. p. 35-52.

SOUZA, Carlos Affonso. O debate sobre personalidade jurídica para robôs: Errar é humano, mas o que fazer quando também for robótico? **Jota**, 2017. Disponível em: https://www.jota.info/opiniao-e-analise/artigos/o-debate-sobre-personalidade-juridica-para-robos-10102017. Acesso em: 20 set. 2018.

STEIBEL, Fabro; VICENTE, Victor Freitas; JESUS, Diego Santos Vieira de. Possibilidades e potenciais da utilização da inteligência artificial. *In*: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência artificial e direito:** ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, 2019. p. 53-64.

STONE, Peter. Artificial Intelligence and life in 2030: report of the 2015-2016. **Stanford University**, 2016. Disponível em: https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fnl.pdf. Acesso em: 20 jun. 2020.

TACCA, Adriano; ROCHA, Leonel Severo. Inteligência artificial: reflexos no sistema do direito. **NOMOS**: Revista do Programa de Pós-Graduação em Direito da UFC, v. 38, n. 2, jul./dez. 2018. Disponível em: http://periodicos.ufc.br/nomos/article/view/20493/95963. Acesso em: 18 set. 2020.

TASIOULAS, John. First steps towards an ethics of robots and artificial intelligence. **Journal of Practical Ethics**, London: King's College London Law School Research Paper

Forthcoming, v. 7, n. 1, jun. 2019. Disponível em: https://ssrn.com/abstract=3413639. Acesso em: 15 jul. 2020.

TEIXEIRA, João de Fernandes. **Inteligência artificial**. São Paulo: Paulus, 2014.

TEPEDINO, Gustavo; SILVA, Rodrigo da Guia. Desafios da inteligência artificial em matéria de responsabilidade civil. **Revista Brasileira de Direito Civil - RBDCivil**, Belo Horizonte, v.21, p. 61-86, jul./set. 2019b. DOI 10.33242/rbdc.2019.03.004. Disponível em: https://rbdcivil.ibdcivil.org.br/rbdc/article/viewFile/465/308. Acesso em: 18 set. 2020.

TEPEDINO, Gustavo; SILVA, Rodrigo da Guia. Inteligência artificial e elementos da responsabilidade civil. *In*: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência artificial e direito:** ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, 2019a. p. 293-324.

TOCCHETTO, Gabriel Zanatta; GRUBBA, Leilane Serratine. Humano como conceito, humano como objeto, debate sobre novas tecnologias e o conceito de "ser humano". *In*: XXVII CONGRESSO NACIONAL DO CONPEDI PORTO ALEGRE, RS. **Anais eletrônicos** [...]. 2018, Porto Alegre, Direito, governança e novas tecnologias I [Recurso eletrônico on-line] Org. CONPEDI/UNISINOS. Coord.: Têmis Limberger; Valter Moura do Carmo; Aires Jose Rover. Florianópolis: CONPEDI, 2018. p. 264-282. Disponível em: http://conpedi.danilolr.info/publicacoes/34q12098/9l053031/60b14FP2P3449F1p.pdf. Acesso em: 18 set. 2020.

TOLLE, Eckhart. **Um Novo Mundo:** o despertar de uma nova consciência. Tradução de Henrique Monteiro. Rio de Janeiro: Sextante, 2007.

TRINDADE, André Fernando dos Reis. **Para entender Luhmann**. Porto Alegre: Livraria do Advogado Editora, 2008.

TURING, Alan. Computing machinery and intelligence. **Mind:** a quarterly review of psychology and philosophy. Oxford: Oxford University Press, v. 59, n. 236, p. 433-460, out. 1950. DOI 10.1093/mind/LIX.236.433. Disponível em: https://phil415.pbworks.com/f/TuringComputing.pdf. Acesso em: 26 jun. 2020.

UNIÃO EUROPEIA. **Directiva 95/46/CE do Parlamento Europeu e do Conselho da União Europeia de 24 de outubro de 1995**. Relativa à protecção das pessoas singulares no que diz respeito ao tratamento de dados pessoais e à livre circulação desses dados. Estrasburgo, p. 31-50, 1995. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:31995L0046&from=PT. Acesso em: 17 jul. 2020.

UNIÃO EUROPEIA. **Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho da União Europeia 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). Bruxelas, 2016. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679&from=PT. Acesso em: 17 jul. 2020.

VERONESE, Alexandre; SILVEIRA, Alessandra; LEMOS, Amanda Nunes Lopes Espiñeira. 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. **Inteligência artificial e direito:** ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, 2019. p. 233-264.

VIEGAS, Cláudia Mara de Almeida Rabelo. Inteligência artificial: uma análise da sua aplicação no Judiciário Brasileiro. *In*: ALVES, Isabella Fonseca. **Inteligência artificial e processo**. Belo Horizonte, São Paulo: D'Plácido, 2020. p. 135-156.

WEISSBERGER, Alan. Are the Internet of Things (IoT) & Internet of Everything (IoE) the Same Thing? **Viodi**, maio 2014. Disponível em: https://viodi.com/2014/05/23/are-the-Internet-of-things-iot-Internet-of-everything-iot-the-same-thing/. Acesso em: 24 jun 2020.

WOLKART, Erik Navarro. Tecnologia e precedentes: do portão de Kafka ao panóptico digital pelas mãos da jurimetria. *In*: ALVES, Isabella Fonseca. **Inteligência artificial e processo**. Belo Horizonte, São Paulo: D'Plácido, 2020. p. 7-20.

XAVIER, Luciana Pedroso; SPALER, Mayara Guibor. Patrimônio de afetação: uma possível solução para os danos causados por sistemas de inteligência artificial. *In*: FRAZÃO, Ana; MULHOLLAND, Caitlin. **Inteligência artificial e direito:** ética, regulação e responsabilidade. São Paulo: Thomson Reuters Brasil, 2019. p. 541-566.