

UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE DIREITO
PROGRAMA DE PÓS-GRADUAÇÃO EM DIREITO

JÚLIA MELO CARVALHO RIBEIRO

**Regulação da inteligência artificial à luz dos desafios impostos
pela tecnologia à responsabilidade civil**

BELO HORIZONTE
2022

JÚLIA MELO CARVALHO RIBEIRO

**Regulação da inteligência artificial à luz dos desafios impostos
pela tecnologia à responsabilidade civil**

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

Orientador(a): Prof. Dr. Marcelo de Oliveira Milagres

BELO HORIZONTE
2022

Ficha catalográfica elaborada pela bibliotecária Meire Luciane Lorena Queiroz - CRB-6/2233.

R484r Ribeiro, Júlia Melo Carvalho
Regulação da inteligência artificial à luz dos desafios impostos pela
tecnologia à responsabilidade civil / Júlia Melo Carvalho Ribeiro. - 2022.
155 f.

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

Bibliografia: f. 142-155.

1. Direito civil - Teses. 2. Responsabilidade (Direito) - Teses.
3. Inteligência artificial - Teses. I. Milagres, Marcelo de Oliveira.
II. Universidade Federal de Minas Gerais - Faculdade de Direito. III. Título.

CDU: 347.51(81)



ATA DA DEFESA DA DISSERTAÇÃO DA ALUNA JULIA MELO CARVALHO RIBEIRO

Realizou-se, no dia 26 de agosto de 2022, às 14:00 horas, na Faculdade de Direito - UFMG, da Universidade Federal de Minas Gerais, a defesa de dissertação, intitulada *Regulação da inteligência artificial à luz dos desafios impostos pela tecnologia à responsabilidade civil*, apresentada por JULIA MELO CARVALHO RIBEIRO, número de registro 2020652573, graduada no curso de DIREITO/NOTURNO, como requisito parcial para a obtenção do grau de Mestre em DIREITO, à seguinte Comissão Examinadora: Prof(a). Marcelo de Oliveira Milagres - Orientador (UFMG), Prof(a). Edgard Audomar Marx Neto (Universidade Federal de Minas Gerais), Prof(a). Leonardo Netto Parentoni (UFMG).

A Comissão considerou a dissertação:

Aprovada, tendo obtido a nota 95.

Reprovada

Finalizados os trabalhos, lavrei a presente ata que, lida e aprovada, vai assinada por mim e pelos membros da Comissão.

Belo Horizonte, 26 de agosto de 2022.

MARCELO DE OLIVEIRA MILAGRES:02377846610
Assinado de forma digital por MARCELO DE OLIVEIRA MILAGRES:02377846610
Dados: 2022.08.31 21:04:01 -03'00'

Prof(a). Marcelo de Oliveira Milagres (Doutor) nota 95.

EDGARD AUDOMAR MARX NETO
Assinado de forma digital por EDGARD AUDOMAR MARX NETO
Dados: 2022.09.02 02:26:36 -03'00'

Prof(a). Edgard Audomar Marx Neto (Doutor) nota 95.

gov.br
LEONARDO NETTO PARENTONI
Data: 05/09/2022 19:43:05-0300
Verifique em <https://verificador.itl.br>

Prof(a). Leonardo Netto Parentoni (Doutor) nota 95.

RESUMO

Propõe-se estudar o embate entre o *machine learning* e os fundamentos da responsabilidade civil sob a ótica das diversas aplicabilidades e graus de complexidade da tecnologia e, portanto, da impossibilidade de definição genérica de um regime único de responsabilização (objetivo ou subjetivo) pelos danos causados pelo comportamento emergente de sistemas de inteligência artificial. Diante dessa impossibilidade, defende-se que, no atual estágio de desenvolvimento da inteligência artificial, seria bem-vinda a aplicação de um regime regulatório baseado em princípios, circundado de microsistemas normativos de cada área de aplicação da tecnologia, quando as suas peculiaridades, riscos e complexidade demandarem regras específicas. A partir da avaliação dos principais regulamentos sobre inteligência artificial nos Estados Unidos, China, Reino Unido e Brasil, identifica-se como princípios comuns ao desenvolvimento da IA a prestação de contas (*accountability*), a explicabilidade (*explainability*) e a transparência (*transparency*). A partir desses dados, traça-se o paralelo entre esses princípios, a responsabilidade civil no Brasil e sua aplicação aos casos de danos gerados por inteligência artificial.

Palavras-chave: inteligência artificial; responsabilidade civil; regulamentos; princípios; *accountability*.

ABSTRACT

This study proposes to investigate the challenges imposed by machine learning systems on the foundations of tort law in Brazil, from the perspective of the technology's different applications and degrees of complexity, and, therefore, the impossibility of a generic definition of a single liability regime (strict or negligence-based) for damages caused by the emergent behavior of artificial intelligence systems. Faced with this impossibility, it is found that the establishment of a principle-based regulatory regime, surrounded by normative microsystems of each area of application of technology, when necessary, would be welcome in the current stage of development of AI. In order to identify the main principles related to liability and AI on the global stage, the main regulations on artificial intelligence in the United States, China, United Kingdom and Brazil are studied. From this analysis, it is identified that accountability, explainability, and transparency are essential principles to regulate the AI liability issue. Given this information, a parallel is drawn between these principles, tort law in Brazil, and its application to cases of damage caused by artificial intelligence.

Keywords: artificial intelligence; tort law; regulations; principles; accountability.

LISTA DE FIGURAS

Figura 1 – Definições de “accountability” em diversos dicionários.....	120
Figura 2 – Retorno do Google Trends para pesquisa global pelo termo “Explainable AI”.....	127
Figura 3 – Gráfico sobre a possibilidade de interpretação de um modelo algorítmico X precisão do modelo.....	128
Figura 4 – Representação do contrafactual para explicação algorítmica.....	132

LISTA DE ABREVIACOES E SIGLAS

AI	Artificial Intelligence
AIDP	New Generation of Artificial Intelligence Development Plan
ART	Artigo
CC	Cdigo Civil
CDC	Cdigo de Defesa do Consumidor
COMPAS	Correctional Offender Management Profiling for Alternative Sanction
DTIBR	Centro de Pesquisa em Direito, Tecnologia e Inovao
EBIA	Estratgia Brasileira de Inteligncia Artificial
EUA	Estados Unidos da Amrica
GDPR	General Data Protection Regulation
HAI	Institute for Human Centered Artificial Intelligence
HFT	High Frequency Trading
IA	Inteligncia Artificial
IBM	International Business Machines Corporation
LIME	Local interpretable model-agnostic explanations
LGPD	Lei Geral de Proteo de Dados
MIT	Massachusetts Institute of Technology
ML	Machine Learning
NIST	National Institute of Standards and Technology
OECD	Organizao para a Cooperao e Desenvolvimento Econmico
OXAI	Oxford Artificial Intelligence Society
PIB	Produto Interno Bruto
PL	Projeto de Lei
PLS	Projeto de Lei do Senado
PLN	Processamento de Linguagem Natural
PWC	PricewaterhouseCoopers
P&D	Pesquisa e Desenvolvimento
SHAP	Shapley Values
UK	United Kingdom
UVA	Unmanned Aerial Vehicle

SUMÁRIO

INTRODUÇÃO	10
1 INTELIGÊNCIA ARTIFICIAL	15
1.1 BREVE EXPOSIÇÃO SOBRE A EVOLUÇÃO DA TECNOLOGIA	15
1.2 DEFINIÇÃO DE INTELIGÊNCIA ARTIFICIAL	24
1.3 TIPOS DE INTELIGÊNCIA ARTIFICIAL	28
1.4 ESTADO DA ARTE: PRINCIPAIS APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL ATUALMENTE.....	37
2 RESPONSABILIDADE CIVIL E OS DESAFIOS IMPOSTOS PELA INTELIGÊNCIA ARTIFICIAL	44
2.1 A CULPA	46
2.2 O NEXO CAUSAL	54
2.3 POSSÍVEIS RESPOSTAS AO DESAFIO POSTO	58
2.3.1 Criação de personalidade jurídica para sistemas de inteligência artificial	58
2.3.2 Aplicação da responsabilidade objetiva.....	63
2.3.2.1 Atividade de risco.....	63
2.3.2.2 Responsabilidade indireta	68
2.3.2.3 Código de Defesa do Consumidor e a responsabilidade pelo defeito do produto ou serviço	72
2.4 INVIABILIDADE DA PREVISÃO DE UM REGIME DE RESPONSABILIDADE CIVIL ÚNICO PARA OS CASOS DE DANOS GERADOS POR IA.....	78
3 ESTUDO DE CASO: REGULAMENTOS DOS ESTADOS UNIDOS, DA CHINA E DO REINO UNIDO.....	93
3.1 ESTADOS UNIDOS DA AMÉRICA.....	95
3.2 CHINA	100
3.3 REINO UNIDO	105
3.4 REFLEXOS NO BRASIL	109
3.5 PRINCÍPIOS: PRESTAÇÃO DE CONTAS, EXPLICABILIDADE E TRANSPARÊNCIA	119
3.6 REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL SOB A ÓTICA DA PREVENÇÃO.....	135
CONCLUSÃO	139
REFERÊNCIAS	142

INTRODUÇÃO

Em 1965, foi cunhado o termo inteligência artificial – IA para definir o que, à época, Jonh McCarthy disse ser “that of making a machine behave in ways that would be called intelligent if a human were so behaving”¹. De lá para os dias de hoje, a inteligência artificial passou por períodos de esquecimento (os chamados *AI winters*) alternados com períodos de entusiasmo, sendo o mais recente destes gerado pelo aumento na capacidade de processamento dos computadores e a maior disponibilidade de dados na rede. Nos últimos anos observou-se a intensa difusão do uso da inteligência artificial, que, atualmente, pode-se dizer, tem presença pervasiva na vida de milhões de pessoas. Seus usos vão dos mais corriqueiros (como na definição do conteúdo que é mostrado em redes sociais a um determinado usuário, nas respostas de assistentes de voz, ou no desenvolvimento de jogos online) até os mais complexos, como os aclamados carros autônomos, os robôs de cirurgia de alta precisão, a automação de controle de pragas na agricultura, entre tantas outras aplicações, que seria impossível citá-las todas.

Sabe-se que a expressão inteligência artificial alberga diferentes técnicas de computação e variadas possibilidades de aplicação da tecnologia. Para tentar minimamente estreitar o objeto deste estudo, identifica-se como área de interesse o *machine learning*², ou aprendizado de máquina, responsável pelo grande salto de qualidade da inteligência artificial e também pelas inquietações no ramo da responsabilidade civil. Com o advento do *machine learning*, pode-se afirmar que tipos mais sofisticados de inteligência artificial têm a capacidade de aprender a partir da sua própria experiência e, conseqüentemente, responder a estímulos de forma original, sem que o seu programador possa prever seus *outputs*. Eis aqui o desafio em torno do qual este trabalho gravita, qual seja, o de se discernir, na esfera civil, quem responde e sob

¹ A máquina que se comporta de tal forma que seria chamada de inteligente caso fosse um humano assim se comportando (tradução nossa). (MCCARTHY, John; MINSKY, Leon; ROCHESTER, Nathanael; SHANNON, Claude. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence. p. 11, 1955. Disponível em: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. Acesso em: 05 out. 2021, p. 11.)

² “Machine learning is the ‘field of study that gives computers the ability to learn without being explicitly programmed’, according to computer pioneer Arthur Samuel. This does not mean that the machine programs itself per se, but rather that the machine is able to generate and store associations and facts, even if it is outside the confines of its programming”. (MCCARTHY, John; MINSKY, Leon; ROCHESTER, Nathanael; SHANNON, Claude. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, 1955. Disponível em: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. Acesso em: 05 out. 2021. p. 9.)

quais fundamentos responde pelos danos causados em decorrência de condutas emergentes da inteligência artificial.

Considerando os fundamentos da responsabilidade civil, percebe-se que a sua lógica é deturpada pelo *machine learning* em razão da diluição do nexos causal entre a conduta danosa e a programação inicial da máquina pelo seu programador e, portanto, da dificuldade de se imputar a responsabilidade ao programador/fornecedor do sistema.

Face a esses desafios, vários estudiosos passaram a buscar soluções para o embate entre a autonomia da inteligência artificial e os fundamentos da responsabilidade civil, dentre os quais cita-se, apenas a título de exemplo, Caitlin Mulholland³, Eduardo Magrani⁴ e, no cenário internacional, Laurance White e Samir Chopra⁵. Dessas tentativas surgiram teorias favoráveis à designação de personalidade jurídica à inteligência artificial para que ela própria passasse a ser o centro de imputação de responsabilidade, à aplicação da responsabilidade objetiva sob seus diversos fundamentos, entre outras propostas recorrentemente encontradas na doutrina especializada, que serão analisadas no decorrer deste trabalho.

Dada a complexidade e novidade do tema, pode-se apontar pontos favoráveis e fragilidades específicas de cada uma dessas propostas, mas o pano de fundo deste trabalho será estruturado sobre a crítica de dois principais e frequentes aspectos das soluções trazidas ao desafio ora elucidado. O primeiro desses aspectos refere-se à perspectiva fantasiosa que não raramente é acolhida na propositura de soluções ao embate entre a

³ Ao discorrer sobre a delegação total de processos decisórios por meio de automação de inteligência artificial, a autora aventa o resgate da teoria do risco do desenvolvimento “como forma de embasar a atribuição de responsabilidade aos agentes empresariais que desenvolvem ou exploram a IA, justificando-a no princípio da solidariedade social”. (MULHOLLAND, Caitlin: Responsabilidade civil e processos decisórios autônomos em sistemas de Inteligência Artificial (IA): autonomia, imputabilidade e responsabilidade, In: FRAZÃO, Ana. MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Revista dos Tribunais, 2019. p. 335.)

⁴ Esse autor, em conjunto com Priscilla Silva e Rafael Viola, afirma o seguinte sobre a questão da responsabilidade civil decorrente de danos causados pela inteligência artificial: “Inicialmente, é preciso reconhecer a necessidade de identificar, do ponto de vista social, o melhor sistema de responsabilidade civil que tenha a capacidade de proteger mais eficazmente a pessoa e seu patrimônio, de tal maneira que não reduza a sua proteção a um discurso dogmaticamente estéril. Diante da alta complexidade da atividade desenvolvida no âmbito das AIs, afigura-se mais adequado o sistema de responsabilidade objetiva que teria a capacidade de melhor proteger a vítima”. (MAGRANI, Eduardo; SILVA, Priscilla; VIOLA, Rafael. In: FRAZÃO, Ana. MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Revista dos Tribunais, 2019. p. 145.)

⁵ No livro intitulado *A Legal Theory for Autonomous Artificial Agents* os autores defendem a possibilidade de se responsabilizar objetivamente os proprietários ou usuários da Inteligência Artificial, pois, pensado-se sob a lógica econômica, a tecnologia pode gerar danos a terceiros e, portanto, a responsabilidade objetiva incentivaria o uso da inteligência artificial apenas quando os seus benefícios superarem os custos sociais que seu emprego pode causar. (CHOPRA, Samir; WHITE, Laurence F. *A Legal Theory for Autonomous Artificial Agents*. Ann Arbor: University of Michigan Press, 2011. p. 130.)

inteligência artificial e a responsabilidade civil. Essas soluções miram problemas advindos de robôs com capacidades equiparáveis a dos humanos e com um excepcional grau de autonomia e periculosidade. Não é pretensão deste trabalho limitar a perspectiva da tecnologia ao que ela representa atualmente. Este, inclusive, seria um grande equívoco, que colocaria o Direito em risco de se tornar defasado e inútil em um curto espaço de tempo, dada a evolução tecnológica exponencial⁶. Por outro lado, é necessário que o estudo jurídico se baseie em uma visão genuína e realista do objeto a ser analisado.

Segundo Carlos Affonso Souza e Vinicius Padrão, o impacto da ideia de inteligência artificial vinda da ficção científica no Direito está longe de ser desprezível. Ao discorrer sobre o assunto, eles traçam um paralelo entre o “turco” do século XVIII e a inteligência artificial na atualidade. Explica-se: o “turco” era um artefato aparentando uma espécie de robô, vestido de trajes orientais que encantava as cortes europeias da época por sua incrível habilidade no jogo de xadrez⁷.

Acontece que o Turco nunca possuiu qualquer autonomia, não sabia jogar xadrez e nem era provido de qualquer espécie de inteligência artificial. A partir de uma esperta configuração da mesa sobre a qual se apoiava o tabuleiro estava escondida uma pessoa que, na verdade, operava os movimentos do jogador mecânico⁸.

Segundo os autores, o deslinde da farsa, em 1820, frustrou várias pessoas que viam ali um futuro tecnológico promissor.

O paralelo que se traça com a inteligência artificial decorre do fato de que, aparentemente, muitos estão sob o efeito do “turco”, admirados pelas maravilhas da inteligência artificial vinda diretamente das telas de cinema, “gerando um fascínio no

⁶ Segundo Erik Brynjofsson e Andrew McAfee, no livro, “A Segunda Era das máquinas”, todo o progresso observado até os dias de hoje é apenas uma pequena amostra do que está por vir. Os referidos autores fiam a sua teoria em uma ideia central, qual seja, a ideia de desenvolvimento tecnológico exponencial. Segundo esta teoria, o progresso em termos tecnológicos não fluiria em uma ordem linear, mas sim, exponencial, a qual o raciocínio humano é incapaz de apreender, segundo alguns estudos. (FERREIRA, Matheus Costa. Progresso digital exponencial: o direito no ponto de inflexão In: PARENTONI, Leonardo (Coord.). *Direito, Tecnologia e Inovação*: volume 1. Belo Horizonte: D’Plácido, 2018.)

⁷ PADRÃO, Vinicius; SOUZA, Carlos Affonso Souza. Responsabilidade civil e os “novos turcos”: o papel da conduta humana nos danos relacionados à inteligência artificial, p. 664. In: SILVA, Rodrigo da Gaia; TEPEDINO, Gustavo (Coord.). *O Direito Civil na era da inteligência artificial*. São Paulo: Thomson Reuters Brasil, 2020. p. 663-680.

⁸ PADRÃO, Vinicius; SOUZA, Carlos Affonso Souza. Responsabilidade civil e os “novos turcos”: o papel da conduta humana nos danos relacionados à inteligência artificial, p. 664. In: SILVA, Rodrigo da Gaia; TEPEDINO, Gustavo (Coord.). *O Direito Civil na era da inteligência artificial*. São Paulo: Thomson Reuters Brasil, 2020. p. 663-680.

espectador que pode obscurecer o tratamento correto da tecnologia em seu contemporâneo estágio de desenvolvimento”⁹.

A segunda e principal crítica que se faz ao tema inteligência artificial e responsabilidade civil é o seu tratamento sob a concepção única da tecnologia, ignorando suas diferentes aplicações, complexidades e os graus de riscos que ela gera. Diante disso, despontam as seguintes indagações, às quais essa pesquisa pretende responder: é possível a aplicação de um regime único de responsabilidade (objetivo ou subjetivo) para os casos de danos causados pela inteligência artificial? Qual é a melhor forma de se regular a inteligência artificial em termos de responsabilidade civil?

Para alcançar os objetivos ora traçados, realizou-se pesquisa teórica a partir da análise da bibliografia nacional e estrangeira especializada sobre responsabilidade civil e inteligência artificial. Foram utilizadas também fontes de dados primários para o enriquecimento da pesquisa, tais como a legislação, em especial o Código Civil e o Código de Defesa do Consumidor brasileiros, além de regulamentos estrangeiros sobre o tema.

É diante destas fontes de informação que se empreendeu pesquisa interpretativa-prospectiva¹⁰, de modo a analisar as relações jurídicas já postas sobre o tema, testá-las sob a perspectiva do Direito brasileiro e, por fim, prospectar alternativas normativas ou interpretativas em relação à aplicação da responsabilidade civil frente às novas tendências tecnológicas trazidas pelo *machine learning*.

Além disso, dada a complexidade do tema e sua intercessão com outras áreas do conhecimento, notadamente com a ciência da computação, o presente estudo tem caráter multidisciplinar¹¹, uma vez que articula essas áreas do conhecimento, buscando nelas elementos para se alcançar conclusões na esfera jurídica.

Este trabalho é organizado em três capítulos, sendo que no primeiro será identificado o que se entende por inteligência artificial, a partir da análise da sua evolução histórica até chegar as suas aplicações e desafios atuais, passando pela diferenciação das suas técnicas e dos seus estágios de desenvolvimento.

⁹PADRÃO, Vinicius; SOUZA, Carlos Affonso Souza. Responsabilidade civil e os “novos turcos”: o papel da conduta humana nos danos relacionados à inteligência artificial, p. 664. In: SILVA, Rodrigo da Gaia; TEPEDINO, Gustavo (Coord.). *O Direito Civil na era da inteligência artificial*. São Paulo: Thomson Reuters Brasil, 2020. p. 663-680.

¹⁰GUSTIN, Miracy Barbosa de Sousa; DIAS, Maria Tereza Fonseca. *(Re)pensando a pesquisa jurídica: teoria e prática*. 4. ed. Belo Horizonte: Del Rey, 2013. p. 28.

¹¹ GUSTIN, Miracy Barbosa de Sousa; DIAS, Maria Tereza Fonseca. *(Re)pensando a pesquisa jurídica: teoria e prática*. 4. ed. Belo Horizonte: Del Rey, 2013. p. 72

Feito isso, no segundo capítulo passou-se à apreciação dos desafios impostos pela inteligência artificial, mais precisamente, pelo *machine learning*, ao instituto da responsabilidade civil. Para tanto, foram focalizados dois principais elementos da responsabilidade civil que são impactados pelas novidades trazidas pelo aprendizado de máquina: a culpa, em suas diversas dimensões, e o nexo de causalidade, com prestígio à análise da teoria positivada no Código Civil, qual seja, do nexo direito e imediato. No segundo capítulo também foram pormenorizadas algumas das principais propostas da doutrina aos desafios impostos pela inteligência artificial à responsabilidade civil, sendo elas: a criação de personalidade jurídica ao sistema de inteligência artificial, a aplicação da teoria do risco, da responsabilidade indireta e da responsabilidade pelo fato do produto ou serviço.

A partir da constatação das dificuldades enfrentadas pelas possíveis soluções elencadas acima, no terceiro capítulo, foi avaliada a melhor forma de regulação da IA em termos de responsabilidade civil, considerando o atual estágio de desenvolvimento da tecnologia e toda a polissemia do termo inteligência artificial. Foi dada especial atenção à regulação principiológica, partindo do estudo de caso da regulação sobre inteligência artificial em três países que se destacam no cenário mundial em relação ao desenvolvimento e ao emprego da tecnologia: os Estados Unidos, a China e o Reino Unido. Nessa análise, buscou-se princípios aplicáveis ao desenvolvimento e ao uso da inteligência artificial que se relacionam com o tema da responsabilidade civil, para, na sequência, submetê-los à crítica, considerando os desafios e características da tecnologia.

Por fim, foram apresentadas as notas conclusivas deste trabalho, ocasião em que foram sintetizadas as conclusões obtidas ao longo da dissertação sob o binômio da *concisão e precisão*.

1 INTELIGÊNCIA ARTIFICIAL

1.1 BREVE EXPOSIÇÃO SOBRE A EVOLUÇÃO DA TECNOLOGIA

A pesquisa em inteligência artificial surgiu da confluência de várias áreas do conhecimento, como matemática, filosofia, economia e estatística, bem como, ao longo do tempo, antes mesmo de o ramo ser intitulado como inteligência artificial, vários estudiosos fizeram grandes contribuições para o seu desenvolvimento¹².

Para os fins desta pesquisa, tem-se como um bom marco inicial do desenvolvimento da tecnologia o final dos anos 1940 e início dos anos 1950, quando o matemático britânico Alan Turing publicou ilustres artigos que lançaram algumas bases para a inteligência artificial. Ainda que de forma conceitual e partindo, muitas vezes, de um exercício de futurologia, o chamado “pai da computação” desenvolveu pesquisas quanto à possibilidade de uma máquina apresentar comportamento inteligente. Nesses estudos, já era possível identificar, por exemplo, ideias sobre redes neurais e aprendizado de máquina. No artigo “Intelligent Machinery”, de 1948, Turing afirma que:

It would be quite unfair to expect a machine straight from the factory to compete on equal terms with a university graduate. The graduate has had contact with human beings for twenty years or more [...] If we are trying to produce an intelligent machine, and are following the human model as closely as we can, we should begin with a machine with very little capacity to carry out elaborate operations or to react in a disciplined manner to orders [...] Then by applying appropriate interference, mimicking education¹³.

No mesmo artigo, Turing fala também sobre a semelhança do cérebro humano com o modelo de máquina idealizado à época. Segundo ele, essas máquinas seriam como um modelo mais simples de sistema nervoso, como um arranjo aleatório de neurônios¹⁴.

¹² Para um estudo aprofundado sobre as várias áreas do conhecimento que contribuíram para evolução da inteligência artificial, vide: RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010, p. 5-16.

¹³ Seria muito injusto esperar que máquinas saindo direto da fábrica competissem de igual para igual com um universitário, que teve contato com os humanos já por vinte anos ou mais. [...] se estamos tentando produzir máquinas inteligentes e estamos tentando seguir o modelo humano, devemos começar com uma máquina com muito pouca capacidade de elaborar operações ou reagir de forma disciplinada às ordens. [...] e apenas posteriormente devemos aplicar a interferência apropriada, mimetizando educação [...] (TURING, Alan M. *Intelligent Machinery*. National Physical Laboratory. 1948, p.14, tradução nossa.)

¹⁴ “The A-type unorganized machines are of interest as being about the simplest model of a nervous system with a random arrangement of neurons”. (TURING, Alan M. *Intelligent Machinery*. National Physical Laboratory, 1948, p.10.)

Ainda nos anos 40, outros personagens importantes para a história da inteligência artificial, como Wiener, Warren McCulloch, Walter Pitts, e John von Neumann organizaram uma série de trabalhos que exploravam modelos computacionais cognitivos, que tiveram grande influência no desenvolvimento da inteligência artificial naquela década¹⁵. McCulloch e Pitts trouxeram importantes contribuições ao tema das redes neurais, desenvolvendo novas teorias sobre o funcionamento do cérebro, a partir da ideia de que os neurônios e as sinapses operariam sob o racional da lógica matemática¹⁶.

Não se poderia deixar de mencionar também o aclamado artigo de Turing, denominado “Computing machinery and intelligence”, publicado em 1950¹⁷, em que ele propõe o jogo da imitação, ou “Teste de Turing”, como ficou conhecido. Turing inicia o artigo fazendo a seguinte provocação: “*As máquinas conseguem pensar?*” Ciente da dificuldade de se conceituar o que é o pensamento, ou a inteligência, ele próprio reformula a questão inicialmente proposta, sugerindo uma abordagem mais objetiva do tema, que é justamente por meio do jogo da imitação.

Nesse modelo teórico, um computador e um humano seriam interrogados por um terceiro, também humano. Todas as respostas seriam dadas por escrito, por meio de um teclado, sem o interrogador poder vê-los ou ouvi-los. Ambos, computador e humano, tentariam convencer o interrogador de que são humanos. Se, ao final das perguntas, o interrogador não conseguisse fazer a distinção entre máquina e humano, então, a máquina seria considerada como inteligente.

Segundo Russel e Norvig, houve vários exemplos de trabalhos que poderiam se enquadrar como estudo de inteligência artificial, antes mesmo de esse ramo ter se destacado das demais correntes de estudo da matemática e da ciência da computação, mas a visão de Alan Turing talvez tenha sido a mais influente, principalmente em razão das lições contidas no artigo “Computing Machinery and Intelligence”, em que, além do Teste de Turing, são aludidas outras questões centrais para o desenvolvimento da inteligência

¹⁵ “Wiener, Warren McCulloch, Walter Pitts, and John von Neumann organized a series of influential conferences that explored the new CYBERNETICS mathematical and computational models of cognition. Wiener’s book *Cybernetics* (1948) became a bestseller and awoke the public to the possibility of artificially intelligent machines”. (RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010, p. 15.)

¹⁶ “Because of the “all-or-none” character of nervous activity, neural events and the relations among them can be treated by means of propositional logic”. (MCCULLOCH, Warren. S.; PITTS, Walter. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, v. 5, p. 115-133 (1943), p. 115.)

¹⁷ TURING, Alan M. Computing Machinery and Intelligence. *Mind*, v. 49, p. 433-460. 1950, p. 1.

artificial, como o aprendizado de máquina, o aprendizado por reforço e os algoritmos genéticos¹⁸.

Apenas seis anos após a publicação do artigo de Turing, no verão americano de 1956, ocorreria um outro marco importante para o desenvolvimento da Inteligência Artificial: quatro cientistas, dentre eles Jonh McCarthy e Marvin Minsky¹⁹, realizariam o *Dartmouth Summer Project on Artificial Intelligence*. O projeto tinha como objetivo reunir dez pesquisadores, durante os dois meses daquele verão, a fim de conquistar progressos relevantes no ramo da inteligência artificial.

Na proposta do trabalho, foi deixado claro que os estudos seriam fundamentados na ideia de que, uma máquina poderia ser arquitetada para simular o aprendizado e a inteligência²⁰. Dentre os assuntos tratados no evento, já se identificava os temas de redes neurais, processamento de linguagem natural, aprendizado de máquina e automação de processos²¹.

Apesar de o evento em si ter alcançado avanços moderados, o projeto é considerado um dos grandes marcos para a inteligência artificial em razão de os pesquisadores que participaram do evento terem se colocado na vanguarda da inteligência artificial pelas décadas subsequentes, dominando os ambientes do MIT, IBM, universidades de Carnegie Mellon e Stanford, entre outros²².

¹⁸ “There were a number of early examples of work that can be characterized as AI, but Alan Turing’s vision was perhaps the most influential. He gave lectures on the topic as Early as 1947 at the London Mathematical Society and articulated a persuasive agenda in his 1950 article “Computing Machinery and Intelligence.” Therein, he introduced the Turing Test, machine learning, genetic algorithms, and reinforcement learning. He proposed the Child Programme idea, explaining “Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulated the child’s?” (RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010, p. 17.)

¹⁹ Marvin Misky foi um cientista, graduado em Harvard, Andover e Princeton, e foi professor do MIT; Jonh McCarthy foi um professor de matemática em Dartmouth. Os outros dois pesquisadores que idealizaram o projeto de Dartmouth foram Nathan Rochester, cientista da IBM, e Claude Shannon, conhecido como “o pai da teoria da comunicação”, que na época era cientista da Bell Telephone Laboratories.

²⁰ “The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it”. (MCCARTHY, John *et al.* A proposal for the Dartmouth Summer Research Project on Artificial Intelligence. 1955, p. 2. Disponível em: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. Acesso em: 05 out. 2021.)

²¹ MCCARTHY, John *et al.* A proposal for the Dartmouth Summer Research Project on Artificial Intelligence. 1955, p. 2. Disponível em: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. Acesso em: 05 out. 2021.

²² “The Dartmouth workshop did not lead to any new breakthroughs, but it did introduce all the major figures to each other. For the next 20 years, the field would be dominated by these people and their students and colleagues at MIT, CMU, Stanford, and IBM.” (RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010, p. 18.)

Tão importante quanto isso foi o fato de o termo inteligência artificial ter sido cunhado naquela oportunidade. A criação do nome foi atribuída a Jonh McCarthy, que o teria concebido para diferenciar o seu ramo de estudo, baseado em lógica simbólica²³, de outros ramos de pesquisa já consolidados, como a própria estatística²⁴⁻²⁵.

Fato é que o termo, que insinua uma simbiose entre a tecnologia e um atributo predominantemente humano, qual seja, a inteligência, fez com que o tema extrapolasse o âmbito científico computacional, aguçando o interesse da indústria cinematográfica, das ciências sociais e da mídia. Esse truque de *marketing*, intencional ou não, é considerado como um dos grandes propulsores do interesse pela tecnologia.

Vale lembrar que o ambiente de desenvolvimento da inteligência artificial à época era impulsionado por grandes investimentos financeiros, por matemáticos extremamente criativos e muitas ideias especulativas sobre o que seria possível no futuro, conforme descreve Meredith Broussard²⁶. Ilustrativa é a entrevista concedida por Marvin Minsky, um dos idealizadores do evento de 1956, em que ele narra a facilidade em se obter investimentos para os seus projetos no MIT:

Jonh McCarthy and I had started working on artificial intelligence in about 1958, 1959, when we both came to MIT. We had a couple of students working on it. Jerry Wiesner came by once and said, how are you doing? We said, we're doing fine, but it would be nice if we could support three or four more graduate students. [...] Two days later we had this little labo of three or four rooms, and

²³ De acordo com Jerry Kaplan, a lógica simbólica é o “ramo da matemática que estuda a representação de conceitos e declarações como símbolos, para, em seguida, definir formas de manipular esses símbolos, transformando-os novamente em conceitos e declarações, por meio de raciocínio dedutivo ou indutivo”. (KAPLAN, Jerry. *Artificial Intelligence: What everyone needs to know*. Oxford: Oxford University Press, 2016, p. 14, tradução nossa). Um exemplo de uso da lógica simbólica na área de inteligência artificial são os casos de processamento de linguagem natural, em que a linguagem natural é transformada em símbolos, para posterior recomposição dos símbolos em sentenças.

²⁴ “McCarthy selected the term artificial intelligence for the conference in part to distinguish his and his colleagues’ work from the more established field of cybernetics— ‘the scientific study of control and communication in the animal and the machine’”. (KAPLAN, Jerry. *Artificial Intelligence: What everyone needs to know*. Oxford: Oxford University Press, 2016, p. 13.)

²⁵ “The first answer is that AI from the start embraced the idea of duplicating human faculties such as creativity, self-improvement, and language use. None of the Other fields were addressing these issues. The second answer is methodology. AI is the only one of these fields that is clearly a branch of computer science (although operations research does share an emphasis on computer simulations), and AI is the only field to attempt to build machines that will function autonomously in complex, changing environments”. (RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010, p. 18.)

²⁶ BROUSSARD, Meredith. *Artificial Intelligence: How computers misunderstand the world*. Cambridge: MIT Press, 2018, p. 71.

a large pile of Money that IBM had given to MIT for the advancement of computer science and nobody knew what to do with it. So they gave it to us²⁷.

Todo esse entusiasmo com os avanços da inteligência artificial também ganhou a atenção de Hollywood e da ficção científica de forma geral. O mesmo Marvin Minsky era frequentado por Isaac Asimov, autor da obra de ficção *Eu, robô*, bem como foi consultado por Arthur C. Clarke na construção do personagem Hal de *2001: uma odisseia no espaço*.

A criatividade e a ousadia de Minsky, somadas à sua proximidade com autores das principais obras ficcionais da época, muitas vezes fazia com que ciência e fantasia se misturassem. Cita-se, por exemplo, a anedota contada pelo próprio Minsky, em que ele, convencido por Arthur Clarke sobre a sua ideia de criação de um elevador espacial (algo que pudesse trazer e levar coisas do espaço para a Terra e vice-versa), mobilizou eminentes cientistas do Laboratório Nacional Lawrence Livermore, nos Estados Unidos, durante todo um período de seis meses, para o desenvolvimento do referido elevador. O projeto, evidentemente, fracassou²⁸.

Citam-se aqui estes bastidores do desenvolvimento da tecnologia para ilustrar o ambiente inicial em que a Inteligência Artificial se desenvolveu, o que influencia fortemente em como a tecnologia é vista atualmente e, inclusive, a perspectiva sobre a qual são estudadas as suas implicações na responsabilidade civil.

Logo após a realização do projeto de Dartmouth, os seus participantes estavam muito otimistas em relação ao desenvolvimento da inteligência artificial. Hebert Simon teria afirmado que “a capacidade das máquinas de pensar, aprender e criar aumentaria tão rapidamente que, em um futuro visível, a gama de problemas com os quais a máquina conseguirá lidar seria coextensiva com a qual a mente humana lida²⁹.” Marvin Minsky

²⁷ Jonh McCarthy e eu começamos a trabalhar com inteligência artificial por volta de 1958, 1959, quando ambos viemos para o MIT. Tínhamos alguns alunos trabalhando conosco nos projetos. Jerry Wiesner apareceu uma vez e perguntou: como vai você? Dissemos que estávamos indo bem, mas que seria bom se pudéssemos ter o apoio de mais três ou quatro alunos de graduação. [...] Dois dias depois, tínhamos este pequeno laboratório de três ou quatro salas, e uma grande pilha de dinheiro que a IBM havia dado ao MIT para o avanço da ciência da computação e ninguém sabia o que fazer com o dinheiro. Então eles o deram para nós. (BROUSSARD, Meredith. *Artificial Intelligence: How computers misunderstand the world*. Cambridge: MIT Press, 2018, p. 71, tradução nossa.)

²⁸ BROUSSARD, Meredith. *Artificial Intelligence: How computers misunderstand the world*. Cambridge: MIT Press, 2018, p. 71.

²⁹ RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010, p. 21.

também teria apostado, em 1970, que dentro de três a oito anos existiriam sistemas com inteligência geral, tal qual ao do homem”³⁰.

Ainda em 1956, o sistema de Arthur Samuel que jogava damas foi apresentado na televisão, gerando forte impressão na população em relação aos avanços da tecnologia. Nessa época, McCarthy e Minsky também alcançaram alguns importantes avanços. McCarthy, no MIT, desenvolveu a linguagem computacional *Lisp*, que foi altamente difundida para o uso na inteligência artificial. Já Minsky desenvolvia, junto aos seus alunos em Stanford, os micromundos, como eram chamados os sistemas que conseguiam resolver problemas simples, como os de lógica.

Quando se tentou abranger os sistemas utilizados nos micromundos de Minsky para realidades um pouco mais intrincadas é que se observou um dos primeiros períodos de baixa da inteligência artificial. A forma como os problemas eram resolvidos nos micromundos, ou seja, experimentando diferentes combinações em etapas até a solução ser encontrada, simplesmente não era suficiente para tratar problemas mais complexos³¹.

Outro problema enfrentado na época foi a falta de conhecimento prévio dos sistemas sobre os assuntos que lhes eram apresentados. Essa dificuldade é bem ilustrada quando se pensa em tradução automática, conforme expõem Russel e Norvig: por volta de 1957, o governo dos Estados Unidos financiou programas para tradução de artigos científicos em Russo, quando do lançamento do Sputnik. O programa frustrou, tendo em vista que a conversão simples das palavras em russo suas correspondentes em inglês, não era suficiente para tornar os textos inteligíveis. Sem nenhum histórico dos idiomas, o sistema não logrou êxito³².

A Inteligência Artificial viu um cenário de reviravolta do período de retração entre as décadas de 1970 e 1980, quando os sistemas especialistas começaram a ser desenvolvidos. Estes funcionavam como mecanismos de inferência, baseados nos conceitos de lógica simbólica e partindo de uma base de dados considerável sobre uma área de conhecimento específica. Para alimentar essa base de dados, era necessário coletar informações de profissionais especialistas na área. Essas informações eram trabalhadas

³⁰ DARRACH, Brad. Meet Shaky, the first electronic person. The fascinating and fearsome reality of a machine with a mind of its own. *Life Magazine*, 1970. Disponível em: https://books.google.com.br/books?id=2FMEAAAAMBAJ&printsec=frontcover&hl=pt-BR&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false. Acesso em: 27 jun. 2022.

³¹ RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010. p. 21.

³² RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010. p. 21.

com regras e inferências do sistema para encontrar soluções. Segundo Jerry Kaplam, “a ideia era capturar e duplicar a escassa experiência humana em forma computável, na esperança de que esse recurso se tornasse disponível de forma mais ampla e econômica”³³.

Essa aplicação teve forte apelo para as indústrias que, a partir de 1980, passaram a investir na tecnologia. De acordo com Russel e Norvig, quase toda companhia de maior porte nos Estados Unidos tinha o seu próprio time focado em inteligência artificial e estava ou usando, ou desenvolvendo sistemas especialistas³⁴. Nesse cenário, surgiram também várias empresas menores de tecnologia, as *startups* da época, que buscavam desenvolver seus próprios sistemas especialistas para posterior implementação em grandes empresas. Os sistemas eram utilizados em áreas como medicina, finanças e produção.

Contudo, os sistemas especialistas também apresentaram seus problemas, considerando, principalmente, que eram extremamente específicos, o que impossibilitava sua aplicação em outras categorias do conhecimento que não a originalmente programada. Além disso, esses sistemas se mostraram caros. Primeiramente, porque era necessário que profissionais da área de interesse alimentasse o programa com o seu conhecimento e, em segundo lugar, pois o programa não tinha a habilidade de aprender pela experiência, sendo necessária sua constante atualização³⁵.

Depois de grande euforia na década de oitenta, os sistemas especialistas caíram em desuso, diante da incapacidade de empresas de cumprirem promessas extravagantes relacionadas ao potencial desses sistemas³⁶. Veio, então, um novo “inverno” da inteligência artificial.

Após esse período de desaquecimento, que segundo alguns autores durou até aproximadamente 1993³⁷, a tecnologia entrou em uma outra onda positiva, o que foi possível graças a uma série de fatores. Cita-se, por exemplo, a retomada às redes neurais com o desenvolvimento da chamada retroprogramação, por David Rumelhart e Ronald J.

³³ KAPLAN, Jerry. *Artificial Intelligence: What everyone needs to know*. Oxford: Oxford University Press, 2016, p. 22.

³⁴ RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010, p. 24.

³⁵ TAULLI, Tom. *Introdução à Inteligência Artificial*. Tradução A Press Media, LLC. São Paulo: Novatec, 2020, p. 31.

³⁶ RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010, p. 24.

³⁷ TAULLI, Tom. *Introdução à Inteligência Artificial*. Tradução A Press Media, LLC. São Paulo: Novatec, 2020, p. 32.

Willians³⁸. Russel e Norvig citam também a adoção do método científico pelos pesquisadores de inteligência artificial como um dos recentes impulsionadores da tecnologia. Como os autores mencionam, o campo de estudo da IA desenvolveu-se a partir de uma ruptura com métodos científicos clássicos, conforme já mencionado em relação à criação do próprio nome inteligência artificial, para diferenciar os estudos de outros mais tradicionais – o que, inclusive, foi um dos motivos que permitiu a aproximação de outros ramos de conhecimento e das artes à inteligência artificial, levando o seu estudo a especulações de menor conteúdo científico. Porém, com o desenvolvimento da tecnologia, observou-se a sua reaproximação aos métodos clássicos, como a estatística e a teoria da informação³⁹. A partir disso, as hipóteses trazidas no ramo da inteligência artificial passaram a ser rigorosamente testadas para posterior aceitação ou rejeição, o que robusteceu a pesquisa em IA, permitindo a melhor investigação e compreensão dos seus problemas.

Por fim, o mais recente período de entusiasmo em torno da inteligência artificial foi gerado pela maior disponibilidade de dados na rede (que são o principal combustível dos sistemas de IA, como se verá no próximo tópico) e pelo aumento na capacidade de processamento dos computadores. Pode-se dizer, assim, que um dos maiores saltos da inteligência artificial das últimas duas décadas se deu em razão do desenvolvimento, não da tecnologia em si, mas das tecnologias adjacentes à inteligência artificial. Sobre esse fato, Russel e Norvig afirmam:

Throughout the 60-year history of computer science, the emphasis has been on the algorithm as the main subject of study. But some recent work in AI suggests that for many problems, it makes more sense to worry about the data and be less picky about what algorithm to apply.⁴⁰

Apenas a título de exemplo, ainda segundo os autores, isso significa que um algoritmo pouco complexo, que tenha acesso a cem milhões de palavras sem qualquer

³⁸ TAULLI, Tom. *Introdução à Inteligência Artificial*. Tradução A Press Media, LLC. São Paulo: Novatec, 2020, p. 33.

³⁹ RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010, p. 35.

⁴⁰ Durante os 60 anos de história da ciência da computação, a ênfase dos estudos sempre foi no algoritmo. Mas alguns trabalhos mais recentes sobre IA sugerem que, para muitos problemas, faz mais sentido se preocupar com os dados e ser menos exigente em relação a qual algoritmo aplicar. (RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010, p. 27, tradução nossa.)

rotulação prévia, tem melhor performance que o melhor algoritmo já conhecido que tenha acesso a um milhão de palavras⁴¹.

Jerry Kaplan também reforça esse entendimento, afirmando que “pode até ser que um novo algoritmo, ou novo conceito, gere um progresso significativo no ramo da ciência da computação, mas, na maior parte das vezes, é o contrário – os avanços em armazenamento, rede, disponibilidade de dados, ou comunicação é que geram oportunidades para desenvolvimento de novas técnicas de IA”⁴².

Quanto à capacidade dos hardwares, Turing já afirmava que “a capacidade de memória de uma máquina, mais do que qualquer outra coisa, determina a complexidade do seu possível comportamento” (tradução nossa)⁴³.

Vale lembrar aqui da sentença proferida por Gordon Moore em 1965, que posteriormente ficou conhecida como Lei de Moore. Ele, à época, comandava a *Fairchild Semiconductor*, empresa que produziu o primeiro circuito integrado (*chip*) comercializável nos Estados Unidos. Em artigo publicado para a *Electronics Magazine*, Moore afirma:

The complexity for minimum component costs has increased at a rate of roughly a factor of two per year. Certainly over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least ten years⁴⁴.

Com essa sentença, Moore prenuncia que a capacidade de chips de processamento dobraria a cada um ano e meio, durante, pelo menos, dez anos, sendo que o preço dos chips decresceria de forma inversamente proporcional ao aumento de sua capacidade. Segundo Erik Brynjolfsson e Andrew McAfee, o único erro de Moore foi ter sido conservador em relação ao tempo em que sua “lei” seria aplicada, considerando que, pelo menos nas quatro décadas subsequentes à sua afirmação, o crescimento exponencial da

⁴¹ RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010, p. 27.

⁴² KAPLAN, Jerry. *Artificial Intelligence: What everyone needs to know*. Oxford: Oxford University Press, 2016, p. 138.

⁴³ TURING, Alan M. *Intelligent Machinery*. National Physical Laboratory, 1948, p. 110.

⁴⁴ A complexidade dos custos dos componentes mínimos aumentou a uma taxa de, aproximadamente, dois por ano. Certamente, no curto prazo, pode-se esperar que essa taxa se mantenha, ou aumente. No longo prazo, a taxa de aumento é um pouco mais incerta, embora não haja razão para acreditar que não permanecerá quase constante por pelo menos dez anos. (MOORE. Gordon E. Cramming more components onto integrated circuits. *Electronics Magazine*, v. 38, p. 2, 1965. Tradução nossa. Disponível em: <https://newsroom.intel.com/wp-content/uploads/sites/11/2018/05/moores-law-electronics.pdf>. Acesso em: 05 out. 2021.)

capacidade de computação foi observado⁴⁵. Matheus Ferreira corrobora esse entendimento a partir dos seguintes exemplos:

Não é difícil observar os impactos da Lei de Moore no progresso tecnológico no mundo real. O poder computacional em um telefone celular em 2008 era no mínimo mil vezes maior que todo o poder computacional do MIT em 1965 – e um milhão de vezes mais barato. Em 2017, o Iphone 6 era 120 milhões de vezes mais rápido que o Apollo Guidance Computer, computador responsável por direcionar o homem à lua em 1969⁴⁶.

Com a maior disponibilidade de dados estruturados e com o avanço dos hardwares, os caminhos para a disseminação da inteligência artificial foram abertos de forma considerável, até que se alcançasse o cenário atual, que será melhor explorado a partir de exemplos de aplicações da inteligência artificial mais adiante neste trabalho.

1.2 DEFINIÇÃO DE INTELIGÊNCIA ARTIFICIAL

Apesar de o termo inteligência artificial ter nascido ainda em 1956, como se elucidou no tópico anterior, até os dias de hoje, não há consenso em relação à sua definição. Como já mencionado neste trabalho, Jonh McCarthy, a quem se atribui a criação do termo, o conceituou como “that of making a machine behave in ways that would be called intelligent if a human were so behaving”⁴⁷.

Nesse ponto, também são valiosos os estudos de Russel e Norvig, que traçaram as quatro maiores correntes de definição da tecnologia, sendo que cada uma se aproxima mais de uma abordagem de pesquisa sobre a inteligência artificial, quais sejam: (i) a capacidade do algoritmo de pensar como um humano⁴⁸; (ii) a demonstração de

⁴⁵ BRYNJOLFSSON, Erik; MCAFEE, Andrew. *Second Machine Age*. Nova York: W.W. Norton & Company, 2014, p. 32.

⁴⁶ FERREIRA, Matheus Costa. Progresso digital exponencial: o direito no ponto de inflexão In: PARENTONI, Leonardo (Coord.). *Direito, Tecnologia e Inovação*: volume 1. Belo Horizonte: D'Plácido, 2018, p. 43.

⁴⁷ A máquina que se comporta de tal forma que seria chamada de inteligente caso fosse um humano assim se comportando (tradução nossa). MCCARTHY, John; MINSKY, Leon; ROCHESTER, Nathanael; SHANNON, Claude. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence. p. 11, 1955. Disponível em: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. Acesso em: 05 out. 2021.

⁴⁸ Essa abordagem parte da visão da ciência cognitiva, que busca compreender exatamente como a mente humana funciona, para, então, representá-la em um modelo científico de programa computacional.

comportamento humano pelo sistema⁴⁹; (iii) a capacidade da máquina de pensar racionalmente (sendo que o pensamento racional não é aquele mais fiel à forma humana de se pensar, mas, sim, o que visa à obtenção do melhor resultado)⁵⁰; e (iv) a capacidade de agir, ou seja, de se comportar racionalmente⁵¹. A partir disso, os autores dão exemplo de possíveis definições de inteligência artificial sob esses quatro eixos interpretativos:

Pensamento humano	Pensamento racional
<p>“The exciting new effort to make computers think... <i>machines with minds</i>, in the full and literal sense” (HAUGELAND, 1985)⁵².</p> <p>“[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning...” (BELLMAN, 1978)⁵³.</p>	<p>“The study of mental faculties through the use of computational models” (CHARNIAK; MCDERMOTT, 1985)⁵⁴.</p> <p>“The study of the computations that make it possible to perceive, reason, and act” (WINSTON, 1992)⁵⁵.</p>
Comportamento humano	Comportamento racional
<p>“The art of creating machines that perform functions that require intelligence when performed by people” (KURZWEIL, 1990)⁵⁶.</p>	<p>“Computational Intelligence is the study of the design of intelligent agents” (POOLE <i>et al.</i>, 1998)⁵⁸.</p>

⁴⁹ A segunda corrente ainda parte da comparação da tecnologia com o humano, mas não a partir de um estudo da mente, mas da possibilidade de a tecnologia demonstrar um comportamento humano. Nesse caso, a ação (o que parece ser) vale mais do que o pensamento e a estrutura da mente. Russel e Norvig citam o Teste de Turing como um dos exemplos dessa abordagem.

⁵⁰ Aqui já não mais se preocupa com a fidelidade do sistema à forma de pensamento humano, mas sim que os outputs da máquina sejam ideais dentro do que ela se propõe, o que está intimamente ligado à abordagem de estudo da tecnologia baseada em lógica.

⁵¹ A quarta forma de se conceituar inteligência artificial parte do tratamento da tecnologia como um agente racional (ou seja, daquele que age racionalmente). Esse agente, por sua vez, é aquele que atua de modo a atingir o melhor resultado esperado dentro de uma determinada situação.

⁵² O entusiasmante esforço de fazer que os computadores pensem... máquinas com mentes, no sentido pleno e literal. (HAUGELAND, John. *Artificial Intelligence: The Very Idea*. 1985. MIT Press: Cambridge *apud* RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 3. ed. New Jersey: Prentice-Hall, 2010, p. 2, tradução nossa.)

⁵³ [A automação de] atividades que associamos ao pensamento humano, como a tomada de decisão, a resolução de problemas, a aprendizagem... (BELLMAN, Richard. *An Introduction to Artificial Intelligence: Can Computers Think?* Boyd & Fraser Publishing Company: 1978 *apud* RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 3. ed. New Jersey: Prentice-Hall, 2010, p. 2, tradução nossa.)

⁵⁴ O estudo das faculdades mentais por meio do uso de modelos computacionais. (CHARNIAK, Eugene; MCDERMOTT, Drew. *Introduction to Artificial Intelligence*, 1985. Addison-Wesley Publishing Company *apud* RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 3. ed. New Jersey: Prentice-Hall, 2010, p. 2, tradução nossa.)

⁵⁵ O estudo da ciência da computação que permite com que máquina percebam, raciocinem e ajam. (WINSTON, P. H. (1992). *Artificial Intelligence*. 3. ed. ADDISON-WESLEY *apud* RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 3. ed. New Jersey: Prentice-Hall, 2010, p. 2, tradução nossa.)

⁵⁶ A arte de criar máquinas que realizam funções que requerem inteligência quando realizadas por pessoas. (KURZWEIL, R. *The Age of Intelligent Machines*, 1990. MIT Pres *apud* RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 3. ed. New Jersey: Prentice-Hall, 2010, p. 2, tradução nossa.)

⁵⁸ A Inteligência Computacional é o estudo do design de agentes inteligentes. (POOLE, David; MACKWORTH, Alan; GOEBEL, Randy. *Computational intelligence: A logical approach*. 1998. Oxford University Press: New York *apud* RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 3. ed. New Jersey: Prentice-Hall, 2010, p. 2, tradução nossa.)

“The study of how to make computers do things at which, at the moment, people are better” (RICH; KNIGHT, 1991) ⁵⁷ .	“AI...is concerned with intelligent behavior in artifacts” (NILSSON, 1998) ⁵⁹ .
--	--

Para definições mais contemporâneas da tecnologia, cita-se o relatório *Artificial Intelligence and Life In 2030*, publicado em 2016 pela universidade de Stanford, que contou com a colaboração de várias pesquisadores da área, como Ryan Calo e Erik Brynjolfsson, em que a inteligência artificial é definida como “a atividade dedicada a tornar as máquinas inteligentes, sendo a inteligência, por sua vez, a qualidade que permite que algo funcione adequadamente de acordo com a percepção do seu ambiente”⁶⁰. Cita-se, também, o *Dicionário Oxford*, que atualmente define inteligência artificial como “a teoria e o desenvolvimento de sistema de computação aptos a executar tarefas que normalmente requerem inteligência humana, tais como percepção visual, reconhecimento de fala, tomada de decisão, tradução entre idiomas”⁶¹.

Apesar de a inteligência artificial muito ter se desenvolvido dos anos 50 até atualmente, ganhando novas modalidades e definições, nota-se que, no presente, a conceituação mais difundida para este tipo de sistema ainda passa por uma perspectiva antropológica, uma vez que compara o funcionamento do programa de computação a atributos biológicos, majoritariamente humanos. Jerry Kaplan é um dos críticos desta aproximação entre a tecnologia e a inteligência humana, por entender que sequer é possível definir esta última e que, segundo ele, a relação entre sistemas de inteligência artificial e organismos biológicos é, na maior parte das vezes, apenas metafórica e inspiracional⁶².

⁵⁷ O estudo de como fazer com que os computadores façam coisas em que, no momento, os humanos são melhores. (RICH, Elaine; KNIGHT, KevIn: *Artificial Intelligence*. 2. ed. 1991. McGraw-Hill Education: New York *apud* RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 3. ed. New Jersey: Prentice-Hall, 2010, p. 2, tradução nossa.)

⁵⁹ IA foca no comportamento inteligência de artefatos. (NILSSON, Nils. John. *Artificial Intelligence: A New Synthesis*. 1998. Morgan Kaufmann Publishers: San Francisco *apud* RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 3. ed. New Jersey: Prentice-Hall, 2010, p. 2, tradução nossa.)

⁶⁰ STONE, Peter *et al.* *Artificial Intelligence and Life in 2030. One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel*. Stanford University, Stanford, CA, 2016. Disponível em: <https://ai100.stanford.edu/2016-report> . Acesso em: 05 out. 2021.

⁶¹ INTELIGÊNCIA ARTIFICIAL. *Dicionário online de Oxford*. Disponível em: <https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095426960> . Acesso em: 05 out. 2021.

⁶² KAPLAN, Jerry. *Artificial Intelligence: What everyone needs to know*. Oxford: Oxford University Press, 2016, p. 16.

Para além disso, conforme já exposto neste trabalho, acredita-se que essa aproximação, entre o desenvolvimento de uma tecnologia e as capacidades humanas, teve um grande apelo para a imaginação e a ficção, que muitas vezes representam a inteligência artificial como uma espécie de super-humano, apto a feitos extraordinários. Essa percepção da inteligência artificial logo se difundiu e, certamente, impactou na forma como algumas áreas do conhecimento, como o Direito, percebem a tecnologia.

Conforme elucida Matthew U. Scherer, a falta de assertividade sobre a conceituação da inteligência artificial leva a um grande desafio, inclusive no campo regulatório. Segundo ele, para se regular, deve-se conhecer com precisão o objeto regulado⁶³. Na tentativa de elucidar a que se refere os estudos ora propostos, e também para que não se caia na armadilha de estudar um objeto indefinido, para o presente trabalho, ter-se-á como ponto de partida a definição de inteligência artificial proposta na Resolução do Parlamento Europeu (2020/2014(INL)), de 20 de outubro de 2020, sobre responsabilidade civil e IA⁶⁴, para conceitualizá-la como “um sistema baseado em software, integrado a um dispositivo físico (ainda que este seja um computador pessoal ou um smartphone), que apresenta um comportamento que simula a inteligência humana, colhendo e tratando dados, analisando e interpretando o seu ambiente e tomando decisões – com um determinado nível de autonomia – para atingir objetivos específicos”⁶⁵.

Entende-se que esta é uma boa definição de inteligência artificial para os objetivos perseguidos por esta dissertação pelos seguintes motivos: (i) na definição proposta, a comparação com a inteligência humana é apenas do ponto de vista da “simulação”, o que afasta a discussão sobre a tecnologia ser efetivamente inteligente, ter autoconsciência, dentre outros debates que fogem do escopo deste trabalho; (ii) ainda que se tenha uma

⁶³ SCHERER, Matthew U. Regulating Artificial Intelligence Systems. *Harvard Journal of Law & Technology*, v. 29, n. 2, p. 358, primavera 2016. Disponível em: <https://ssrn.com/abstract=2609777>. Acesso em: 08 de maio de 2021.

⁶⁴ UNIÃO EUROPEIA. Parlamento Europeu. Regime de responsabilidade civil aplicável à inteligência artificial. Resolução do Parlamento Europeu, de 20 de outubro de 2020, que contém recomendações à Comissão sobre o regime de responsabilidade civil aplicável à inteligência artificial (2020/2014(INL)). Bruxelas, 20 out. 2017. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_PT.html. Acesso em: 08 de jun. de 2021.

⁶⁵ A definição precisa de sistemas de IA dada pela referida resolução, conforme seu art. 3º, a) é: “um sistema baseado em software ou integrado em dispositivos físicos e que apresenta um comportamento que simula inteligência, nomeadamente recolhendo e tratando dados, analisando e interpretando o seu ambiente e tomando medidas – com um determinado nível de autonomia – para atingir objetivos específicos”. UNIÃO EUROPEIA. Parlamento Europeu. Regime de responsabilidade civil aplicável à inteligência artificial. Resolução do Parlamento Europeu, de 20 de outubro de 2020, que contém recomendações à Comissão sobre o regime de responsabilidade civil aplicável à inteligência artificial (2020/2014 (INL)). Bruxelas, 20 out. 2017. (Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_PT.html. Acesso em: 05 out. 2021.)

carga subjetiva na conceituação da inteligência humana, a definição proposta busca identificar ao que ela se refere, apontando como inteligente a capacidade de tratar dados, analisar e interpretar o ambiente ao entorno para a tomada de decisões, visando ao alcance de um objetivo; e (iii) o fato de a definição mencionar a autonomia da inteligência artificial, que é um pré-requisito fundamental para se trazer à tona o embate entre a tecnologia e o instituto da responsabilidade civil, como se verá mais adiante.

Apesar dos esforços de se definir a inteligência artificial na sua completude, é importante ressaltar que essa área de estudo comporta várias subcategorias, cada uma com a sua particularidade. Considerando que essas peculiaridades impactam, inclusive, em como o direito deve abordar o tema e, por consequência, nas conclusões deste estudo, passa-se a expor algumas das espécies abrangidas pelo gênero inteligência artificial.

1.3 TIPOS DE INTELIGÊNCIA ARTIFICIAL

Assim como há várias definições para inteligência artificial, existem também várias formas de a categorizar. Neste texto, serão abordadas duas dimensões da classificação da inteligência artificial, que facilitarão na localização do problema que esta pesquisa busca endereçar. Essas dimensões são: (i) a baseada nas suas capacidades e estágio de evolução; e (ii) a relacionada aos ramos e aos domínios de aplicação e desenvolvimento da tecnologia.

Em relação à primeira dimensão, divide-se a inteligência artificial em fraca (no inglês, *narrow* ou *weak*), forte ou geral (*Artificial General Intelligence – AGI*) e em superinteligência (*Artificial Superintelligence – ASI*).

Atualmente, tem-se apenas sistemas fracos de inteligência artificial, ou seja, aqueles que são programados para executarem uma tarefa específica e que, portanto, não conseguem ser eficientes na resolução de outros problemas, diferentes daqueles para os quais ela foi programada inicialmente para resolver. A Alexa pode ser usada como um bom exemplo de inteligência artificial fraca. A assistente de voz da Amazon é capaz de tocar uma música de acordo com a solicitação do seu usuário, por outro lado, caso se deseje saber, por exemplo, qual é o melhor trajeto de casa ao trabalho, a Alexa não saberá a resposta. Pode-se dizer também, exemplificativamente, que o “AlphaGo, que aprendeu

a jogar Go analisando milhares de partidas e depois jogando contra si mesmo, não teria a capacidade de aprender a jogar xadrez”⁶⁶.

Pensando-se então na inteligência artificial fraca, que é a existente atualmente, pode-se dividi-la, ainda, em dois subgrupos: a IA reativa e a IA de memória limitada. A inteligência artificial reativa é o tipo mais simples da tecnologia, correspondente às máquinas que operam exclusivamente baseadas nos dados presentes, sem a capacidade de formar inferências dos dados atuais para avaliar possíveis situações futuras. O sistema Deep Blue da IBM, que venceu uma partida de xadrez contra Garry Kasparov, ex-campeão mundial de xadrez, é um exemplo de inteligência artificial reativa. Ou seja, o sistema não aprendeu novas jogadas de xadrez, ele apenas reproduziu aquelas que já lhe haviam sido apresentadas por meio dos dados que alimentaram o sistema. Esses algoritmos reagem da mesma forma sempre que lhes é apresentado uma situação já conhecida. Conclui-se, assim, que esse tipo de inteligência artificial é previsível, respondendo de uma forma esperada aos seus *inputs*, logo, não representam problemas em termos de responsabilidade civil.

Ao contrário da IA reativa, a IA de memória limitada é capaz de buscar informações e dados do passado para fazer inferências sobre o futuro, aprimorando, assim, as suas respostas aos diferentes estímulos que recebe no decorrer do tempo. Esses sistemas são movidos a algoritmos de aprendizado que lhes possibilitam ler o ambiente ao seu entorno e gerar novas formas de resolução de problemas a partir da sua experiência. Os carros autônomos podem ser um exemplo desse tipo de inteligência artificial. Ao trafegar, os carros são deparados com uma série de novas situações que exigem condutas imediatas, que são feitas considerando as experiências passadas, a partir dos dados que o software armazena e analisa, para tomar, cada vez, melhores decisões no trânsito.

Já a inteligência artificial forte, ainda inexistente, seria aquela capaz de se adaptar e encontrar soluções em tarefas distintas, em diferentes ramos do conhecimento. É a também chamada inteligência artificial geral, que corresponderia ao funcionamento da inteligência humana, capaz de perceber, associar, prever e planejar. Lembra-se que vários sistemas fracos acoplados não geram a inteligência forte, uma vez que, para se alcançar esta, seriam necessárias interações entre os vários tipos de conhecimentos que a máquina possui, fazendo com que ela gerasse correlações e inferências entre eles e generalizações

⁶⁶ LOUKIDES, Mike; LORICA, Ben. *What's Artificial Intelligence?* Sebastopol: O'Reilly Media, 2016, p. 2.

de aprendizados passados, da mesma forma em que o cérebro humano apreende e relaciona experiências diversas. Comparando a inteligência artificial fraca e a forte, Meredith Broussard afirma que a inteligência artificial forte seria próxima àquela criada em filmes de Hollywood, enquanto a fraca seria apenas “statistics on steroids”⁶⁷.

A superinteligência, por sua vez, corresponde ao estágio de desenvolvimento da inteligência artificial que superaria a inteligência humana. Segundo Nick Bostrom, a superinteligência seria “qualquer intelecto que exceda em muito o desempenho cognitivo de seres humanos em, virtualmente, todos os domínios de interesse”⁶⁸. Ainda de acordo com o mesmo autor, a superinteligência poderia ser aferida de algumas formas. Pode haver, por exemplo, a superinteligência em razão da maior agilidade no raciocínio e na ação⁶⁹, a partir da total emulação do cérebro humano em um hardware extremamente veloz⁷⁰; a superinteligência que qualitativamente seja melhor do que a humana, ainda que seja apenas tão ágil quanto esta; e, ainda, a superinteligência coletiva, definida como “um sistema composto de um grande número de intelectos menores, de forma que seu desempenho total supere significativamente, em diversas áreas gerais do conhecimento, qualquer sistema cognitivo atual”⁷¹.

O sistema superinteligente seria capaz de se adaptar às emoções, da mesma forma que os humanos são capazes de o fazer, bem como de entender as necessidades do seu interlocutor humano, por meio do discernimento do seu processo de pensamento. Somado a isso, espera-se também da superinteligência não só a capacidade de apreender tudo ao seu redor, mas também de ser autoconsciente, de reconhecer a si mesma, de ter crenças e desejos, assim como o humano. Apesar de haver previsões de alguns estudiosos, como Nick Bostrom e Raymond Kurzweil⁷², sobre a viabilidade da superinteligência em um

⁶⁷ BROUSSARD, Meredith. *Artificial Intelligence: How computers misunderstand the world*. Cambridge: MIT Press, 2018, p. 32.

⁶⁸ BOSTROM, Nick. *Superinteligência: caminhos, perigos e estratégias para um novo mundo*. Rio de Janeiro: DarkSide, 2014, p. 55.

⁶⁹ BOSTROM, Nick. *Superinteligência: caminhos, perigos e estratégias para um novo mundo*. Rio de Janeiro: DarkSide, 2014, p. 108.

⁷⁰ BOSTROM, Nick. *Superinteligência: caminhos, perigos e estratégias para um novo mundo*. Rio de Janeiro: DarkSide, 2014, p. 109.

⁷¹ BOSTROM, Nick. *Superinteligência: caminhos, perigos e estratégias para um novo mundo*. Rio de Janeiro: DarkSide, 2014, p. 111.

⁷² Raymond Kurzweil é um premiado cientista e futurologista que escreveu o livro *The Singularity Is Near*, publicado em 2005, no qual ele afirma que a Singularidade – considerada como o fenômeno em que máquina mais inteligentes que os humanos poderiam se replicar em máquinas cada vez mais complexas, o que ocorreria *ad infinitum* – seria possível já nas próximas décadas. Para maiores informações sobre a Singularidade vide: KURZWEIL, Ray. *The Singularity is Near: When humans transcend biology*. Nova York: Penguin Books, 2005.

futuro próximo, o ponto de vista adotado neste trabalho é que o estágio de desenvolvimento atual da inteligência artificial está distante de alcançar algo que se pareça com a superinteligência, pelo que esta não será objeto de estudo por ora.

Passando-se para a análise da inteligência artificial sob a perspectiva dos seus ramos de desenvolvimento, tem-se as seguintes principais divisões: (i) aprendizado de máquina e aprendizado profundo; (ii) sistemas especialistas; (iii) processamento de linguagem natural – PLN; e (iv) robótica.

Em relação ao aprendizado de máquina (*machine learning*), pode-se conceituá-lo como a habilidade de sistemas de IA de aprenderem com a própria experiência, possibilitando condutas originais por meio da realização do estudo e inferências de dados correspondentes a situações passadas. Conforme explica Evan Joseph Zimmerman, isso não significa que o computador programa a si mesmo, mas que a máquina é capaz de gerar e armazenar associações e fatos independente da sua programação explícita⁷³. De forma objetiva, Stuart Russel define esse mecanismo como “o ramo da IA que estuda formas de fazer com que os computadores melhorem sua performance com base na experiência”⁷⁴.

Apesar de ter tido origem em pesquisas na área da ciência cognitiva e da psicologia sobre a aprendizagem conceitual e de reforço, o *machine learning*, no seu atual grau de desenvolvimento, distancia-se consideravelmente de teorias sobre a mente humana, sendo carregado em estudos estatísticos e de probabilidade. Nesse sentido, Margaret Boden afirma:

Esse campo [*machine learning*] [...] depende atualmente de técnicas matemáticas assustadoras, porque as representações do conhecimento utilizadas envolvem a teoria da probabilidade e estatística (pode-se dizer que a psicologia ficou muito para trás). É claro que alguns sistemas modernos de aprendizado de máquina têm pouca ou nenhuma semelhança com aquilo que provavelmente acontece na cabeça do ser humano⁷⁵.

Essa visão vai ao encontro da perspectiva da inteligência artificial que se busca evidenciar neste trabalho, qual seja, a perspectiva científica e matemática da inteligência

⁷³ “This does not mean that the machine programs itself per se, but rather that the machine is able to generate and store associations and facts, even if it is outside the confines of its programming”. (ZIMMERMAN, Evan J. *Machine Minds: Frontiers in Legal Personhood*, 2015. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2563965. Acesso em: 08 de maio de 2021, p. 20.)

⁷⁴ RUSSEL, Stuart. Q&A: The Future of Artificial Intelligence. University of Berkeley. 2016. Disponível em: <https://people.eecs.berkeley.edu/~russell/research/future/q-and-a.html>. Acesso em: 05 out. 2021.

⁷⁵ BODEN, Margaret. *Inteligência Artificial: uma brevíssima introdução*. Tradução Fernando Santos. São Paulo: Editora Unesp, 2020, p. 68.

artificial, despojada da carga fantasiosa que a acompanha desde o início do seu desenvolvimento. Entende-se que, a partir dessa visão, as discussões sobre a interpretação jurídica e a regulação da inteligência artificial seriam mais assertivas.

Para se esclarecer como os algoritmos são capazes de aprender, é importante mencionar as formas como eles podem ser treinados. Citam-se aqui as formas mais tradicionais de treinamento que são, basicamente, por meio de aprendizado supervisionado, não supervisionado e de reforço⁷⁶.

No aprendizado supervisionado, os dados que serão apresentados ao sistema são previamente rotulados. Os rótulos são os chamados exemplos e não exemplos. Supõe-se que se deseja ensinar um sistema a identificar cachorros. Nesse caso, as imagens que serão apresentadas serão rotuladas como “cachorros” ou “não cachorros”. Assim, deve haver uma grande quantidade de dados disponíveis rotulados, para que o algoritmo consiga aprender a identificar aquele dado chamado de “exemplo”. O programador, então, dá *feedbacks* ao sistema, indicando se os resultados esperados foram ou não alcançados. O aprendizado deste tipo de sistema se dá pela elaboração de hipóteses, por meio de estatística e probabilística, das características que são relevantes para se identificar o objeto desejado. Assim, “o sistema observa alguns exemplos de pares *input-output* e aprende a função que o guia do *input* para o *output* desejado”⁷⁷. O grande desafio em relação à aprendizagem supervisionada é que a maioria dos dados disponíveis na rede não está rotulada e que, ao mesmo tempo, rotular quantidades massivas de dados é uma tarefa dispendiosa em termos de tempo e de pessoas dedicados a essa atividade.

No aprendizado não supervisionado, por sua vez, os *inputs* não são rotulados, ou seja, não são fornecidos os resultados desejados ou os *feedbacks*. Nesse caso é o próprio algoritmo que identifica padrões entre os dados disponíveis, sendo usado para “descobrir conhecimento”, uma vez que “os programadores não precisam saber quais padrões ou agrupamentos existem nos dados: o sistema descobre sozinho”, conforme afirma Margaret Boden⁷⁸. A principal tecnologia relacionada ao descobrimento de padrões é o

⁷⁶ Lembra-se que há uma infinidade de algoritmos de *machine learning* e de fundamentos para a sua criação (classificador Naive Bayes, K-Nearest Neighbor, Regressão Linear) que, na sua maioria, podem ser enquadrados nas categorias de supervisionado, não supervisionado e por reforço.

⁷⁷ RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010, p. 695.

⁷⁸ BODEN, Margaret. *Inteligência Artificial: uma brevíssima introdução*. Tradução Fernando Santos. São Paulo: Editora Unesp, 2020, p. 69

aprendizado profundo (*deep learning*), subespécie do *machine learning*, que utiliza redes neurais para processar dados e encontrar padrões⁷⁹⁻⁸⁰.

A última forma de aprendizado de máquina que será tratada neste trabalho é o aprendizado por reforço (*reinforcement learning*), em que um sistema é colocado em um ambiente complexo e deve aprender a se comportar por meio de “análogos de recompensa e punição”⁸¹, ou seja, a partir de *feedbacks* que indicam se a ação tomada foi positiva ou negativa. O aprendizado por reforço, que também pode se valer de técnicas de *deep learning*, significou um salto de qualidade, principalmente, para o ramo de jogos e para a robótica. Quanto aos jogos, Russel e Norvig afirmam:

For example, in game playing, it is very hard for a human to provide accurate and consistent evaluations of large numbers of positions, which would be needed to train an evaluation function directly from examples. Instead, the program can be told when it has won or lost, and it can use this information to learn an evaluation function that gives reasonably accurate estimates of the probability of winning from any given position⁸².

Pode-se dizer o mesmo em relação à robótica: tarefas que são executadas por humanos de forma simples, como transitar em um espaço sem se chocar com objetos a sua volta, significam um grande desafio no treinamento de robôs. O método de reforço é útil também nesse tipo de inteligência artificial, uma vez que catalogar todo tipo de situação que um robô pode encontrar ao se locomover em um determinado ambiente poderia impossibilitar o seu desenvolvimento.

⁷⁹ “A tecnologia *deep learning* é uma subárea do *machine learning*. Esse tipo de sistema permite o processamento de enormes quantidades de dados para encontrar relacionamentos e padrão que os seres humanos são muitas vezes incapazes de detectar. A palavra “*deep*” (em português, “profundo”) refere-se ao número de camadas ocultas na rede neural, as quais fornecem grande parte do poder de aprendizagem. [...] No nível básico, uma rede neural artificial (Artificial Neural Network – ANN) é uma função que inclui unidades (que também podem ser chamadas de neurônios, perceptrons ou nós). Cada unidade terá um valor e um peso, que indicam a sua importância relativa, e irá para a camada oculta. A camada oculta usa uma função cujo resultado se torna a saída”. TAULLI, Tom. Introdução à Inteligência Artificial. Tradução A Press Media, LLC. São Paulo: Novatec, 2020, p. 98-99.

⁸⁰ Para melhor visualização do funcionamento do *deep learning*, vide: MOORE, Samuel; SCHNEIDER, David; STRICKLAND, Eliza. How Deep Learning Works: Inside the neural networks that power today's AI. *IEEE Spectrum*, 2021. Disponível em: <https://spectrum.ieee.org/what-is-deep-learning>. Acesso em: 27 jun. 2022.

⁸¹ BODEN, Margaret. *Inteligência Artificial: uma brevíssima introdução*. Tradução Fernando Santos. São Paulo: Editora Unesp, 2020, p. 69.

⁸² Fornecer avaliações precisas e consistentes de um grande número de estratégias e jogadas é uma tarefa complexa para os humanos. Em vez disso, o programa pode ser informado quando ganhou ou perdeu e pode usar essas informações para aprender uma função de avaliação que forneça estimativas razoavelmente precisas da probabilidade de ganhar em qualquer posição. (RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010, p. 831, tradução nossa.)

Nota-se que o ramo de *machine learning* se confunde com a categorização da inteligência artificial de memória limitada, pois, apenas a partir de algoritmos de aprendizagem, a máquina é capaz de se desenvolver ao longo do tempo, em vez de apenas repetir padrões anteriores, decorrentes da programação. É exatamente este o ramo da inteligência artificial que se pretende estudar neste trabalho, tendo em vista que, em primeiro lugar, o *machine learning* é uma das categorias de IA mais difundidas e promissoras, visto o seu sucesso em conquistar progressos inéditos em complexidade e habilidades dos sistemas de inteligência artificial. Conforme mencionam Mike Loukides e Ben Lorica:

Over the past few years, we've seen many applications that qualify as AI, in one sense or another. Almost anything that falls under the rubric of "machine learning" qualifies as artificial intelligence: indeed, "machine learning" was the name given to the more successful parts of AI back when the discipline fell into disrepute⁸³.

Além disso, o desenvolvimento do *machine learning* significou uma mudança de paradigma da computação. Passou-se de softwares que se "resumiam a uma lista de comandos que seguia um fluxo lógico de execução"⁸⁴ para um outro tipo de sistema cuja programação deixa espaços abertos para o aprendizado da máquina. Ou seja, os programadores deixam de ser apenas os instrutores de tudo aquilo que os softwares são capazes de fazer, para serem receptores de informações novas descobertas pelo software.

E, por fim, o fato que é mais relevante para esta pesquisa, pensando-se na interface da tecnologia com a responsabilidade civil, o *machine learning* é o tipo de IA que impõe desafios ao referido instituto legal, uma vez que os sistemas dotados de aprendizado de máquina podem apresentar conduta inesperada quando analisada a sua programação original, gerando dificuldade em se aferir o nexo de causalidade entre a programação e o eventual dano causado, bem como dificuldade na identificação do agente responsável pelo dano. O embate entre a inteligência artificial e a responsabilidade civil será melhor explicitado no segundo capítulo deste trabalho.

⁸³ Nos últimos anos, vimos muitos aplicativos que se qualificam como IA, em um sentido ou outro. Quase tudo que se enquadra na rubrica de "aprendizado de máquina" é qualificado como inteligência artificial: na verdade, "aprendizado de máquina" era o nome dado às partes mais bem-sucedidas da IA quando a disciplina caiu em descrédito. (LOUKIDES, Mike; LORICA, Ben. *What's Artificial Intelligence?* Sebastopol: O'Reilly Media, 2016, p. 8, tradução nossa.)

⁸⁴ TAULLI, Tom. *Introdução à Inteligência Artificial*. Tradução A Press Media, LLC. São Paulo: Novatec, 2020, p. 65

Apenas para que o leitor possa também ter uma breve compreensão a respeito dos demais ramos da inteligência artificial, sintetiza-se aqui algumas das principais características do processamento de linguagem natural e da robótica, sendo importante salientar que ambos esses ramos podem fazer uso de algoritmos de aprendizagem. Em razão disso, não se descarta a menção e o estudo de robôs e sistemas de PLN neste trabalho, o que também justifica as breves explicações sobre esses ramos na presente pesquisa. Quanto aos sistemas especialistas, visto que eles já foram abordados no tópico anterior e que atualmente são muito pouco utilizados, eles não serão aprofundados neste capítulo.

A robótica é o ramo da inteligência artificial que desenvolve os robôs, que, por sua vez, podem ser conceituados como “agentes físicos capazes de performar tarefas a partir da manipulação do mundo físico”⁸⁵, ou, ainda, de acordo com Neil Richards e William Smart:

A robot is a constructed system that displays both physical and mental agency but is not alive in the biological sense. That is to say, a robot is something manufactured that moves about the world, seems to make rational decisions about what to do, and is a machine⁸⁶.

Além de alguns robôs se valerem de aprendizado de máquina para se desenvolverem, o que os atrai também para o objeto de estudo deste trabalho, para fins de responsabilidade civil, é menos importante fazer a diferenciação entre a inteligência artificial que opera exclusivamente por meio de aplicações em *hardwares* tradicionais, tal como um aparelho de celular ou um computador pessoal, e aqueles que se corporificam em “hospedeiros” não tão usuais, como são os carros autônomos e os *UVA's* (*Unmanned aerial vehicle*), considerando que, em ambas as situações, o uso do sistema de inteligência artificial pode gerar danos ao seu usuário ou a terceiros.

O Processamento de Linguagem Natural mereceu uma categoria específica da inteligência artificial por comportar vários processos que são muito próprios ao desafio de se compreender, traduzir e reproduzir a linguagem natural. Dentre esses desafios, citam-se, por exemplo, (i) a possibilidade de palavras e expressões serem ambíguas; (i) o

⁸⁵ RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010, p. 971

⁸⁶ Um robô é um sistema que tem ação tanto mental quanto no mundo físico. Isso significa dizer que o robô é algo fabricado que se move geograficamente e parece tomar decisões racionais sobre o que fazer e, além disso, é uma máquina. (RICHARDS, Neil M.; SMART, Willian D. How should the law think about robots? In: CALO, Ryan; FROOMKIN, A. Michael; KERR, Ian. *Robot Law*. Northhampton: Edward Elgar Pub, 2016, p. 6, tradução nossa).

fato de haver pronúncias muito parecidas para palavras semanticamente desconexas; (iii) os diferentes sotaques e dialetos existentes; e (iv) os erros gramaticais que se cometem principalmente quando do uso da linguagem falada.

Apesar de todas essas adversidades, o PLN tem se desenvolvido rapidamente, o que se pode perceber com a melhora dos retornos do Google Tradutor, a disseminação de assistentes de voz, como Alexa e Cortona, e também dos *chatbots* empregados para fins comerciais, para atendimento ao cliente, por exemplo. Isso se deve ao emprego do *deep learning* nesse ramo, mas também a uma sistematização do processo de apreensão da linguagem natural, que passa por várias etapas com o objetivo de decomposição das frases para sua posterior recomposição de forma apreensível pelo sistema. Assim, toda frase que é apresentada a um sistema de PLN passa por um processo de identificação de sentenças e depois de palavras, de exclusão de palavras que tem um menor grau de contribuição para a compreensão da frase (como os artigos), de redução da palavra à sua raiz (como por exemplo, “agendamento” e “agendado” tem como raiz a palavra “agenda”), até que cada frase seja minuciada em um formato em que os computadores são capazes de “ler”⁸⁷. Apesar de esse processo poder se assemelhar às etapas de aprendizado da linguagem nos humanos, segundo Margaret Boden:

[A] PLN atual depende mais de músculos (capacidade computacional) do que cérebro (análise gramatical). A matemática – especificamente a estatística – superou a lógica, e a tradução automática [...] substituiu a análise sintática. [...] O motor de busca do Google, por exemplo, busca termos avaliados por relevância – que é calculada estatisticamente, não semanticamente (isto é, sem compreensão)⁸⁸.

Identificado o que é inteligência artificial, para o presente trabalho, e o tipo de inteligência artificial que será estudado, é importante apresentar o estágio de desenvolvimento e as aplicabilidades atuais da inteligência artificial, bem como o que se pode dela esperar no futuro, o que será feito no próximo tópico.

⁸⁷Cada fase desse processo tem a sua denominação correspondente na programação de PLN. Por exemplo, a *tokenização* é o ato de segmentação da sentença em vários elementos, como substantivos, adjetivos e verbos. A *estematização* refere-se à redução da palavra à sua raiz, excluindo-se os prefixos e sufixos, enquanto à *lematização* aproxima às palavras pela semântica (por exemplo, “bom” e “melhor” são incluídas no mesmo “pacote” de palavras divididas por lemas).

⁸⁸BODEN, Margaret. *Inteligência Artificial: uma brevíssima introdução*. Tradução Fernando Santos. São Paulo: Editora Unesp, 2020, p. 88-91.

1.4 ESTADO DA ARTE: PRINCIPAIS APLICAÇÕES DA INTELIGÊNCIA ARTIFICIAL ATUALMENTE

Neste tópico, ao se expor o presente estágio de desenvolvimento da inteligência artificial e de suas aplicabilidades, pretende-se, além de elucidar a sua importância para a sociedade atual, aprofundar na diferenciação entre a inteligência artificial das obras ficcionais e o que já se tem atualmente em funcionamento e uso. Conforme afirma Meredith Broussard:

To most people, the phrase artificial intelligence suggests something cinematic – maybe Commander Data, the lifelike cyborg from *Star Trek: The Next Generation*; perhaps Hal 9000 from *2001: A Space Odyssey*; or Samantha, the AI system from the movie *Her*, or Jarvis, the AI majordomo that helps Iron Man in the Marvel comics and movies. Regardless, here’s what’s important to remember: those are imaginary. [...] A disproportionate number of the people who make tech fall into the camp of desperately wanting Hollywood robots to be real. When Facebook’s Mark Zuckerberg built an AI-based home automation system, he named it Jarvis⁸⁹.

Para se evitar distorções a respeito do objeto a ser estudado nesta pesquisa, é importante dizer que, para além de robôs humanoides, a inteligência artificial está presente em tarefas corriqueiras do dia a dia, atuando, muitas vezes, de forma imperceptível, mas extremamente impactante. E é exatamente pelo fato de já estar tão propagada, e repercutindo de várias formas na vida de milhões de pessoas (inclusive gerando danos), é que a inteligência artificial merece ser objeto deste estudo.

Como a ênfase desta pesquisa será nos algoritmos de aprendizado, serão expostos aqui alguns exemplos de aplicação desta tecnologia. Em primeiro lugar, citam-se as tarefas que foram referidas acima como “corriqueiras”, ou seja, que estão muito presentes nas rotinas das pessoas e que, de tão normalizadas, pode-se até se esquecer que, por trás delas, está operando um algoritmo de *machine learning*. São elas, por exemplo, os resultados das buscas feitas no Google; a tradução simultânea realizada por algumas plataformas, ou mesmo pelo próprio Google Tradutor; a seleção automática do feed de

⁸⁹ Para a maioria das pessoas, a expressão inteligência artificial sugere algo cinematográfico – talvez o Commander Data, o ciborgue de *Star Trek: A nova geração*; talvez Hal 9000 de *2001: Uma Odisseia no Espaço*; ou Samantha, o sistema de IA do filme *Her*, ou Jarvis, o mordomo de IA que ajuda o Homem de Ferro nos quadrinhos e filmes da Marvel. Independentemente disso, aqui está o que é importante lembrar: esses são imaginários. [...] Um número desproporcional de pessoas que estudam tecnologia se volta à tentativa desesperada de fazer os robôs de Hollywood virarem realidade. Quando Mark Zuckerberg construiu um sistema de automação residencial baseado em IA, ele o chamou de Jarvis. (BROUSSARD, Meredith. *Artificial Intelligence: How computers misunderstand the world*. Cambridge: MIT Press, 2018. p. 31, tradução nossa.)

notícias, de fotos, ou de tuítes que aparecerão aos usuários de redes sociais; os filtros de *spam* no e-mail; e a recomendação de filmes em plataformas de *streaming*. O *machine learning* também é utilizado para estimular a compra de produtos online, aproximando o potencial comprador do produto que mais provavelmente o agrada, considerando seu perfil traçado pelo algoritmo. *E-commerces* de roupas e calçados, passagens aéreas, hospedagem, cursos, entre outros, são exemplos de bens que muitas vezes são ofertados por meio de plataformas que fazem uso de *machine learning* para melhorar a percepção dos gostos de cada usuário. O mesmo ocorre com a divulgação de anúncios online, que são hiperdirecionados a um usuário alvo, que tem maior probabilidade de se interessar pelo produto anunciado.

Não à toa, boa parte das empresas que mais faturam atualmente fazem grande uso do *machine learning* como ferramenta de auxílio ao aumento de usuários ou de vendas. Exemplificativamente, Pedro Domingos afirma que:

A Amazon não consegue codificar precisamente os gostos de todos os seus clientes em um programa de computador, e o Facebook não sabe como escrever um programa que selecione as melhores atualizações a serem exibidas para cada usuário. O Walmart vende milhões de produtos e tem bilhões de decisões para tomar todo dia; se seus programadores tentassem escrever um programa para tomar todas as decisões, nunca terminariam. Em vez disso, essas empresas aplicam algoritmos de aprendizado às montanhas de dados acumulados e deixam que eles adivinhem o que os clientes querem⁹⁰.

Apesar de esse tipo de emprego da tecnologia, à primeira vista, parecer não representar maiores riscos aos indivíduos que a utilizam, a verdade é que, sob um olhar mais atento, percebe-se que podem sobrevir danos dos usos mais comezinhos do aprendizado de máquina. Cita-se, a título de exemplo, o direcionamento de anúncios de produtos mais caros, ou mesmo a indisponibilidade de alguns produtos, de acordo com o perfil de algum usuário, ou, de maneira mais geral, a própria criação dos filtros-bolha⁹¹ – que aprisionam o indivíduo em um ambiente virtual em que todas as opiniões são convergentes com a sua, sendo-lhes furtada a chance de conviver com outras possibilidades de pensamento, produtos, serviços e debates, sem que ele próprio se dê conta de tal limitação. Os danos decorrentes desse efeito colateral do uso do *machine*

⁹⁰ DOMINGOS, Pedro. *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Nova York: Basic Books, 2015, p. 35.

⁹¹ Segundo Eli Pariser, que cunhou o termo, filtros-bolha são “um universo único de informação para cada um de nós”. (PARISER, Eli. *The filter bubble: What the Internet is hiding from you*. Nova York: Penguin, 2011, p. 9.)

learning em aplicações online revela-se um grande desafio, uma vez que, em razão da dificuldade de se aferir e dimensionar o dano, estes acabam sendo não indenizáveis.

Pensando-se em outras aplicações do *machine learning* que têm o potencial de gerar ainda maiores impactos na vida das pessoas, são mencionados os usos referentes a aplicativos de relacionamento, em que algoritmos fazem todo o trabalho de busca pelo par perfeito, e também o uso da tecnologia para a procura de emprego e nos processos de recrutamento e seleção, hoje já amplamente utilizados⁹².

Mais polêmicos ainda e com maior potencialidade danosa são os sistemas de avaliação de risco de reincidência de um réu em um processo criminal, a exemplo do *COMPAS – Correctional Offender Management Profiling for Alternative Sanctions*, software que já foi utilizado pelas cortes americanas de Wisconsin, Flórida e Nova York, para auxílio na tomada de decisão em relação à condenação de cada réu. O uso desse sistema foi investigado e duramente criticado, haja vista a comprovação de que o *COMPAS* reproduzia vieses raciais, o que refletia na maior probabilidade de um acusado pertencente a uma minoria étnica ter penas mais severas do que réus brancos⁹³.

Além dos usos de *machine learning*, que são comuns à boa parte da sociedade, como os serviços online, existem também aqueles que exercem atividades específicas em cada ramo do conhecimento. Citam-se os *High Frequency Trading – HFT*⁹⁴ no mercado de capitais, os algoritmos capazes de realizarem análise de imagens e diagnóstico médicos

⁹² Cita-se, por exemplo, o polêmico caso da Amazon, que utilizava inteligência artificial para o recrutamento de colaboradores, que posteriormente se mostrou tendenciosa a apenas recrutar homens, discriminando as mulheres que participavam dos processos seletivos. Disponível em: <https://exame.com/negocios/robo-criado-pela-amazon-para-buscar-candidatos-discriminava-mulheres/>. Acesso em: 05 out. 2021.

⁹³ Para mais informações sobre o caso COMPAS, vide: <https://irisbh.com.br/por-que-a-ciencia-precisa-falar-mais-sobre-etica-uma-estudo-do-caso-compas-e-analise-da-existencia-de-machine-bias-em-decisoes-automatizadas/>. Acesso em: 05 out. 2021.

⁹⁴ Conforme afirma Isac Silveira da Costa sobre os HTFs: “[...] além do já referido elevado número de mensagens intradiárias aos sistemas de negociação, do uso de algoritmos e de infraestrutura tecnológica de minimização de latência na comunicação com os sistemas de negociação, os estudos dos reguladores norte-americanos caracterizam a atividade de HFT como sendo predominantemente de *day trading*, com a abertura e fechamento de posições no mesmo dia e baixo saldo residual ao final do pregão e, ainda, resultando em elevada taxa de cancelamento de ofertas e curta duração das posições, procurando captar pequenos lucros (mas com alta possibilidade de sucesso) em um grande número de transações.” (COSTA, Isac Silveira da. High frequency trading (HFT) é uma forma de insider trading? In: PARENTONI, Leonardo (Coord.). *Direito, Tecnologia e Inovação*: volume 1. Belo Horizonte: D’Plácido, 2018, p. 533.)

na área da saúde⁹⁵ e o suporte da tecnologia na elaboração de decisões judiciais, no ramo jurídico⁹⁶.

Passando-se, por fim, aos exemplos de emprego do *machine learning* na robótica, tem-se, por exemplo, os carros autônomos, que, apesar de não estarem atualmente difundidos em escala comercial, já são alvo de grande debate acerca da responsabilidade civil pelos danos causados por acidentes os envolvendo, tendo, inclusive, já ocorrido, em março de 2018, um acidente fatal por atropelamento durante um teste do carro autônomo da Uber⁹⁷, no estado do Arizona, nos Estados Unidos da América.

Diante de todos esses exemplos, corrobora-se aqui o entendimento de Margaret Boden sobre a inteligência artificial:

Pode-se dizer que ela tem sido extraordinariamente bem-sucedida, pois seu raio de ação prático também é extraordinariamente amplo. A IA representa uma grande quantidade de aplicações, para realizar inúmeras tarefas específicas e utilizadas em praticamente todos os setores da vida, tanto por leigos como por profissionais. Muitas dessas aplicações apresentam um desempenho melhor do que a maioria dos especialistas humanos. Nesse sentido, o progresso tem sido espetacular⁹⁸.

A ideia de se expor neste capítulo alguns dos possíveis usos do *machine learning* se justifica uma vez que se pesquisará adiante como a responsabilidade civil poderia lidar com os danos ocorridos em razão do uso da tecnologia, questionando-se, inclusive, se seria possível determinar um único regime de responsabilidade civil para a inteligência artificial de forma genérica. Assim, faz-se necessário, desde já, expor as diversas facetas e utilidades da tecnologia em análise e mercados em que esta pode estar inserida, pois

⁹⁵ Para uma análise aprofundada sobre o uso de inteligência artificial na área da saúde, vide: PRINCE II, William Nicholson. *Artificial Intelligence in Health Care: Applications and Legal Issues*. *U of Michigan Public Law Research Paper*, Michigan, n. 599, nov. 2017. Disponível em: <https://ssrn.com/abstract=3078704>. Acesso em: 05 out. 2021.

⁹⁶ Além do já mencionado COMPAS, Kai-Fu Lee menciona a plataforma iFlyTek que “assumiu a liderança na aplicação de inteligência artificial nos tribunais, construindo ferramentas e executando um programa-piloto com base em Xangai que usa dados de casos anteriores para ajudar os juízes nas provas e sentenças. Um sistema de referência cruzada de provas usa o reconhecimento de fala e o processamento de linguagem natural para comparar todas as provas apresentadas – testemunhos, documentos e material de apoio – e procurar padrões factuais contraditórios. Ele alerta então o juiz para essas disputas, permitindo investigações e esclarecimentos adicionais por parte dos membros do tribunal”. (LEE, Kai-Fu. *AI Superpowers: China, Silicon Valley, and the new world order*. Nova York: Houghton Mifflin Harcourt, 2018, p. 142.)

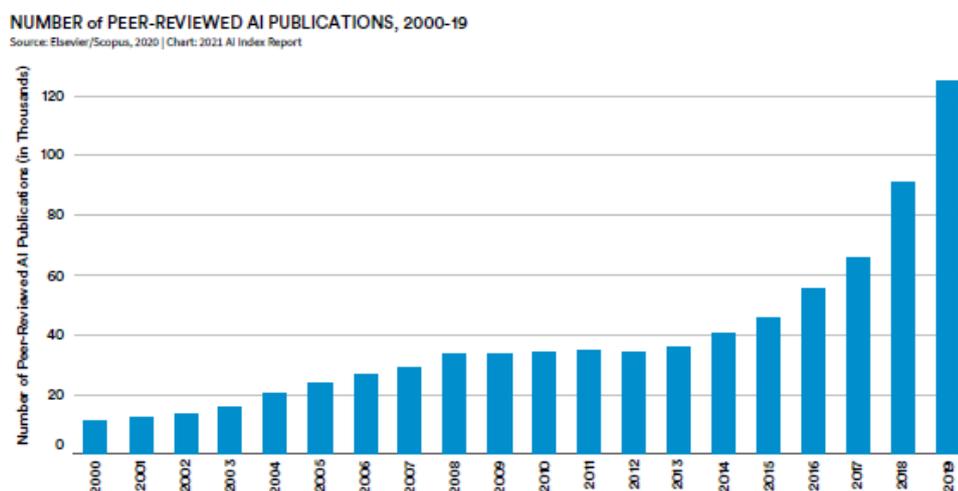
⁹⁷ Para mais informações sobre o acidente envolvendo o carro autônomo da Uber, vide: <https://www.bbc.com/news/technology-54175359>. Acesso em 05 out. 2021.

⁹⁸ BODEN, Margaret. *Inteligência Artificial: uma brevíssima introdução*. Tradução Fernando Santos. São Paulo: Editora Unesp, 2020, p. 37.

tudo isso refletirá na análise jurídica da tecnologia do ponto de vista da responsabilidade civil.

Por fim, analisando-se alguns dados dos últimos anos e algumas tendências da inteligência artificial, entende-se que existem bons motivos para se acreditar em um futuro promissor da tecnologia. Em relatório publicado em 2021 pelo *Institute for Human Centered Artificial Intelligence (HAI)* da Universidade de Stanford, são disponibilizados alguns dados que indicam o crescimento consistente da IA nos últimos anos, bem como apontam para a continuidade dessa tendência. Ao analisarem os rumos tomados pela pesquisa e desenvolvimento (P&D) em inteligência artificial – que é um bom termômetro para se avaliar a sua propensão de expansão para os próximos anos – notaram que os investimentos em P&D no ramo cresceram dramaticamente nos últimos vinte anos, impulsionados principalmente pela grande aposta feita na inteligência artificial pelas potências mundiais, China e Estados Unidos⁹⁹. Dentre os dados apresentados sobre P&D, é disponibilizado um gráfico que indica o volume de publicações acadêmicas sobre inteligência artificial nas últimas duas décadas, conforme abaixo:

Gráfico 1



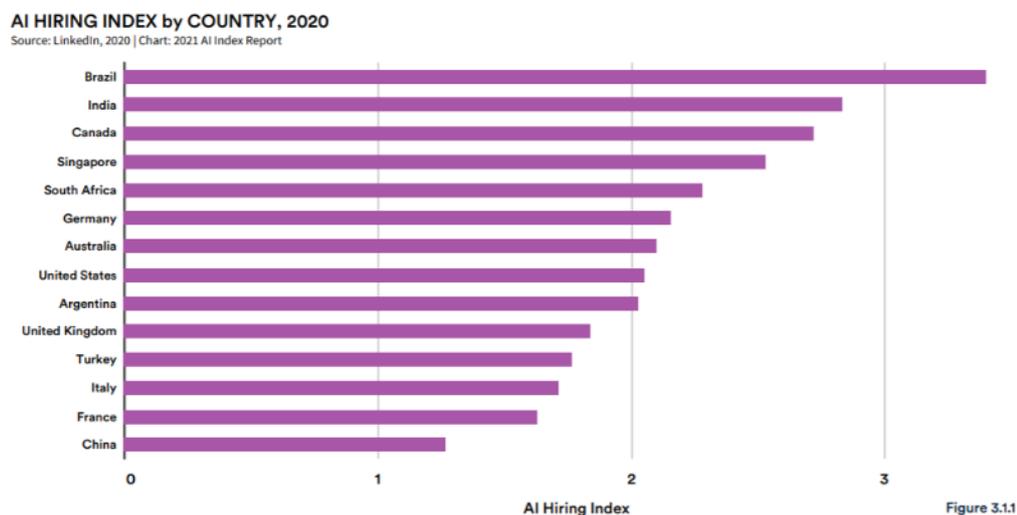
Fonte: ARTIFICIAL INTELLIGENCE INDEX REPORT 2021. Stanford University Human-Centered Artificial Intelligence, p. 16. Disponível em: https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf. Acesso em: 05 out. 2021, p. 18.

O mesmo relatório também indica a disponibilidade de postos de trabalho na área da inteligência artificial no ano de 2020, conforme dados coletados no LinkedIn. Dentre

⁹⁹ ARTIFICIAL INTELLIGENCE INDEX REPORT 2021. Stanford University Human-Centered Artificial Intelligence, p. 16. Disponível em: https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf. Acesso em: 05 out. 2021.

os países investigados, o Brasil, curiosamente, ocupa o primeiro lugar em vagas de trabalho abertas nessa área:

Gráfico 2



Fonte: ARTIFICIAL INTELLIGENCE INDEX REPORT 2021. Stanford University Human-Centered Artificial Intelligence, p. 16. Disponível em: https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf. Acesso em: 05 out. 2021, p. 84.

Analisando-se também investimentos corporativos em inteligência artificial, nota-se grande otimismo do mercado em relação à tecnologia. Fundos de *Venture Capital* especializados em inteligência artificial e *startups* na área vêm atingindo números recordes de investimento e receita, respectivamente¹⁰⁰. A aposta do mercado em IA foi potencializada em 2020 e 2021, em razão da pandemia do Covid-19. De acordo com pesquisa feita no mercado americano pela PWC, em 2021:

This trend is crystal clear. US companies are ramping up their AI investments. Fifty-two percent of our survey respondents have accelerated their AI adoption plans in the wake of the COVID-19 crisis. The results will be felt for years to come. These “accelerating” companies cite their top changes as new use cases for AI (40%) and increased AI investments (also 40%). Of all the participants

¹⁰⁰ Mais informações sobre os bons resultados de fundos e *startups* em IA, cf.: <https://www.forbes.com/sites/jeanbaptiste/2019/02/12/venture-capital-funding-for-artificial-intelligence-startups-hit-record-high-in-2018/>. Acesso em: 05 out. 2021.

in our survey, 86% say that AI will be a “mainstream technology” at their company in 2021¹⁰¹.

No mesmo sentido, o relatório anual de resultados da IBM de 2020 atribuiu parte significativa da receita da companhia no referido ano à busca dos clientes pelo emprego da inteligência artificial. É citado que as empresas estão se movimentando em grande escala em busca da adoção de inteligência artificial, sendo que 40 mil clientes recorreram à IBM, buscando ferramentas para destravar potencial valor da empresa decorrente do tratamento de dados por meio da IA¹⁰².

Pelo exposto neste tópico, conclui-se que a inteligência artificial não apenas já está amplamente difundida na sociedade atual, como há indícios de uma tendência de crescimento para os próximos anos, considerando os atuais investimentos em pesquisa e desenvolvimento, a abertura de novos postos de trabalho e o otimismo do mercado em relação ao uso da inteligência artificial para melhoria de tomada de decisões ou expansão nos negócios. Esse fato leva à necessidade de avaliação de como se deve encarar a tecnologia do ponto de vista jurídico, em relação a riscos e à mitigação ou à prevenção de danos eventualmente causados pela tecnologia. Nos próximos capítulos, explorar-se-á a interface da tecnologia com a responsabilidade civil e serão melhor analisados e esclarecidos os desafios que a IA impõe ao referido instituto.

¹⁰¹ “Essa tendência é cristalina. As empresas americanas estão aumentando seus investimentos em IA. Cinquenta e dois por cento dos entrevistados em nossa pesquisa aceleraram seus planos de adoção de IA após a crise do COVID-19. Os resultados serão sentidos nos próximos anos. Essas empresas ‘em aceleração’ citam suas principais mudanças como novos casos de uso para IA (40%) e aumento de investimentos em IA (também 40%). De todos os participantes da nossa pesquisa, 86% dizem que a IA será uma ‘tecnologia *mainstream*’ em sua empresa em 2021 (tradução nossa).” Disponível em: <https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions.html>. Acesso em: 05 out. 2021.

¹⁰² 2020 IBM ANNUAL REPORT. Disponível em: <https://www.ibm.com/annualreport/>. Acesso em: 05 out. 2021.

2 RESPONSABILIDADE CIVIL E OS DESAFIOS IMPOSTOS PELA INTELIGÊNCIA ARTIFICIAL

Tal como a ciência da computação se viu defronte de um novo paradigma da computação, com a possibilidade de o programa de computador aprender com a experiência e disponibilizar informação nova ao seu próprio programador, a responsabilidade civil também encara um novo desafio com o advento da inteligência artificial, em particular, com o aprendizado de máquina.

Como já se viu, o *machine learning* é um tipo de codificação que deixa “espaços abertos” à aprendizagem. Assim, a lógica linear da programação, em que um comando A culmina, necessariamente, em um resultado B, é deturpada. Em outras palavras, à medida que o sistema colhe dados, ele pode fazer assimilações inéditas, gerando resultados originais, ou seja, *outputs* não programados.

Essa característica não se confunde com falhas na programação, *bugs*, acessos indevidos (hackeamento), tampouco trata-se de falta de informação ao usuário sobre o funcionamento do sistema. Para estes problemas, existem respostas já consagradas no ordenamento jurídico pátrio. A questão que se impõe é a da autonomia da inteligência artificial, ou do seu comportamento emergente, como preferem alguns autores:

Como ainda não se chegou a um estado de desenvolvimento tecnológico em que as máquinas tenham, de fato, comportamentos dotados de uma autonomia total, Calo “prefere chamar essa característica de *comportamento emergente*, de modo a prestigiar o fato de que o elemento principal no debate não é a *autonomia*, mas o fato de que o robô passa a se comportar de modo distinto a partir do input que recebe do ambiente”. Argumenta Calo que o termo *autonomia* “sugere que robôs estão, de alguma forma, tomando uma decisão para agir de uma maneira específica. Pouco se ganha, e provavelmente muito é perdido, ao se imaginar que os robôs contemporâneos manifestam algo como intenção/vontade”¹⁰³.

Com isso em vista, basta lançar o olhar aos requisitos para aferição da responsabilidade civil para se deparar com os desafios impostos pela inteligência artificial a este instituto. Eles podem ser resumidos nas seguintes indagações: quem responde pelos danos causados por condutas emergentes de sistemas de inteligência artificial e sob qual critério de imputação responde?

¹⁰³ MEDON, Filipe. *Inteligência Artificial e Responsabilidade Civil: autonomia, riscos e solidariedade*. Editora JusPodivm: São Paulo, 2022, p. 147-148.

A dificuldade apresenta-se, resumidamente, no enfraquecimento do liame causal entre a conduta de um agente humano, qual seja o desenvolvedor do sistema ou seu fornecedor de modo geral, e o dano causado. Sem que o desenvolvedor/fornecedor do sistema tenha incorrido em qualquer ilícito, a inteligência artificial, por si só, é capaz de gerar danos ao seu usuário ou a terceiros. Por outro lado, pela falta de imputabilidade, não se pode responsabilizar o próprio sistema, mesmo quando ele deixa de ser mera ferramenta nas mãos de humanos e passa a atuar por si mesmo.

Aliás, a tentativa de resolução das questões trazidas pelos danos causados pela inteligência artificial é ainda mais desafiadora, uma vez que o próprio instituto da responsabilidade em si é carregado de dúvidas, de novas teorias e de mudanças vertiginosas no decorrer da história. Segundo Rui Stoco, a responsabilidade civil “é o instituto de Direito Civil que teve maior desenvolvimento nos últimos cem anos”¹⁰⁴. Nos dizeres deste autor, tal desenvolvimento “exige um repensar das concepções antes tidas por inabaláveis, tal como a ideia de responsabilidade sem culpa”¹⁰⁵. Diante dessa assertiva, nas próximas páginas, será feita uma breve exposição dos dois elementos tradicionais da responsabilidade civil que se correlacionam de forma intrincada com o problema da conduta emergente da inteligência artificial, quais sejam a culpa e o nexo de causalidade. Serão sucintamente avaliadas as mutações desses elementos da responsabilidade civil ao longo do tempo, correlacionando-as às novas questões trazidas pela inteligência artificial. Para tanto, valeu-se de um importante marco teórico, a saber, a teoria da erosão dos filtros da responsabilidade civil de Anderson Schreiber.

Antes, contudo, é importante mencionar que, apesar de a responsabilidade contratual e extracontratual terem os seus elementos centrais comuns e sobrepostos¹⁰⁶, haverá leve inclinação deste trabalho para o estudo da responsabilidade extracontratual, na qual se localizam as principais questões envolvendo os danos causados pela inteligência artificial atualmente. Mais adiante, também será tratada a responsabilidade

¹⁰⁴ STOCO, Rui. *Tratado de responsabilidade civil: doutrina e jurisprudência*. 7. ed. São Paulo: Revista dos Tribunais, 2007, p. 148.

¹⁰⁵ STOCO, Rui. *Tratado de responsabilidade civil: doutrina e jurisprudência*. 7. ed. São Paulo: Revista dos Tribunais, 2007, p. 148.

¹⁰⁶ “Todos os casos de responsabilidade civil obedecem a quatro séries de exigências comuns: a) o dano, que deve ser certo, podendo, entretanto, ser material ou moral; b) e a relação de causalidade, a causal connexion, laço ou relação direta de causa e efeito entre o fato gerador da responsabilidade e o dano são seus pressupostos indispensáveis; c) a força maior e a exclusiva culpa da vítima têm, sobre ação de responsabilidade civil, precisamente porque suprimem esse laço de causa e efeito, o mesmo efeito preclusivo; d) as autorizações judiciais e administrativas não constituem motivo de exoneração de responsabilidade”. (DIAS, José de Aguiar. *Da responsabilidade civil*. 11. ed. Atualizada de acordo com o Código Civil de 2002 e aumentada por Rui Belford Dias. Rio de Janeiro: Renovar, 2006, p. 131.)

sob a ótica do direito do consumidor, visto que esta também é uma seara de grande controvérsia, em razão do uso dos produtos e serviços imbuídos de inteligência artificial pelo consumidor final.

2.1 A CULPA

Ao se dissertar sobre o elemento culpa, o primeiro desafio no qual se esbarra é a o da sua própria conceituação. “Não obstante a simplicidade do seu sentido vulgar”¹⁰⁷, trata-se de conceito de difícil abstração. Aguiar Dias, lembrando das lições de Georges Ripert, afirma que:

É tão difícil definir culpa, que muitos mestres negam a própria possibilidade de o fazer. Ripert é, a este respeito, muito expressivo, pois sustenta, sem disfarce, que não há definição legal da culpa, que não se pode, mesmo, tentar formulá-la¹⁰⁸.

A despeito dessa afirmação de Ripert, muitos doutrinadores envidaram grandes esforços para tentar definir a culpa. Na Modernidade, o seu conceito era carregado de conotação moral, sendo preciso acessar os dados internos do indivíduo para sua identificação. Essa acepção concedia à culpa uma justificativa filosófica que ia ao encontro dos preceitos individualistas da época e da preponderância da autonomia da vontade. Ao mesmo tempo, o elemento anímico da culpa também satisfazia as diretrizes do Direito Canônico, marcado pelas noções de pecado e livre-arbítrio, que ecoaram na Idade Moderna¹⁰⁹. Tal é a importância desse elemento moral e psíquico que, também na contemporaneidade, alguns autores sustentam que para haver culpa, seria imprescindível perquirir o estado de consciência do causador do dano. Conforme explica Schreiber:

De fato, ao definir a noção jurídica de culpa, muitos autores se valiam – e ainda hoje se valem – de elementos psicológicos ou anímicos, típicos de uma avaliação moral e subjetiva da conduta individual. Assim, refere-se Chironi à culpa como um “*stato d’animo*” do agente que se pretende responsável. Salemi refere-se expressamente à relação entre a psiche e o ato lesivo, como característica da culpa. Stoppato alude à culpa como um “*diffetto della intelligenza*”. Pontes de Miranda trata da culpa como “*falta de devida atenção*”. José de Aguiar Dias refere-se à previsibilidade do resultado como elemento

¹⁰⁷ BRIZ, Jaime Santos. *La Responsabilidad Civil*, p. 42 apud PEREIRA, Caio Mário da Silva. *Responsabilidade Civil*. 9. ed. Rio de Janeiro: Forense, 1999, p. 64.

¹⁰⁸ DIAS, José de Aguiar. *Da responsabilidade civil*. 11. ed. Atualizada de acordo com o Código Civil de 2002 e aumentada por Rui Belford Dias. Rio de Janeiro: Renovar, 2006, p. 136-137.

¹⁰⁹ SCHREIBER, Anderson. *Novos paradigmas da Responsabilidade Civil: da erosão dos filtros da reparação à diluição dos danos*. 6. ed. VitalSource Bookshelf, Grupo GEN, 2015.

indispensável ao conceito. Karl Larenz menciona, além da previsibilidade, um “includível” juízo de valor do agente sobre seu próprio comportamento¹¹⁰.

Até mesmo as formas de exteriorização da culpa¹¹¹ adotadas pelo art. 186 do Código Civil brasileiro¹¹²⁻¹¹³, consubstanciadas na dupla formulação da imprudência e da negligência, parece reproduzir, em certa medida, a necessidade de inteligência da moralidade de uma determinada conduta para identificação da culpa¹¹⁴.

Essa definição passou por um longo processo de depuração, sendo-lhe identificada a dificuldade do juízo de reprovabilidade da conduta por um exame de consciência do indivíduo, dependente de um dado psicológico, inalcançável pelos julgadores. Dentre os críticos dessa definição moral de culpa, destacam-se os irmãos Mazeaud, que enfatizaram, então, o elemento externo, de fácil aferição ao intérprete, para preconizar a sua própria definição de culpa *stricto sensu*¹¹⁵ como “o erro de conduta que não teria sido cometido

¹¹⁰ SCHREIBER, Anderson. *Novos paradigmas da Responsabilidade Civil: da erosão dos filtros da reparação à diluição dos danos*. 6. ed. VitalSource Bookshelf, Grupo GEN, 2015, p. 14.

¹¹¹ “A falta de cautela exterioriza-se através da imprudência, da negligência e da imperícia. Não são, como se vê, espécies de culpa, nem elementos desta, mas formas de exteriorização da conduta culposa”. (FILHO, Sérgio Cavalieri. *Programa de Responsabilidade Civil*. 9. ed. São Paulo: Atlas, 2010, p. 37.)

¹¹² “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. *Lei n. 10.406*, de 10 de janeiro de 2002. Institui o Código Civil. Artigo 927, parágrafo único. Brasília, 2002. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/2002/110406.htm. Acesso em: 23 maio 2022.)

¹¹³ Nota-se que o art. 186 do Código Civil ao em vez de mencionar a culpa, fala apenas em ato ilícito, o substrato da culpa. Quanto a isso: “Em lugar de *faute*, fonte de interminável controvérsia, [...] o nosso legislador preferiu adotar a noção de ato ilícito. [...] O que se percebe desde logo é que o legislador brasileiro desprezou a distinção entre delitos e quase-delitos, no que fez bem pelo menos se tiver em atenção o faro de que, em outros países, constitui essa questão pretexto de interminável polêmica. O nosso legislador foi ainda ao extremo da prudência ao deixar de definir a culpa”. (DIAS, José de Aguiar. *Da responsabilidade civil*. 11. ed. Atualizada de acordo com o Código Civil de 2002 e aumentada por Rui Belford Dias. Rio de Janeiro: Renovar, 2006, p. 151.)

¹¹⁴ “A própria fórmula tríplice da negligência, imprudência e imperícia, incorporada a tantas codificações e apresentada frequentemente como substitutivo à definição de culpa, revela, claramente, o caráter moral e psicológico de que se costuma revestir o conceito”. (SCHREIBER, Anderson. *Novos paradigmas da Responsabilidade Civil: da erosão dos filtros da reparação à diluição dos danos*. 6. ed. VitalSource Bookshelf, Grupo GEN, 2015, p. 15.)

¹¹⁵ Apesar de se fazer referência apenas à definição de culpa em sentido estrito, sabe-se que o dolo passou pelo mesmo processo de despojamento do elemento psíquico, apesar do seu inerente caráter volitivo. Segundo Caio Mário: “Modernamente, o conceito de dolo alargou-se convergindo a doutrina no sentido de caracterizá-lo na conduta antijurídica, sem que o agente tenha o propósito de prejudicar. Abandonando a noção tradicional do *animus nocendi* (ânimo de prejudicar), aceitou que a sua tipificação delimita-se no procedimento danoso, com a consciência do resultado. Para a caracterização do dolo não há mister perquirir se o agente teve o propósito de causar o mal. Basta verificar se ele procedeu consciente de que seu comportamento poderia ser lesivo. Se a prova da intenção implica a pesquisa da vontade de causar o prejuízo, o que normalmente é difícil de se conseguir, a verificação da consciência do resultado pode ser averiguada na determinação de elementos externos que envolvam a conduta do agente”. (PEREIRA, Caio Mário da Silva. *Responsabilidade Civil*. 9. ed. Rio de Janeiro: Forense, 1999, p. 66.)

por uma pessoa avisada, colocada nas mesmas circunstâncias externas do autor do dano”¹¹⁶.

Com efeito, a culpa passou a ser observada sobre padrões relativamente mais objetivos do que o de detecção do estado de consciência do autor do dano, baseando-se não só nas circunstâncias externas, mas também na consagrada ideia de homem médio, ou do *bonus pater familias*. A partir daí, passou-se à chamada culpa normativa, pela qual “sendo juridicamente capaz, qualquer indivíduo passa a ter como padrão de comportamento a diligência do homem médio, seja ele maximamente ou minimamente apto a empregar tal diligência sob o ponto de vista fático”¹¹⁷, segundo Schreiber.

Essa ideia, no entanto, tampouco escapou de severas críticas. Quem seria essa figura abstrata de homem razoável imune a qualquer possibilidade de erro de conduta? Sucedeu-se, assim, um desconforto nesse padrão, na medida em que a sua identificação seria tão ou mais subjetiva quanto à própria aferição de consciência do autor do dano, considerando que ficaria a cargo do próprio magistrado, que, muito provavelmente, teria como parâmetro de homem razoável a sua personalidade própria¹¹⁸.

Questiona-se como identificar esse requisito unificador do *bonus pater familias* diante da complexidade das relações interpessoais e mercadológicas na contemporaneidade, em um contexto de alta desigualdade social e de especialização das atividades econômicas cada vez maior. Essa concretização da definição de homem razoável pelo magistrado passou a ser tida como intolerável, levando a mais uma guinada na noção de culpa e dos critérios para perquirição dos erros de conduta¹¹⁹.

Para se verificar a existência da culpa, passou-se a adotar, então, um requisito ainda mais objetivo que o do homem avisado, ou prudente, identificado nas próprias regras de mercado, em análise periciais, em códigos de conduta, no uso do artifício do

¹¹⁶ DIAS, José de Aguiar. *Da responsabilidade civil*. 11. ed. Atualizada de acordo com o Código Civil de 2002 e aumentada por Rui Belford Dias. Rio de Janeiro: Renovar, 2006, p. 146

¹¹⁷ SCHREIBER, Anderson. *Novos paradigmas da Responsabilidade Civil: da erosão dos filtros da reparação à diluição dos danos*. 6. ed. VitalSource Bookshelf, Grupo GEN, 2015, p. 37.

¹¹⁸ SCHREIBER, Anderson. *Novos paradigmas da Responsabilidade Civil: da erosão dos filtros da reparação à diluição dos danos*. 6. ed. VitalSource Bookshelf, Grupo GEN, 2015, p. 40.

¹¹⁹ “Não apenas as desigualdades sociais, como também a crescente complexidade da vida contemporânea, a especialização dos setores econômicos e o avanço desconcertante das novas tecnologias resultam em que, muitas vezes, o juiz se vê diante de situações às quais não se pode transportar”. (SCHREIBER, Anderson. *Novos paradigmas da Responsabilidade Civil: da erosão dos filtros da reparação à diluição dos danos*. 6. ed. VitalSource Bookshelf, Grupo GEN, 2015, p. 41.)

amicus curiae, entre outros¹²⁰. Dessa forma, o julgador sai da sua subjetividade e encara os problemas sob uma perspectiva externa, que melhor se aproxima e enxerga o cerne do problema. Conforme aduz Maria Celina Bodin:

Através da nova concepção existirão tantos modelos de diligência quanto forem os tipos de conduta (profissional, desportiva, na direção de veículos etc.) presentes no contato humano, de modo que os parâmetros, entre os tipos, serão variáveis (e diz-se que foram “subjetivados” ou relativizados). Isto é o que permite que se estabeleçam padrões – standards – de conduta que exigirão do agente um comportamento judicioso, o qual variará em cada situação, consideradas sua profissão e demais circunstâncias pessoais¹²¹.

Adentrando no contexto da inteligência artificial, a aceção normativa da culpa teria facilitado a análise da reprovabilidade da conduta do desenvolvedor do software, não fosse pela capacidade de autoaprendizado do sistema. Sendo assim, não se descarta a possibilidade de análise de conduta do desenvolvedor/fornecedor com base em *standards* de mercado para avaliar se ele agiu com culpa ou não. Para tanto, seria avaliado se este se cercou dos artifícios disponíveis e que, razoavelmente, sob a perspectiva do mercado, poderia se esperar que ele tivesse empregado para o desenvolvimento de um tal sistema de inteligência artificial. Ocorre que a própria formatação desse tipo de codificação tem como valor intrínseco a possibilidade de o sistema aprender com a própria experiência, enturvecendo o que se pode atribuir à conduta do agente humano por trás da inteligência artificial e o que advém do próprio sistema. Segundo Mafalda Miranda Barbosa:

Ora, é exatamente a centralidade do conceito de culpa que faz com que os sistemas delituais se mostrem insuficientes para lidar com danos causados por entes dotados de inteligência artificial. As características de autonomia e autoaprendizagem de tais entes dificultam o traçar de fronteira entre os danos

¹²⁰ Nesse sentido, Schreiber afirma: “Na construção de tais modelos, as cortes não se têm baseado tão somente na consciência judicial, mas se socorrido, saudavelmente, de parâmetros externos. Mesmo fora do âmbito daqueles casos que tradicionalmente exigem perícia (como os relacionados a erro médico), os magistrados têm buscado recursos na sociedade para a formação dos standards de conduta, valendo-se, por exemplo, de diretrizes emitidas por associações profissionais, de códigos de conduta especializados mesmo desprovidos de valor normativo, da oitiva de assistentes judiciais especializados. Destaca-se, no Brasil, por exemplo, a atuação da Comissão de Valores Mobiliários como *amicus curiae*, cada vez mais solicitada pelas cortes nacionais, ‘porquanto as questões de direito societário, de competência da CVM, regem-se por regras muito específicas, costumeiramente regulamentadas por portarias, instruções, resoluções, pareceres de orientação que, por vezes, escapam do alcance do já deveras atarefado Poder Judiciário’”. (SCHREIBER, Anderson. *Novos paradigmas da Responsabilidade Civil: da erosão dos filtros da reparação à diluição dos danos*. 6. ed. VitalSource Bookshelf, Grupo GEN, 2015, p. 42.)

¹²¹ MORAES, Maria Celina Bodin de. *Danos à pessoa humana: uma leitura civil-constitucional dos danos morais*. Rio de Janeiro: Renovar, 2003, p. 213.

que resultam de um erro humano e aqueles que são devidos ao próprio algoritmo¹²².

Traz-se nesse ponto um outro elemento de importância central quando se estuda a culpa relacionada aos danos gerados pela inteligência artificial, qual seja, a previsibilidade. Conforme Gustavo Tepedino e Rodrigo Guia, “a tão propagada imprevisibilidade dos atos praticados por sistemas de inteligência pode ser associada, à luz da dogmática da responsabilidade civil, à controvérsia histórica sobre a necessidade de previsibilidade do dano indenizável”¹²³.

No Brasil, sabe-se que não há posituação da previsibilidade do dano como um pressuposto da culpa, diferentemente do que ocorre em países como Itália e França, que “aludem à previsibilidade do dano indenizável”¹²⁴.

No Direito norte-americano, a previsibilidade também tem importante papel na aferição da culpa¹²⁵. No caso da inteligência artificial, Curtis Karnow, ao afirmar que o *Restatement of Torts* não é suficiente para lidar com os desafios do *machine learning*, embasa toda a sua argumentação no fato de tanto a responsabilidade objetiva quanto a subjetiva estarem condicionadas à ideia de previsibilidade do dano:

The basic theories of tort law predicate liability on foreseeability, by which I do not mean foreseeability in the general sense, but rather a type of predictable harm to a predictable group of potential victims. This is true whether one looks at negligence or strict liability¹²⁶.

¹²² BARBOSA, Mafalda Miranda. O futuro da responsabilidade civil desafiada pela inteligência artificial: as dificuldades dos modelos tradicionais e caminhos de solução. *Revista de Direito da Responsabilidade*, p. 284, 2020. Disponível em: <https://revistadireitoresponsabilidade.pt/2020/o-futuro-da-responsabilidade-civil-desafiada-pela-inteligencia-artificial-as-dificuldades-dos-modelos-tradicionais-e-caminhos-de-solucao-mafalda-miranda-barbosa/>. Acesso em: 23 maio 2022.

¹²³ SILVA, Rodrigo da Guia; TEPEDINO, Gustavo. Inteligência Artificial e elementos da responsabilidade civil. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Thomsom Reuters Brasil, 2019, p. 306.

¹²⁴ SILVA, Rodrigo da Guia; TEPEDINO, Gustavo. Inteligência Artificial e elementos da responsabilidade civil. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Thomsom Reuters Brasil, 2019, p. 307.

¹²⁵ “No *common law*, onde a previsibilidade (*foreseeability*) tem peso reconhecido na avaliação da negligência, já se destacou inúmeras vezes a incoerência que surge no sistema de responsabilização com a adoção da *thin skull rule*, uma vez que as particulares condições da vítima são, em regra, imprevisíveis para o agente lesivo”. (SCHREIBER, Anderson. *Novos Paradigmas da Responsabilidade Civil: da erosão dos filtros da reparação à diluição dos danos*. 6. ed. VitalSource Bookshelf, Grupo GEN, 2015, p. 73.)

¹²⁶ As teorias básicas do direito de responsabilidade civil baseiam a responsabilidade na previsibilidade, com o que não me refiro à previsibilidade no sentido geral, mas sim a um tipo específico de dano previsível a um grupo também previsível de potenciais vítimas. Isto é verdade quer se considere a responsabilidade subjetiva ou a responsabilidade objetiva [tradução nossa] (KARNOW, Curtis E. A. The application of traditional tort theory to embodied machine intelligence. In: CALO, Ryan; FROMKIN, A. Michael; KERR, Ian. *Robot Law*. Northampton: Edward Elgar Pub, 2016, p. 72).

Apesar da falta de normatização da questão no Direito brasileiro, retornando às diferentes conceituações da culpa, é preciso dizer que sob a sua compreensão moral divisa-se com tranquilidade a previsibilidade como elemento capital do conceito.

Para Aguiar Dias, sob a perspectiva subjetiva da culpa, esta é tida como a “falta de diligência na observância da norma de conduta, isto é, o desprezo, por parte do agente, do esforço necessário para observá-la, com resultado, não objetivado, mas previsível, desde que o agente se detivesse na consideração das consequências eventuais da sua atitude”¹²⁷.

Em apertada síntese sobre a noção moral de culpa, Sílvio Rodrigues também elenca a previsibilidade como requisito tradicionalmente apontado pelos defensores dessa vertente:

O primeiro grupo [de autores que definem a culpa] é constituído por definições que, de certo modo, inspiram-se numa concepção moral de culpabilidade. O ato danoso deve ser imputado a seu autor. Assim, mister se faz não só que haja ele violado uma regra de conduta, mas que, agindo dentro de seu livre-arbítrio, tenha o agente tido a possibilidade de prever, de agir diferentemente, impedindo, se lhe aprouvesse, o evento danoso¹²⁸.

Embora se pareça que, na culpa normativa, esse elemento deva ser prescindido, pois vinculado a um dado interno do indivíduo, é possível encontrar referências à previsibilidade também dentre os adeptos dessa acepção da culpa.

Cavaliere Filho, logo após concluir que a noção de culpa deve ser normativa, “exigindo um juízo de valor em cada caso”¹²⁹, ele a conceitua como “conduta voluntária contrária ao dever de cuidado imposto pelo Direito, com a produção de um evento danoso involuntário, porém previsto ou previsível”¹³⁰. Ele ainda enfatiza que “não havendo previsibilidade, estaremos fora dos limites da culpa, já no terreno do caso fortuito ou força maior”¹³¹.

Ainda sobre a interação entre a noção de culpa normativa e a crucialidade da previsibilidade do dano, Caio Mário, ao citar Jaime Santos Briz, afirma que:

Ao conceituar a culpa como desvio de um modelo ideal de conduta: “modelo representando, umas vezes pela *fides* ou *bona fides*, e outras pela diligência de um *pater familias* cuidadoso, a determinação do erro de conduta”, acrescenta

¹²⁷ DIAS, José de Aguiar. *Da responsabilidade civil*. 11. ed. Atualizada de acordo com o Código Civil de 2002 e aumentada por Rui Belford Dias. Rio de Janeiro: Renovar, 2006, p. 149

¹²⁸ RODRIGUES, Sílvio. *Direito Civil: Responsabilidade Civil*. São Paulo: Saraiva, 2002, p. 145.

¹²⁹ FILHO, Sérgio Cavalieri. *Programa de Responsabilidade Civil*. 9. ed. São Paulo: Atlas, 2010, p. 34.

¹³⁰ FILHO, Sérgio Cavalieri. *Programa de Responsabilidade Civil*. 9. ed. São Paulo: Atlas, 2010, p. 35.

¹³¹ FILHO, Sérgio Cavalieri. *Programa de Responsabilidade Civil*. 9. ed. São Paulo: Atlas, 2010, p. 35.

ele, “é a previsibilidade, sem a qual um bom pai de família não o teria cometido”¹³².

Assim, caso considerada a previsibilidade como um dado informativo da conduta culposa, torna-se ainda mais problemático o reconhecimento da responsabilidade do desenvolvedor/fornecedor do sistema de inteligência artificial, considerando que a imprevisibilidade é ônus indissociável dos sistemas de *machine learning*.

Voltando-se ao desenvolvimento histórico do conceito e aplicação da culpa, embora a sua noção normativa trouxesse mais clareza para a sociedade em relação aos requisitos para identificação do sujeito culpado, fato é que o próprio instituto da responsabilidade subjetiva não foi tido como suficiente para cobrir todas as pretensões ressarcitórias advindas da vida moderna, principalmente em razão dos acidentes de trabalho, que começavam a se avultar no capitalismo industrial. Nas famosas palavras de De Page, na dependência da culpa, “a irresponsabilidade é a regra; a responsabilidade, a exceção”¹³³.

Assim, o processo de aprimoramento do conceito de culpa conviveu com a própria decadência desse elemento da responsabilidade civil, tratando-se de conceito “progressivamente desacreditado”¹³⁴, conforme afirma Aguiar Dias. Em contrapartida a esse descrédito, observou-se a objetivação da responsabilidade civil, inaugurada pela teoria do risco.

Nas palavras de Jossierand, que, ao lado de Salleiles, lançou a ideia de teoria do risco:

Impor à vítima ou a seus herdeiros demonstrações desse gênero [de culpa] é o mesmo que lhes recusar qualquer indenização: um direito só é efetivo quando sua prática está assegurada; não ter direito e tê-lo sem o poder exercer são uma coisa só. A teoria tradicional de responsabilidade repousa manifestamente em bases muito estreitas: cada vez mais se mostrava insuficiente e perempta¹³⁵.

No Brasil, Aguiar Dias, um dos arautos da responsabilidade objetiva, afirma que:

Temos prova no fato de ser precisamente à equidade, ao sentimento de solidariedade social, à revolta em face das desigualdades de fortuna, influido

¹³² BRIZ, Jaime Santos Briz. *La Responsabilidad Civil*, p. 43 apud PEREIRA, Caio Mário da Silva. *Responsabilidade Civil*. 9. ed. Rio de Janeiro: Forense, 1999, p. 68, p. 45.

¹³³ FILHO, Sérgio Cavalieri. *Programa de Responsabilidade Civil*. 9. ed. São Paulo: Atlas, 2010, p. 29.

¹³⁴ DIAS, José de Aguiar. *Da responsabilidade civil*. 11. ed. Atualizada de acordo com o Código Civil de 2002 e aumentada por Rui Belford Dias. Rio de Janeiro: Renovar, 2006, p. 149.

¹³⁵ DIAS, José de Aguiar. *Da responsabilidade civil*. 11. ed. Atualizada de acordo com o Código Civil de 2002 e aumentada por Rui Belford Dias. Rio de Janeiro: Renovar, 2006, p. 79.

nos direitos reclamados pelas partes, as principais invocações com que se procura justificar a teoria do risco¹³⁶.

Nota-se, assim, que face ao sentimento de injustiça gerado pela necessidade de aferição da culpa e pela amplificação do caráter perigoso da vida em sociedade, a responsabilidade objetiva, apesar de não ser isenta a críticas, teve grande aceitabilidade não só na doutrina e na jurisprudência como em diversas legislações pelo mundo. Antes da positivação da cláusula geral da atividade de risco no Código Civil brasileiro, já se mencionava a adesão à essa teoria no projeto de Código Civil húngaro e nos Códigos de países como Rússia, México, Espanha e Portugal¹³⁷.

No Brasil, o parágrafo único do art. 927 do Código Civil de 2002 prevê que haverá responsabilidade independente de culpa quando “a atividade normalmente desenvolvida pelo autor do dano implicar, por sua natureza, risco para os direitos de outrem”¹³⁸.

Tal dispositivo legal é fundado na teoria do risco criado¹³⁹, que tem como notável adepto Caio Mário da Silva Pereira:

É o que se fixa no fato de que, se alguém põe em funcionamento uma qualquer atividade, responde pelos eventos danosos que esta atividade gera para os indivíduos, independentemente de determinar se em cada caso, isoladamente, o dano é devido à imprudência, à negligência, a um erro de conduta, e assim se configura a “teoria do risco criado”¹⁴⁰.

Como se vê, não à toa hipóteses, frequentemente ventiladas sobre a responsabilidade pelos danos causados pela inteligência artificial, baseiam-se na responsabilidade objetiva, seja do fornecedor, do supervisor, ou do próprio usuário do sistema. Nos próximos tópicos, essas possíveis respostas serão analisadas, mas não sem

¹³⁶ DIAS, José de Aguiar. *Da responsabilidade civil*. 11. ed. Atualizada de acordo com o Código Civil de 2002 e aumentada por Rui Belford Dias. Rio de Janeiro: Renovar, 2006, p. 72.

¹³⁷ PEREIRA, Caio Mário da Silva. *Responsabilidade Civil*. 9. ed. Rio de Janeiro: Forense, 1999, p. 21.

¹³⁸ BRASIL. *Lei n. 10.406*, de 10 de janeiro de 2002. Institui o Código Civil. Artigo 927, parágrafo único. Brasília, 2002. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/2002/110406.htm. Acesso em: 23 maio 2022.

¹³⁹ Conforme argumenta Schreiber, no sentido de o dispositivo em questão ser fundamentado na teoria do risco criado: “Diante de todo o exposto, a conclusão mais razoável parece ser a de que a cláusula geral de responsabilidade objetiva dirige-se simplesmente às atividades perigosas, ou seja, às atividades que apresentam grau de risco elevado seja porque se centram sobre bens intrinsecamente danosos (como material radioativo, explosivos, armas de fogo etc.), seja porque empregam métodos de alto potencial lesivo (como o controle de recursos hídricos, manipulação de energia nuclear etc.). Irrelevante, para a incidência do dispositivo, que a atividade de risco se organize ou não sob forma empresarial ou que se tenha revertido em proveito de qualquer espécie para o responsável”. (SCHREIBER, Anderson. *Novos Paradigmas da Responsabilidade Civil: da erosão dos filtros da reparação à diluição dos danos*. 6. ed. VitalSource Bookshelf, Grupo GEN, 2015, p. 25.)

¹⁴⁰ PEREIRA, Caio Mário da Silva. *Responsabilidade Civil*. 9. ed. Rio de Janeiro: Forense, 1999, p. 268.

antes se voltar à análise do nexo de causalidade como elemento essencial da responsabilidade civil.

2.2 O NEXO CAUSAL

Dispensada a culpa, anteriormente tida como o mais importante filtro da responsabilidade civil, esta passa a ter seu foco no elemento causal entre a conduta e o dano¹⁴¹.

Dessa forma, a única possibilidade de o réu se ver livre de determinada condenação seria demonstrar a inexistência do nexo causal entre a sua conduta e o fato danoso. Voltam-se, assim, as atenções a esse segundo filtro da reparação, cuja evolução histórica e diferentes teorias são tão complexas quanto a questão da culpa.

Nas conhecidas palavras de Caio Mário da Silva Pereira quanto ao nexo causal:

Este é o mais delicado elemento da responsabilidade civil e o mais difícil de ser determinado. Aliás, sempre que um problema jurídico vai ter na indagação ou na pesquisa da causa, desponta a sua complexidade maior. Ele compreende, ao lado do aspecto filosófico, dificuldades de ordem prática¹⁴².

Sabe-se que o nexo de causalidade, em conjunto com o ilícito e o dano, é requisito clássico para aferição da responsabilidade civil, ou seja, não basta que haja o ilícito e o dano, é preciso que haja uma relação de causalidade entre ambos, o que caberia à vítima comprovar.

Apesar de o nexo causal ter um elemento naturalístico, ou seja, decorrente da noção natural de causa e efeito (“algo assim como: se chover fica molhado”¹⁴³), deve-se ter em mente que a causalidade jurídica não se resume à causalidade natural ou factual. Isso porque o condicionamento de um determinado dano a uma conduta lesiva não implica, de *per se*, o dever de reparação. Conforme leciona Schreiber:

De fato, reconhece-se, há muito, que o nexo de causalidade natural ou lógico diferencia-se do jurídico, no sentido de que nem tudo que, no mundo dos fatos

¹⁴¹ “O fundamento será, então, este: todo problema de responsabilidade civil resolve-se num problema de causalidade. Todo fato do homem obriga aquele que causou um prejuízo a outrem a repará-lo. Para dizer com Carbonnier, a responsabilidade objetiva “não importa em nenhum julgamento de valor sobre os atos do responsável. Basta que o dano se relacione materialmente com estes atos, porque aquele que exerce uma atividade deve-lhe assumir os riscos” (PEREIRA, Caio Mário da Silva. *Responsabilidade Civil*. 9. ed. Rio de Janeiro: Forense, 1999, p. 19.)

¹⁴² PEREIRA, Caio Mário da Silva. *Responsabilidade Civil*. 9. ed. Rio de Janeiro: Forense, 1999, p. 76.

¹⁴³ FILHO, Sérgio Cavalieri. *Programa de Responsabilidade Civil*. 9. ed. São Paulo: Atlas, 2010, p. 47.

ou da razão, é considerado como causa de um evento pode assim ser considerado juridicamente¹⁴⁴.

Para se evitar uma regressão contínua e infinita das causas sucessivas que culminaram em dano indenizável é que o legislador brasileiro optou pela positivação da teoria da causa imediata, conforme se extrai do art. 403 do Código Civil:

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 processual¹⁴⁵.

Apesar de o dispositivo estar localizado sob o título que diz respeito ao direito das obrigações, e a sua redação trazer a palavra “inexecução”, é amplamente aceito, inclusive com sólido posicionamento jurisprudencial¹⁴⁶, que a teoria da causalidade imediata, ou da interrupção do nexu causal, como também é conhecida, se aplica igualmente à responsabilidade extracontratual.

Essa teoria elege como causa do dano apenas o evento diretamente ligado a ele, evitando, assim, uma sucessividade incansável de acontecimentos que teriam levado ao dano. Várias escolas desenvolveram suas próprias teses a respeito da causa direta e imediata, no intuito de melhor definir o que se entende por esses adjetivos, sendo uma das mais prestigiadas a subteoria da necessariedade causal, segundo a qual a expressão “direto e imediato” passa a ser interpretada não como um requisito de aproximação temporal, mas sim de necessariedade para a causa do evento danoso. Assim, afirma Schreiber:

Desta forma, podem-se identificar danos indiretos, passíveis de ressarcimento, desde que sejam consequência necessária da conduta tomada como causa. De fato, a melhor doutrina conclui, atualmente, que a necessariedade consiste no verdadeiro núcleo da teoria da causalidade direta e imediata, não se excluindo a ressarcibilidade excepcional de danos indiretos, quando derivados necessariamente da causa em questão¹⁴⁷.

Nas palavras de Agostinho Alvim, referência dessa escola:

¹⁴⁴ SCHREIBER, Anderson. *Novos Paradigmas da Responsabilidade Civil: da erosão dos filtros da reparação à diluição dos danos*. 6. ed. VitalSource Bookshelf, Grupo GEN, 2015, p. 55.

¹⁴⁵ BRASIL. *Lei n. 10.406*, de 10 de janeiro de 2002. Institui o Código Civil. Artigo 403. Brasília, 2002. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/2002/110406.htm. Acesso em: 23 mai o 2022.

¹⁴⁶ “Diga-se de passagem que, não obstante se localiza norma no capítulo destinado ao inadimplemento das obrigações negociais, certamente se aplica também no campo da responsabilidade civil aquiliana ou extranegocial”. (FARIAS, Cristiano Chaves de; ROSENVALD, Nelson; NETTO, Felipe Peixoto Braga. *Curso de direito civil: responsabilidade civil*. 5. ed. Salvador: Juspodivm, 2018, p. 417.)

¹⁴⁷ SCHREIBER, Anderson. *Novos Paradigmas da Responsabilidade Civil: da erosão dos filtros da reparação à diluição dos danos*. 6. ed. VitalSource Bookshelf, Grupo GEN, 2015, p. 62.

Segundo o nosso ponto de vista, a Escola mais prestigiada, defendida por Giorgi, Chironi, Polacco e outros, é a da consequência necessária, à qual damos a nossa adesão. [...] Assim, é indenizável todo o dano que se filia a uma causa, ainda que remota, desde que ela lhe seja causa necessária, por não existir outra que explique o mesmo dano¹⁴⁸.

Essa teoria eliminaria, assim, possíveis concausas ao evento danoso, aceitando apenas aquele necessário, sem o qual, em última instância, o dano não teria ocorrido. Em outras palavras, a teoria considera como causa do dano aquela “necessária desse dano, porque ele a ela se filia necessariamente; é causa única, porque opera por si, dispensadas outras causas”¹⁴⁹.

Em que pese essa aparente pacificação e aceitação da teoria, esta vem sendo deturpada, principalmente, pelo que alguns sustentam ser a falta de técnica e de sistematização no emprego da teoria pelos tribunais:

A grande crítica que se faz no Brasil em relação à determinação do nexos causal é dirigida aos magistrados que, em última instância, são aos que incumbem a difícil tarefa de identificar o nexos causal e justificar a imputação dos danos e a sua limitação. A completa falta de critérios jurídicos e a actínia para a definição e pesquisa do nexos causal gerou uma confusão terminológica e uma miscelânea de decisões que ora adoram como justificativa a teoria da causalidade adequada, ora adotam a teoria do dano direto e imediato, ou até mesmo a teoria da equivalência das condições. Mas mesmo assim, ainda existem aqueles que utilizam uma das três denominações, com fundamento teórico de outra teoria¹⁵⁰.

Antes, contudo, de se pensar na incapacidade do magistrado, deve-se avaliar o porquê de a teoria da causalidade imediata parecer, então, não ser a melhor resposta, ou, pelo menos, não ser a resposta única para os pleitos de reparação civil. Conforme Agostinho Alvim:

Formica, partidário embora desta teoria, tal como a expõe Giorgi, entende que o problema do nexos causal é insolúvel à base teórica. E acrescenta: “O limite do nexos causal é, pois, uma questão de fato, para cuja solução, entretanto, todos os critérios sugeridos pela doutrina e especialmente o da necessidade podem ser verdadeiros ou ter um fundo de verdade, mas não podem ser guia seguro. Na imensa variedade dos casos práticos, singulares, sempre diferentes, rebeldes a qualquer classificação ou sotoposição a uma norma geral, a prática e as especiais condições de fato permanecem, ainda, a única orientação segura” (cf. ob. cit., n. 28). No mesmo sentido, Huc e DERNBURG, o qual descreve de um princípio regulador abstrato, e apela para a casuística, dissertando, embora,

¹⁴⁸ ALVIM, Agostinho. *Da inexecução das obrigações e suas consequências*. 5. ed. São Paulo: Saraiva, 1980, p. 346.

¹⁴⁹ ALVIM, Agostinho. *Da inexecução das obrigações e suas consequências*. 5. ed. São Paulo: Saraiva, 1980, p. 356.

¹⁵⁰ MULHOLLAND, Caitlin Sampaio. *A responsabilidade civil por presunção de causalidade*. Rio de Janeiro: GZ Editora, 2010, p. 189.

em face de outra teoria que não a do dano direto e imediato. ENNECCERUS pondera: “A difícil questão de saber até onde vai o nexo causal não se pode resolver nunca, de uma maneira plenamente satisfatória, mediante regras abstratas, mas em casos de dúvida o juiz há de resolver segundo sua livre convicção, ponderando todas as circunstâncias, segundo lhe faculta o § 287 da LPC” [LPC significa: Lei Processual Civil alemã]¹⁵¹.

E, em mais recente doutrina, Cavalieri Filho postula sob seu olhar de magistrado:

À luz do exposto é forçoso concluir que, não obstante as teorias existentes sobre o nexo causal e tudo quanto já se escreveu sobre o tema, o problema da causalidade, como ressaltamos, não encontra solução nessa fórmula simples e unitária, válida para todos os casos. Na minha experiência de magistrado tenho constatado que este é um ponto onde se registra o maior número de divergências entre os julgadores de todos os graus. E assim é porque esta ou aquela teoria fornece apenas um rumo a seguir, posto que a solução do caso concreto sempre exige do julgador alta dose de bom-senso prático e da justa relação das coisas; em suma, é imprescindível um juízo de adequação, a ser realizado com base na lógica do razoável¹⁵².

Considerando, então, que a hermética e acrítica aplicação da teoria da causalidade direta e imediata pode levar a injustiças, observa-se um amálgama de teorias aplicadas pelos julgadores, algumas vezes preterindo a teoria da causa direta e imediata frente a outras não prestigiadas pelo Código Civil, já outras vezes cotejando o art. 403 do Código Civil com diferentes teorias do nexo causal, tal como a teoria da causalidade adequada, teoria da *conditio sine qua non*, a teoria do escopo da norma, a presunção de causalidade, da causalidade alternativa, dentre tantas outras, que não seria prudente querer esmiuçá-las neste trabalho, o qual não pretende esgotar o tema do nexos de causalidade.

Todo esse desapego à teoria da causalidade direta e imediata, ou, pode-se dizer, esforço jurisprudencial para o seu cotejamento, parece demonstrar o enfraquecimento, não apenas da teoria, mas do nexos de causalidade em si, uma vez que este passa a ser analisado com tanta subjetividade e sob uma miríade de possíveis interpretações que o próprio instituto é erodido para dar lugar ao êxito de pretensões ressarcitórias, algumas das vezes, descabidas sob a leitura do art. 403 do Código Civil.

Introduzindo a discussão sobre a inteligência artificial nesse cenário já caótico, de certa forma, pela imprecisão e sobreposição de teorias sobre o nexos causal, percebe-se que, em um primeiro momento, não se pode aferir com clareza o liame causal entre a conduta do agente que desenvolve, comercializa, ou usufrui da inteligência artificial e os

¹⁵¹ ALVIM, Agostinho. *Da inexecução das obrigações e suas consequências*. 5. ed. São Paulo: Saraiva, 1980, p. 371

¹⁵² FILHO, Sérgio Cavalieri. *Programa de Responsabilidade Civil*. 9. ed. São Paulo: Atlas, 2010, p. 53.

danos causados por esta. Vê-se que a dificuldade dos casos envolvendo inteligência artificial se encontra no fato de que a atividade do sistema pode ser emergente e que este aprende com a própria experiência. A conduta do programador, por exemplo, apenas se relacionaria com o eventual dano provocado por conduta emergente da inteligência artificial pelo fato de o desenvolver ter criado o software e propiciado, assim, que ele fosse disponibilizado no mercado. Dessa forma, o evento “desenvolver o produto” não é direta ou imediatamente vinculado ao dano, ou, sequer, necessário para que ele ocorresse. Assim, baseando-se puramente na teoria do dano direto e imediato e na sua subclassificação da necessidade, a vítima restaria desamparada em uma eventual tentativa de responsabilização do desenvolvedor, fabricante ou vendedor do sistema. Nota-se que essa reflexão é válida inclusive para os casos em que se aplica o direito consumerista, pois o que se tem sob essa perspectiva é a dispensa da culpa, e não do nexo de causalidade.

A uma primeira vista e estritamente sob a ótica da teoria do nexo causal adotada pelo Código Civil, o Direito brasileiro não outorga uma resposta legislativa aceita para o caso de danos envolvendo inteligência artificial. Contudo, conforme vem sendo demonstrado, a partir das teorias alternativas e da própria erosão do filtro do nexo de causalidade, é possível que intérpretes do Direito encontrem respostas cognoscíveis ao desafio imposto pela inteligência artificial.

2.3 POSSÍVEIS RESPOSTAS AO DESAFIO POSTO

Considerando as adversidades supracitadas, estudiosos têm desenvolvido teorias para convergir a matéria da responsabilidade civil e o comportamento emergente da inteligência artificial. Neste capítulo, será feita a revisão da bibliografia sobre algumas possíveis alternativas que vêm sendo apresentadas com maior recorrência na literatura jurídica, que podem ser divididas em duas principais vertentes, quais sejam, a criação da personalidade jurídica para o sistema de inteligência artificial e a aplicação da responsabilidade objetiva sob diversos fundamentos.

2.3.1 Criação de personalidade jurídica para sistemas de inteligência artificial

Em 1992, Lawrence Solum escreveu o importante texto “Legal Personhood for Artificial Intelligences”, que traz vários questionamentos sobre a possibilidade de ser

conferida personalidade jurídica aos sistemas de inteligência artificial. O autor parte de uma abordagem comportamental e filosófica, discutindo temas que dizem respeito à possibilidade de a inteligência artificial ser provida de alma, vontade própria, consciência, entre outras características usualmente atribuídas a humanos¹⁵³.

No momento atual, também é possível se deparar com autores que pesquisam a possibilidade de concessão de personalidade jurídica à inteligência artificial sob o viés filosófico e neurocientífico, que inquire sobre o que se considerada como consciência e livre-arbítrio. Cita-se, por exemplo, Evan Zimmerman, que afirma que máquinas dotadas de alto nível de inteligência seriam merecedoras de personalidade, considerando que, no futuro, os sistemas de inteligência artificial, constituídos a exemplo do funcionamento do cérebro humano, se misturariam às pessoas na convivência social, sendo, portanto, merecedores de proteção tal qual os humanos. Porém, como o próprio autor cita em seu texto, ele se refere à superinteligência, a qual não é o foco de estudos deste trabalho¹⁵⁴.

Fato é que, em que pese haver válidas discussões sobre a atribuição de personalidade jurídica a sistemas de inteligência artificial sob a perspectiva da consciência e similitude do comportamento dessas máquinas ao humano, o assunto foi reinaugurado com maior vigor recentemente sob a perspectiva meramente patrimonialista, visando, de um lado, a indenização da vítima por danos provocados pela inteligência artificial, e de outro, a “proteção” do patrimônio do fornecedor do sistema, como uma forma de aquecer o mercado da tecnologia.

Pode-se dizer que um dos fatores contribuintes para a retomada dessas discussões foi a publicação da Resolução do Parlamento Europeu, de 16 de fevereiro de 2017, que contém recomendações à Comissão sobre disposições de Direito Civil sobre Robótica (2015/2103 (INL)). A resolução parte da premissa de que “as normas tradicionais não serão suficientes para suscitar problemas de responsabilidade jurídica pelos danos causados por um robô, uma vez que não seria possível identificar a parte responsável para

¹⁵³ SOLUM, Lawrence. Legal Personhood for Artificial Intelligences. *North Carolina Law Review*, v. 70, 1992. Disponível em: <http://scholarship.law.unc.edu/nclr/vol70/iss4/4>. Acesso em: 23 mai. 2022.

¹⁵⁴ ZIMMERMAN, Evan J. Machine Minds: Frontiers in Legal Personhood. 2015, p. 1. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2563965. Acesso em: 23 maio 2022.

prestar a indenização e para lhe exigir que reparasse os danos causados”¹⁵⁵, nos termos de um dos seus “considerandos”. A proposta contida na Resolução é, então, a de

criar um estatuto jurídico específico para os robôs a longo prazo, de modo 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 independente¹⁵⁶.

A partir dessa publicação, ganhou força o debate acerca da criação da personalidade eletrônica, ou *e-person*, para que os próprios sistemas sejam os sujeitos responsáveis pelos danos que eventualmente causarem.

Remete-se aqui aos problemas identificados nos tópicos anteriores deste trabalho quanto à fragilidade do nexos de causalidade entre o desenvolvedor/fornecedor da inteligência artificial e o dano, além da impossibilidade de imputação do próprio sistema. Uma vez que o problema da responsabilidade civil, quando se trata de agentes inteligentes, é justamente quem responde pelos danos causados, dotar os sistemas de imputabilidade para que eles próprios respondam, de fato, é uma proposta que pode ser aventada como resposta ao desafio. Convém, no entanto, avaliar se essa alternativa é viável do ponto de vista jurídico e prático.

Marcelo Milagres, com o seu olhar múltiplo e perspicaz “de alguém egresso das ciências exatas (engenharia elétrica), do profissional do direito e do [...] usuário de tantas e necessárias técnicas”¹⁵⁷, afirma que “os pressupostos e a viabilidade dessa personificação são discutíveis”¹⁵⁸. Ele afirma, em primeiro lugar, que a capacidade tecnológica não corresponde à capacidade de direito e muito menos à personificação. Esta opção, se juridicamente viável e conveniente, deveria ser tomada pelo legislador, em

¹⁵⁵ UNIÃO EUROPEIA. Parlamento Europeu. *Disposições de Direito Civil sobre robótica*. Resolução do Parlamento Europeu, de 16 de fevereiro de 2017, que contém recomendações à Comissão sobre disposições de Direito Civil sobre robótica (2015/2103(INL)). Estrasburgo, 16 fev. 2017, item AF sob o título de “Responsabilidade”. Disponível em: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_PT.html. Acesso em: 23 maio 2022.

¹⁵⁶ UNIÃO EUROPEIA. Parlamento Europeu. *Disposições de Direito Civil sobre robótica*. Resolução do Parlamento Europeu, de 16 de fevereiro de 2017, que contém recomendações à Comissão sobre disposições de Direito Civil sobre robótica (2015/2103(INL)). Estrasburgo, 16 fev. 2017, Item 59, f. Disponível em: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_PT.html. Acesso em: 23 maio 2022.

¹⁵⁷ MILAGRES, Marcelo de Oliveira. A robótica e as discussões sobre a personalidade eletrônica. In: JÚNIOR, Marcos Ehrhardt Júnior; CATALAN, Marcos; MALHEIROS, Pablo (Coord.). *Direito Civil e Tecnologia*. Belo Horizonte: Forum, 2020, p. 509.

¹⁵⁸ MILAGRES, Marcelo de Oliveira. A robótica e as discussões sobre a personalidade eletrônica. In: JÚNIOR, Marcos Ehrhardt Júnior; CATALAN, Marcos; MALHEIROS, Pablo (Coord.). *Direito Civil e Tecnologia*. Belo Horizonte: Forum, 2020, p. 513.

homenagem à segurança jurídica¹⁵⁹. Contudo, conforme afirma o autor, a personalidade eletrônica não parece ser uma alternativa possível em razão de dois principais motivos: primeiramente, não se poderia criar uma personalidade híbrida para a qual são atribuídas obrigações (de ressarcir), mas não direitos; em segundo lugar, “o reconhecimento dessa personalidade eletrônica dos robôs não ensinaria, por si só, capacidade patrimonial para suportar as consequências de suas decisões”¹⁶⁰. Milagres conclui, então, que a personificação das máquinas “parece trazer mais controvérsias que soluções”¹⁶¹, com o que concorda esta autora.

Nessa mesma toada, Eduardo Nunes de Souza questiona: “Quem deveria ser responsável pela dotação patrimonial que permitiria a criação da *e-person*? O fabricante, o desenvolvedor, alguns dos membros da cadeia de usuários, todos os usuários?”¹⁶². Além de não se ter essa resposta, a hipótese de personalidade aos entes dotados de inteligência artificial é malfadada pelo fato de que a constituição de uma personalidade na história jurídica, como foi feito, por exemplo, com as pessoas jurídicas, é no sentido de limitar a responsabilidade. Assim, a proposta vai de encontro ao seu próprio pressuposto, que é garantir o ressarcimento das pessoas lesadas por condutas da inteligência artificial¹⁶³. Soma-se a isso, conforme o próprio autor afirma, a dificuldade de se aferir um patrimônio proporcional à potencialidade danosa da tecnologia e os entraves que este fato imporá ao mercado:

Trata-se, como se percebe, de uma providência de difícil execução, a começar pelo fato de que a extensão dos possíveis danos que tais ferramentas podem causar é desconhecida. Além disso, a exigência da construção de um vultuoso patrimônio mínimo autônomo para que tais tecnologias pudessem entrar em operação criaria um claro entrave ao desenvolvimento tecnológico,

¹⁵⁹ MILAGRES, Marcelo de Oliveira. A robótica e as discussões sobre a personalidade eletrônica. In: JÚNIOR, Marcos Ehrhardt Júnior; CATALAN, Marcos; MALHEIROS, Pablo (Coord.). *Direito Civil e Tecnologia*. Belo Horizonte: Forum, 2020, p. 512.

¹⁶⁰ MILAGRES, Marcelo de Oliveira. A robótica e as discussões sobre a personalidade eletrônica. In: JÚNIOR, Marcos Ehrhardt Júnior; CATALAN, Marcos; MALHEIROS, Pablo (Coord.). *Direito Civil e Tecnologia*. Belo Horizonte: Forum, 2020, p. 517.

¹⁶¹ MILAGRES, Marcelo de Oliveira. A robótica e as discussões sobre a personalidade eletrônica. In: JÚNIOR, Marcos Ehrhardt Júnior; CATALAN, Marcos; MALHEIROS, Pablo (Coord.). *Direito Civil e Tecnologia*. Belo Horizonte: Forum, 2020, p. 515.

¹⁶² SOUZA, Eduardo Nunes de. Personalidade jurídica e inteligência artificial. In: SILVA, Rodrigo da Guia; TEPEDINO, Gustavo (Coord.). *O Direito Civil na era da inteligência artificial*. São Paulo: Thomson Reuters Brasil, 2020, p. 236.

¹⁶³ SOUZA, Eduardo Nunes de. Personalidade jurídica e inteligência artificial. In: SILVA, Rodrigo da Guia; TEPEDINO, Gustavo (Coord.). *O Direito Civil na era da inteligência artificial*. São Paulo: Thomson Reuters Brasil, 2020, p. 236.

restringindo excessivamente esse setor de mercado, que apenas poderia ser operado por agentes de notável capacidade financeira¹⁶⁴.

Por fim, Eduardo Nunes de Souza traz outro elemento ao debate que vem sendo reiterado neste trabalho: a equiparação que é feita entre humanos e mecanismos inteligentes que, segundo ele, parece decorrer não somente da atual tendência à personificação de coisas, mas pelo fato de alguns desses mecanismos assumirem aparência humana¹⁶⁵.

Assim, a proposição de uma personalidade às máquinas, apesar de se suportar na real dificuldade de identificação do responsável pelos danos causados pela inteligência artificial, parece estar embebida na noção de que os robôs são (ou serão) em alguma medida equiparáveis aos humanos, seja pelas suas capacidades, seja pela sua possível aparência antropomórfica.

Cabe aqui, inclusive, uma crítica à própria Resolução do Parlamento Europeu sobre robótica, cujas primeiras palavras remetem à ficção de Mary Shelley e de Capek:

Considerando que desde o Frankenstein de Mary Shelley ao mito clássico do Pigmaleão, passando pela história do Golem de Praga pelo robô de Karel Čapek, que cunhou o termo, as pessoas têm fantasiado acerca da possibilidade de construir máquinas inteligentes, frequentemente andróides com características humanas¹⁶⁶.

A Resolução também se vale de ficção científica para informar as suas proposições, remetendo às leis de Asimov, no sentido de que estas devem ser observadas pelos “criadores, produtores e operadores de robôs, incluindo robôs com autonomia integrada e autoaprendizagem”¹⁶⁷.

¹⁶⁴ SOUZA, Eduardo Nunes de. Personalidade jurídica e inteligência artificial. In: SILVA, Rodrigo da Guia; TEPEDINO, Gustavo (Coord.). *O Direito Civil na era da inteligência artificial*. São Paulo: Thomson Reuters Brasil, 2020, p. 213-244.

¹⁶⁵ SOUZA, Eduardo Nunes de. Personalidade jurídica e inteligência artificial. In: SILVA, Rodrigo da Guia; TEPEDINO, Gustavo (Coord.). *O Direito Civil na era da inteligência artificial*. São Paulo: Thomson Reuters Brasil, 2020, p. 213-244.

¹⁶⁶ UNIÃO EUROPEIA. Parlamento Europeu. Disposições de Direito Civil sobre robótica. Resolução do Parlamento Europeu, de 16 de fevereiro de 2017, que contém recomendações à Comissão sobre disposições de Direito Civil sobre robótica (2015/2103(INL)). Estrasburgo, 16 fev. 2017, Introdução, item A. Disponível em: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_PT.html. Acesso em: 23 maio 2022.

¹⁶⁷ UNIÃO EUROPEIA. Parlamento Europeu. Disposições de Direito Civil sobre robótica. Resolução do Parlamento Europeu, de 16 de fevereiro de 2017, que contém recomendações à Comissão sobre disposições de Direito Civil sobre robótica (2015/2103(INL)). Estrasburgo, 16 fev. 2017, Introdução, item A. Disponível em: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_PT.html. Acesso em: 23 maio 2022.

O que se conclui é que soluções dessa monta, que envolvem uma certa burocracia e a necessidade de revisão normativa, se justificam quando se imaginam robôs como os relatados na ficção, com altíssimo grau de autonomia e cuja adoção implicaria grande exposição a riscos para o seu usuário (e, ainda assim, haveria dificuldades na implementação dessa escolha jurídica). Assim, apesar de esta parecer ser uma possível solução tanto para o problema do nexos causal quanto o da imputabilidade levantados anteriormente neste trabalho, para a maioria das aplicações da tecnologia seria impensável algo como uma personalidade eletrônica. Basta, para isso, pensar nos usos corriqueiros de sistemas imbuídos de inteligência artificial capazes de lesar seus usuários.

2.3.2 Aplicação da responsabilidade objetiva

Muitos estudiosos são pela aplicação da responsabilidade objetiva aos casos de danos gerados pela inteligência artificial. Essa preferência pela teoria da responsabilidade objetiva (que converge com a tendência, já não tão nova, de desbastar o prestígio da culpa para identificação de responsabilidade) parece emergir da dificuldade em se identificar a reprovabilidade da conduta do desenvolvedor/fornecedor, considerando a autonomia do sistema, que pode ter condutas não programadas originalmente, e também da inimputabilidade do sistema. Porém, os fundamentos para a aplicação desse regime de responsabilidade se diferem entre os autores que o suscitam. No contexto da inteligência artificial, a responsabilidade objetiva, via de regra, é trazida sob três possíveis fundamentos: a atividade de risco; a responsabilidade indireta; e a responsabilidade pelo defeito do produto ou serviço.

2.3.2.1 Atividade de risco

Vale lembrar, conforme já suscitado neste trabalho, que o Brasil, diferentemente de outros países tanto do *Comom Law* quanto do *Civil Law*¹⁶⁸, possui cláusula geral de responsabilidade objetiva pela atividade de risco, insculpida no parágrafo único do art.927 do Código Civil. A cláusula geral de atividade de risco possibilita que o abismo que existe entre o fazer jurídico e os avanços tecnológicos e científicos sejam mitigados, pois a cláusula “se abre para acomodar as transformações para o futuro”¹⁶⁹. Em outras palavras, na medida em que a evolução das atividades humanas e os seus novos riscos são imponderáveis pelo legislador, a cláusula geral de atividade do risco vem em prol da manutenção de uma mínima atualização do Direito, deixando brechas para que a responsabilidade civil dê conta do influxo das novas tecnologias (em sentido *lato*) e seus riscos.

Em contraponto a isso, pondera-se que a cláusula geral fica à mercê de uma possível banalização pelos seus intérpretes, uma vez que se vivencia atualmente a chamada Sociedade do Risco, fruto da superprodução industrial, em que “o estado de exceção ameaça converter-se em normalidade”¹⁷⁰ e em que os riscos deixam de ser individuais para serem coletivos, conforme Ulrich Beck já afirmava em 1986. Considerando, então, o risco como universal e socializado, a cláusula geral de atividade de risco, se utilizada com descuido, serviria para regular qualquer tipo de situação da vida contemporânea¹⁷¹.

Incluindo a inteligência artificial nessa equação, é claro que terão empregos da tecnologia que, a rigor, trarão risco à sociedade, mas, conforme Nelson Rosenvald e Carlos Edison Monteiro, “o simples fato de a tecnologia ser nova não é justificativa suficiente para a introdução de uma disciplina de responsabilidade objetiva”¹⁷². Para esses

¹⁶⁸ “No Direito Europeu, a maioria dos países que compõe a União Europeia não possui, em seus ordenamentos jurídicos internos, cláusulas gerais de responsabilidade civil objetiva nos mesmos moldes da previsão legislativa constante no Código Civil (CC) brasileiro [...]”. (WESENDONCK, Tula. *Inteligência Artificial e Responsabilidade Civil pelos Riscos do Desenvolvimento: um estudo comparado entre as propostas de regulamentação da matéria na união Europeia e o ordenamento vigente brasileiro*. In: BARBOSA, Mafalda Miranda *et al.* (Coord.). *Direito Digital e Inteligência Artificial: diálogos entre Brasil e Europa*. São Paulo: Editora Foco, 2021, p. 205.)

¹⁶⁹ FARIAS, Cristiano Chaves de; ROSENVALD, Nelson; NETTO, Felipe Peixoto Braga. *Curso de direito civil: responsabilidade civil*. 5. ed. Salvador: Juspodivm, 2018, p. 475.

¹⁷⁰ BECK, Ulrich. *Sociedade de Risco: Rumo a uma outra modernidade*. Tradução de Sebastião Nascimento. 2. ed. São Paulo: Editora 34, 2011, p. 28.

¹⁷¹ BECK, Ulrich. *Sociedade de Risco: Rumo a uma outra modernidade*. Tradução de Sebastião Nascimento. 2. ed. São Paulo: Editora 34, 2011, p. 28.

¹⁷² FILHO, Carlos Edison do Rêgo Monteiro; ROSENVALD, Nelson. Riscos e responsabilidades na inteligência artificial e noutras tecnologias digitais emergentes. In: *O Direito Civil na era da inteligência artificial*. São Paulo: Thomson Reuters Brasil, 2020, p. 557.

autores, atualmente se justificaria a aplicação da responsabilidade objetiva pelo risco da atividade exclusivamente aos casos de robôs que se locomovem em espaços públicos¹⁷³. O que se quer dizer com isso é que não são todas as aplicações da inteligência artificial que suscitam riscos, logo, nem sempre essa teoria poderia ser aplicada.

Para Mafalda Miranda Barbosa, em análise do cenário português de responsabilidade civil e inteligência artificial:

Acresce que, se por referência a alguns casos poderá ser possível pensar na perigosidade (ou especial perigosidade, segundo o que nos parece ser a melhor interpretação do preceito) da atividade, pela sua natureza ou pelos meios empregados, nem sempre ela é desvelável quando esteja em causa um mecanismo artificialmente inteligente. Na verdade, como considerar especialmente perigosa a utilização de um *smartphone* ou de um *tablet*? [...] De facto, é possível que alguns robots envolvam um especial perigo, entendido como a probabilidade de causar danos. Mas, dependendo do mecanismo, noutros casos, tal não se afigura plausível. A adequação do critério depende do específico desenho da norma que possa vir prever uma hipótese de responsabilidade objetiva¹⁷⁴.

Na literatura estadunidense, Curtis Karnow também afirma que, embora haja possibilidade de que alguns robôs venham a ser considerados como perigosos (*ultrahazardous*), especialmente no momento da sua introdução no mercado, na maioria das vezes o uso dessa tecnologia na rotina dos indivíduos não implicará risco¹⁷⁵. Assim, o autor entende que a teoria da responsabilidade pela atividade de risco não é suficiente para responder as dúvidas quanto à responsabilização nos casos envolvendo inteligência artificial.

É preciso mencionar, entretanto, que há doutrina dissonante quanto a esse ponto, havendo aqueles que afirmam que o simples fato de a inteligência artificial ser imprevisível, em razão da possibilidade de conduta emergente, a enquadraria no conceito de produto ou atividade perigosa. Cerka *et al.* são autores que defendem esse ponto de vista, afirmando que o gestor (*manager*) dessa fonte de risco deveria ser objetivamente responsável pelos danos que o sistema de inteligência artificial causar, sob a condição do

¹⁷³ FILHO, Carlos Edison do Rêgo Monteiro; ROSENVALD, Nelson. Riscos e responsabilidades na inteligência artificial e noutras tecnologias digitais emergentes. In: *O Direito Civil na era da inteligência artificial*. São Paulo: Thomson Reuters Brasil, 2020, p. 557.

¹⁷⁴ BARBOSA, Mafalda Miranda. O futuro da responsabilidade civil desafiada pela inteligência artificial: as dificuldades dos modelos tradicionais e caminhos de solução. *Revista de Direito da Responsabilidade*, p. 292, 2020. Disponível em: <https://revistadireitoresponsabilidade.pt/2020/o-futuro-da-responsabilidade-civil-desafiada-pela-inteligencia-artificial-as-dificuldades-dos-modelos-tradicionais-e-caminhos-de-solucao-mafalda-miranda-barbosa/>. Acesso em: 23 maio 2022.

¹⁷⁵ KARNOW, Curtis E. A. The application of traditional tort theory to embodied machine intelligence. In: CALO, Ryan; FROOMKIN, A. Michael; KERR, Ian. *Robot Law*. Northampton: Edward Elgar Pub, 2016, p. 67.

deep pocket, segundo a qual alguém que estabelece uma atividade de risco, e que auferir lucros com tal, deve ser a responsável por compensar os danos causados à sociedade pelas externalidades provocadas por essa determinada atividade¹⁷⁶.

David Vladeck é outro autor que aposta na responsabilização objetiva do fornecedor do sistema de inteligência artificial sob o argumento de que este tem maior capacidade econômica para suportar os custos de indenização decorrentes do dano causado pelo robô: “After all, it is not unreasonable that the costs of inexplicable accidents be borne, at least in part, by those who benefit from riskreducing, innovative products”¹⁷⁷. Ainda sobre o regime de responsabilidade objetiva, o autor traz dois outros interessantes fatores em prol da sua aplicação: (i) esse regime pouparia ambas as partes de um processo de indenização por perdas e danos, reduzindo, então, os custos de transação; e (ii) do ponto de vista do incentivo à inovação, um sistema mais previsível de responsabilização é mais vantajoso que sistemas que geram insegurança jurídica aos *players* do mercado tecnológico¹⁷⁸.

Partindo da premissa que vem sendo reiterada neste trabalho de que a inteligência artificial são várias, não seria razoável incluir todos os seus contornos em uma única resposta de teoria do risco, considerando, inclusive, que muitas das suas atividades já estão altamente acomodadas na sociedade, sem que as pessoas se sintam colocadas em situação de perigo por elas.

Cumprido ressaltar que dessa avaliação, de quais empregos da inteligência artificial representariam risco, surge outra questão recorrentemente trazida na doutrina. Trata-se da plasticidade e do caráter transitório do que é considerado perigoso ou não. Pedro Manuel Pimenta Mendes, ao analisar as possíveis soluções no ordenamento jurídico português para os casos de danos envolvendo inteligência artificial¹⁷⁹, afirma que a

¹⁷⁶ ČERKA, Paulius; GRIGIENĖ, Jurgita; SIRBIKYTĖ, Gintarė. Liability for damages caused by Artificial Intelligence. *Computer Law & Security Review*, v. 31, n. 3, p. 376-389, 2015. Disponível em: https://is.muni.cz/el/1422/podzim2017/MV735K/um/ai/Cerka_Grigiene_Sirbikyte_Liability_for_Damages_caused_by_AI.pdf. Acesso em: 23 maio 2022.

¹⁷⁷ “Afiml, nŁo Ł desarrazoado que os custos de acidentes inexplicŁveis sejam suportados, pelo menos em parte, por quem se beneficia dos produtos que impŁem tais riscos” [traduŁŁo nossa]. (VLADDECK, David C. Machines Without Principals: Liability Rules and Artificial Intelligence. *Washington Law Review*, v. 89, n. 1, p. 146-147, 2014.)

¹⁷⁸ “Afiml, nŁo Ł desarrazoado que os custos de acidentes inexplicŁveis sejam suportados, pelo menos em parte, por quem se beneficia dos produtos que impŁem tais riscos” [traduŁŁo nossa]. (VLADDECK, David C. Machines Without Principals: Liability Rules and Artificial Intelligence. *Washington Law Review*, v. 89, n. 1, 2014, p. 146.)

¹⁷⁹ Em Portugal, o art. 493, 2, do CŁdigo Civil traz a regra de responsabilidade por culpa presumida nos casos de atividade de risco, e nŁo de responsabilidade objetiva. O autor estuda, entŁo, hipoteticamente a possibilidade de aplicaŁŁo de responsabilidade objetiva.

novidade e a consequente noção de risco trazidas por um determinado sistema de inteligência artificial podem ser provisórias. Assim, ao ser lançado um determinado novo produto de inteligência artificial, que à primeira vista parece ser perigoso, com o uso reiterado e com eventuais atualizações de software para aprimorá-lo, estes deixariam de ser considerados como atividade de risco¹⁸⁰⁻¹⁸¹.

Por fim, citam-se os autores que receiam que a aplicação da responsabilidade objetiva pela atividade de risco refrearia a inovação, ao depositar ônus excessivo ao fornecedor (*lato sensu*) da tecnologia. Como exemplo dessa corrente doutrinária, citam-se Peter Asaro, que acredita que determinados mercados não propiciarão margem de lucro suficiente para incentivar grandes companhias a aceitarem o risco do desenvolvimento da atividade¹⁸², e Patrick Hubbard que, em análise sobre aeronaves não tripuladas, afirma:

The lack of certainty concerning the application of the abnormally dangerous activity doctrine could inhibit innovation because the potential increased liability could have two effects. First, the expanded potential liability of innovators could negatively affect their decision to engage in experiments to develop, for example, a type of U.A.S. Second, the increased liability could also reduce the demand for a robotic vehicle like a U.A.S. because purchasers and users would need to worry about potential greater liability for personal injury¹⁸³.

¹⁸⁰ MENDES, Pedro Manuel Pimenta. Inteligência Artificial e Responsabilidade Civil: as possíveis “soluções” do ordenamento jurídico português. In: BARBOSA, Mafalda Miranda *et al.* (Coord.) *Direito Digital e Inteligência Artificial: diálogos entre Brasil e Europa*. São Paulo: Editora Foco, 2021, p. 227.

¹⁸¹ No mesmo sentido, Curtis Karnow afirma que: “What counts as ultrahazardous changes over time as we become accustomed to the activity, as it spreads, and presumably as it becomes safer”. / “O que se considera como atividade de risco muda conforme o tempo, à medida que as pessoas se acostumam com ela e que a atividade se torna mais corriqueira, e também à medida que se torna mais segura” [tradução nossa]. (KARNOW, Curtis E. A. The application of traditional tort theory to embodied machine intelligence. In: CALO, Ryan; FROMKIN, A. Michael; KERR, Ian. *Robot Law*. Northampton: Edward Elgar Pub, 2016, p. 67.)

¹⁸² “Or, we could pursue one of the heavy-handed liability schemes, such as strict liability, that would regulate the industry to some extent, but also limit innovation to those areas where there are sufficient profits to motivate large capital companies to enter the market and accept the risks”. (ASARO, Peter. The Liability Problem for Autonomous Artificial Agents. *AAAI Spring Symposium Series*, mar. 2016. Disponível em: <https://www.aaai.org/ocs/index.php/SSS/SSS16/paper/view/12699>. Acesso em: 23 maio 2022.)

¹⁸³ “As incertezas quanto à aplicação da doutrina da atividade de risco podem inibir a inovação, considerando que o aumento da responsabilização em potencial pode ter dois efeitos. Primeiro, o aumento da responsabilidade em potencial pode afetar negativamente a decisão do mercado de se envolver em experimentos para desenvolver, por exemplo, um tipo de aeronave autônoma não tripulada. Em segundo lugar, o aumento da responsabilidade também poderia reduzir a demanda por um veículo autônomo, porque os consumidores e usuários precisariam se preocupar com a maior potencialidade de responsabilização pessoal” [tradução nossa]. (HUBBARD, F. Patrick. Allocating the risk of physical injury from “sophisticated robots”: Efficiency, fairness, and innovation. In: CALO, Ryan; FROMKIN, A. Michael; KERR, Ian. *Robot Law*. Northampton: Edward Elgar Pub, 2016, p. 39.)

Diante dos prós e contras à aplicação da teoria do risco, não se exclui por completo a possibilidade de aplicação do art. 927, parágrafo único do Código Civil, como uma possível resposta aos novos desafios trazidos pela inteligência artificial para alguns dos seus empregos. Essa alternativa venceria o problema da identificação da culpa e também do nexo de causalidade entre a conduta do fornecedor e do dano, uma vez que o liame causal se identificaria pelo simples fato de o fornecedor disponibilizar à sociedade produto ou serviço perigoso. Ainda assim, é preciso ter em mente que esta não poderia ser uma resposta única à questão da inteligência artificial, como vem sendo repisado neste trabalho, em razão das suas diferentes aplicabilidades e níveis de risco.

2.3.2.2 Responsabilidade indireta

Dirigindo-se para análise da segunda hipótese em que poderia haver a aplicação da responsabilidade objetiva nos casos de danos causados por inteligência artificial, parte-se para análise da doutrina que defende a aplicação da responsabilidade indireta. Esta abarca a responsabilidade por fato de outrem, prevista no art. 932 do Código Civil, a responsabilidade pelo fato do animal, conforme art. 936 do Código Civil, e, por fim, a responsabilidade pelo fato da coisa, que, apesar de não ter previsão expressa no Código Civil, exceto pelo caso específico de coisas que caem de edifícios (art. 938 do CC), é bem assimilada pela jurisprudência e pela doutrina como hipótese de responsabilidade objetiva.

A origem da teoria da responsabilidade indireta é controversa, havendo aqueles que defendem que ela se sustenta sobre os pilares da teoria do risco, e outros que assentam a responsabilidade indireta como decorrência de “dever objetivo de guarda e vigilância”¹⁸⁴. Segundo Aguiar Dias, sobre a responsabilidade por fato de outrem:

O excelente Sourdart já havia ensinado coisa semelhante, ao explicar que a razão da responsabilidade por fato de outrem está em que certas pessoas incumbe o dever de velar sobre o procedimento de outras, cuja inexperiência ou malícia possa causar danos a terceiros. É lícito, pois, afirmar sob esse aspecto, que a responsabilidade por fato de outrem não representa derrogação ao princípio da personalidade da culpa, porque o responsável é legalmente considerado em culpa, pelo menos em razão da imprudência ou negligência expressa na falta de vigilância sobre o agente do dano¹⁸⁵.

¹⁸⁴ FILHO, Sérgio Cavalieri. *Programa de Responsabilidade Civil*. 9. ed. São Paulo: Atlas, 2010, p. 193.

¹⁸⁵ DIAS, José de Aguiar. *Da responsabilidade civil*. 11. ed. Atualizada de acordo com o Código Civil de 2002 e aumentada por Rui Belford Dias. Rio de Janeiro: Renovar, 2006, p. 742.

O mesmo autor tem conclusões semelhantes em relação à responsabilidade pelo fato da coisa:

Ora, não há nada tão incongruente como expressar em responsabilidade por fato da coisa a que deriva de acidentes ocorridos com veículos ou objetos em nossa propriedade ou sob nossa guarda, porque a coisa não é capaz de fatos; e todos que ocorrem, causando danos a terceiros, por intermédio de coisas, são, realmente, consequência ou falta de vigilância ou prudência¹⁸⁶.

Sob essa perspectiva é que alguns estudiosos de Direito e tecnologia sugerem a expansão do rol taxativo da responsabilidade objetiva pelo fato de outrem, e também da responsabilidade pelo fato das coisas e dos animais para albergar os casos em que a inteligência artificial causa dano. Nesse caso, o supervisor, o usuário, ou quem quer que tenha o dever de guarda sobre o sistema, deveria responder pelos danos causados por ele.

Na literatura estrangeira, Ugo Pagallo, célebre estudioso do Direito e inteligência artificial, propõe em seu livro *The Laws of Robots: Crimes, Contracts and Torts* que “this kind of liability crucially depends on the different kinds of robots with which we are dealing: a robot nanny, a robot toy, a robot chauffeur, a robot employee, and so forth”¹⁸⁷. A partir dessa assertiva, o autor, além de inclinar o leitor de sua obra a acreditar que a inteligência artificial deve ser regulada de acordo com o mercado em que se insere, faz uma proposta de aplicação da responsabilidade indireta cujo responsável, em última instância, será aquele que deveria cumprir com o dever de “guarda” da inteligência artificial.

Essa alternativa parece interessante, trazendo uma solução relativamente simples e intuitiva para os casos de danos causados por inteligência artificial, inclusive, sob o ponto de vista patrimonial. Marcelo Milagres afirma que: “Se o cenário ideal é o de que quem deu causa ao ilícito deve por ele responder, é inevitável avaliar essa capacidade patrimonial. Nesse aspecto, a reconhecida e antiga técnica da responsabilidade indireta parece, também, ser a resposta”¹⁸⁸.

¹⁸⁶ DIAS, José de Aguiar. *Da responsabilidade civil*. 11. ed. Atualizada de acordo com o Código Civil de 2002 e aumentada por Rui Belford Dias. Rio de Janeiro: Renovar, 2006, p. 578.

¹⁸⁷ Ademais, esse tipo de responsabilidade depende crucialmente dos diferentes tipos de robôs com os quais estamos lidando: robô-babá, robô-brinquedo, robô-motorista, robô-empregado, dentre outras empregabilidades [tradução nossa]. PAGALLO, Ugo. *The Laws of Robots: Crimes, Contracts, and Torts*. Heidelberg: Springer Netherlands, 2013. p. 16.

¹⁸⁸ MILAGRES, Marcelo de Oliveira. A robótica e as discussões sobre a personalidade eletrônica. In: JÚNIOR, Marcos Ehrhardt Júnior; CATALAN, Marcos; MALHEIROS, Pablo (Coord.). *Direito Civil e Tecnologia*. Belo Horizonte: Forum, 2020. p. 517.

Ao se fazer a analogia dos casos de responsabilidade pelo fato de outrem, pelo fato das coisas ou animais com o emprego de sistemas de inteligência artificial nas mais variadas formas, seja como um agente social, por exemplo, no cuidado de crianças ou idosos, seja no emprego pela indústria, o que se percebe é que, em muitos casos, haveria o deslocamento da responsabilidade do programador ou fornecedor da tecnologia para o seu usuário. Esse deslocamento se torna especialmente relevante no caso da inteligência artificial a qual se refere neste trabalho: aquela com capacidade de aprender com novas experiências e ter comportamentos emergentes. Nesse sentido, Rosenvald e Carlos Edison Monteiro afirmam que:

Essa renovada perspectiva baseada na responsabilidade civil no âmbito do direito de família é particularmente útil para entender como a responsabilidade legal dos usuários – e não apenas de projetistas e fabricantes de robôs – deve ser compreendida no campo das tecnologias digitalmente emergentes. Com a crescente capacidade de aquisição de conhecimento e habilidades de máquinas, mediante interação com cuidadores humanos, eventualmente, em uma demanda de direito de regresso, será mesmo possível a discussão de que o proprietário ou usuário de robôs estão em melhor posição para entender o que está acontecendo com a máquina (*cheapest cost avoider*), de modo a evitar seu comportamento danoso, independentemente da conduta daquele robô ser típica de atitudes razoavelmente previsíveis de robôs semelhantes¹⁸⁹.

Há, no entanto, questões a serem mais bem definidas em caso de opção pela aplicação da responsabilidade indireta no cenário de danos gerados pela inteligência artificial, especialmente para os casos análogos à responsabilidade pelo fato de outrem. Conforme aduz Cavalieri Filho, a objetivação da responsabilidade se dá apenas em relação à pessoa que tem o dever de guarda e vigília, a quem o causador do dano esteja ligado por um vínculo jurídico. A responsabilidade do autor do dano, por sua vez, é subjetiva. Com isso, se quer dizer que “o ato deve ser tal que, se praticado por alguém imputável, configuraria a violação de um dever, a culpa estaria caracterizada se o ato ilícito fosse praticado por alguém imputável”¹⁹⁰. Sendo assim, o autor conclui:

Em qualquer dessas hipóteses será preciso a prova de uma situação que, em tese, em condições normais, configure a culpa do filho menor, do pupilo, do curatelado, como também do empregado (se for caso de responsabilidade subjetiva). O ato em condições de ser considerado culposos se nas mesmas condições tivesse sido praticado por pessoa imputável, exsurge o dever de

¹⁸⁹ FILHO, Carlos Edison do Rêgo Monteiro; ROSENVALD, Nelson. Responsabilidade Civil indireta e inteligência artificial. In: BARBOSA, Mafalda Miranda *et al.* (Coord.). *Direito Digital e Inteligência Artificial*: diálogos entre Brasil e Europa. São Paulo: Editora Foco, 2021, p. 189.

¹⁹⁰ FILHO, Sérgio Cavalieri. *Programa de Responsabilidade Civil*. 9. ed. São Paulo: Atlas, 2010, p. 194.

indenizar dos pais, tutor, curador, empregador, etc., independentemente de qualquer culpa destes¹⁹¹.

Em outros termos, deve-se dispensar o elemento “imputabilidade” dos requisitos do ilícito culposo e realizar o teste; caso o agente fosse imputável, haveria culpa. Se sim, os pais, os curadores, os tutores ou os empregadores seriam responsabilizados objetivamente.

Questiona-se, então, qual seria a racionalidade a ser aplicada para aferir se o sistema de inteligência artificial incorreu em uma conduta culposa. Seria a sua comparação com o parâmetro de conduta humana nas mesmas circunstâncias? Considerar-se-ia como paradigma outros sistemas com as mesmas atribuições do causador do dano para se avaliar o padrão aceitável de mercado?

Mais uma vez recorrendo a Nelson Rosenvald e Carlos Edison Monteiro, os autores afirmam o seguinte quanto a esta questão:

Com efeito, a noção de responsabilidade indireta exige que o auxiliar tenha se comportado mal... No caso de uma máquina ou tecnologia, quais parâmetros de referência de “condutas” devem ser avaliadas? A referência para avaliar o desempenho por tecnologia autônoma deveria ser principalmente aquela determinada pelo desempenho de tecnologia comparável disponível no mercado¹⁹².

Essa é uma alternativa aventada que fugiria do problema do nexo de causalidade, pois não seria necessário comprovar o nexo entre a conduta do fornecedor e o dano causado pelo sistema autônomo. O elemento causal estaria bem definido entre a conduta do próprio sistema e o dano, respondendo o terceiro apenas por lhe ser devido o dever de guarda sobre o sistema.

Observa-se, assim, que a responsabilidade indireta é uma possível resposta aos casos de danos provados por sistemas de inteligência artificial, apesar de que, para se ter maior segurança jurídica, não dependendo apenas de jurisprudências oscilantes sobre a matéria, essa alternativa demandaria alterações normativas para que os casos de responsabilidade por fato de outrem, dos animais ou das coisas inanimadas, pudessem ser aplicados também para os casos de inteligência artificial.

¹⁹¹ FILHO, Sérgio Cavalieri. *Programa de Responsabilidade Civil*. 9. ed. São Paulo: Atlas, 2010, p. 193.

¹⁹² FILHO, Carlos Edison do Rêgo Monteiro; ROSENVALD, Nelson. Responsabilidade Civil indireta e inteligência artificial. In: BARBOSA, Mafalda Miranda *et al.* (Coord.). *Direito Digital e Inteligência Artificial: diálogos entre Brasil e Europa*. São Paulo: Editora Foco, 2021, p. 186.

O que se percebe é que esta é uma solução bem localizada para casos específicos, em que é consagrado o dever de guarda, o que reitera o fato de que a responsabilidade civil pelos danos causados por inteligência artificial não é única, mas sim dependente da relação com o ambiente em que se encontra e suas funções na sociedade.

2.3.2.3 Código de Defesa do Consumidor e a responsabilidade pelo defeito do produto ou serviço

Por fim, um outro viés da responsabilidade objetiva usualmente explorado nos casos envolvendo inteligência artificial diz respeito ao defeito do produto ou serviço. Considerando que muitos sistemas de inteligência artificial serão utilizados e comercializados sob a lógica consumerista, em caso de danos causados pelos produtos ou serviços imbuídos de inteligência artificial, aplicar-se-iam o art. 12¹⁹³ e o art. 14¹⁹⁴ do Código de Defesa do Consumidor – CDC.

O defeito causador de acidente de consumo é, então, “o elemento gerador da responsabilidade civil objetiva no regime do Código”¹⁹⁵. Como previsto no próprio CDC, o defeito se verifica quando o produto ou o serviço não oferece a segurança que dele o consumidor legitimamente pode esperar. É claro que qualquer produto colocado em circulação ou serviço oferecido carregam em si, intrinsecamente, algum grau de periculosidade, seja ele mínimo ou considerável. Segundo Herman Benjamin, “o direito, de regra, só atua quando a insegurança ultrapassa o patamar da normalidade e da

¹⁹³ “Art. 12. O fabricante, o produtor, o construtor, nacional ou estrangeiro, e o importador respondem, independentemente da existência de culpa, pela reparação dos danos causados aos consumidores por defeitos decorrentes de projeto, fabricação, construção, montagem, fórmulas, manipulação, apresentação ou acondicionamento de seus produtos, bem como por informações insuficientes ou inadequadas sobre sua utilização e riscos; § 1º O produto é defeituoso quando não oferece a segurança que dele legitimamente se espera, levando-se em consideração as circunstâncias relevantes, entre as quais: I - sua apresentação; II - o uso e os riscos que razoavelmente dele se esperam; III - a época em que foi colocado em circulação.” (BRASIL. *Lei n. 8.078*, de 11 de setembro de 1990. Institui o Código de Defesa do Consumidor. Brasília, 1990. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/18078compilado.htm. Acesso em: 23 maio 2022.)

¹⁹⁴ “Art. 14. O fornecedor de serviços responde, independentemente da existência de culpa, pela reparação dos danos causados aos consumidores por defeitos relativos à prestação dos serviços, bem como por informações insuficientes ou inadequadas sobre sua fruição e riscos. § 1º O serviço é defeituoso quando não fornece a segurança que o consumidor dele pode esperar, levando-se em consideração as circunstâncias relevantes, entre as quais: I - o modo de seu fornecimento; II - o resultado e os riscos que razoavelmente dele se esperam; III - a época em que foi fornecido.” (BRASIL. *Lei n. 8.078*, de 11 de setembro de 1990. Institui o Código de Defesa do Consumidor. Brasília, 1990. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/18078compilado.htm. Acesso em: 23 maio 2022.)

¹⁹⁵ BENJAMIN, Antonio Heman de Vasconcellos. Fato do produto ou serviço. In: BENJAMIN, Antonio Heman de Vasconcellos; BESSA, Leonardo Roscoe; MARQUES, Cláudia Lima. *Manual de Direito do Consumidor*. 9. ed. São Paulo: Thomson Reuters, 2020, p. 194.

previsibilidade do risco, consubstanciando-se em verdadeiro defeito”¹⁹⁶. Ainda segundo o autor, a excessividade do grau de periculosidade e a imprevisibilidade do perigo, que permite qualificar um produto ou serviço como defeituoso, devem ser assim compreendidas pelo consumidor, aqui considerado como um ente em abstrato, ou seja, não o consumidor-vítima, mas a sociedade de consumo como um todo¹⁹⁷.

Dessa forma, nos casos envolvendo inteligência artificial, o dano decorrente de algum comportamento fora do padrão legitimamente esperado para aquele tipo de sistema (inclusive decorrente de condutas emergentes) ensejaria o dever de reparar do fornecedor, considerado em toda a extensão do seu conceito previsto no art. 3º do CDC. De acordo com Cláudia Lima Marques:

O art. 3º do CDC bem especifica que o sistema de proteção do consumidor considera como fornecedores todos os que participam da cadeia de fornecimento de produtos e da cadeia de fornecimento de serviços (o organizador da cadeia e os demais partícipes do fornecimento direto e indireto, mencionados genericamente como “toda pessoa física ou jurídica, pública ou privada, nacional ou estrangeira, bem como os entes despersonalizados, que desenvolvem atividades de [...] prestação de serviços”), não importando sua relação direta ou indireta, contratual ou extracontratual com o consumidor¹⁹⁸.

Toda essa abrangência do conceito de fornecedor, com efeito, traria maior tranquilidade em relação ao questionamento sobre quem responde pelos danos causados por condutas do sistema de inteligência artificial na relação de consumo. Nesse sentido, Gustavo Tepedino e Rodrigo Guia afirmam que:

Aduz-se, ainda, a possibilidade de aplicação do regime da responsabilidade pelo fato do produto ou serviço previsto pelo Código de Defesa do Consumidor (CDC). Afinal, a inteligência artificial pode ser utilizada no âmbito de atividade de fornecimento de produtos ou serviços ao mercado de consumo. Caso se configure relação de consumo à luz da disciplina do CDC, torna-se indubitosa a possibilidade de responsabilização de todos os fornecedores integrantes da cadeia de consumo pelos danos decorrentes de fato do produto

¹⁹⁶ BENJAMIN, Antonio Heman de Vasconcellos. Fato do produto ou serviço. In: BENJAMIN, Antonio Heman de Vasconcellos; BESSA, Leonardo Roscoe; MARQUES, Cláudia Lima. *Manual de Direito do Consumidor*. 9. ed. São Paulo: Thomson Reuters, 2020, p. 184.

¹⁹⁷ BENJAMIN, Antonio Heman de Vasconcellos. Fato do produto ou serviço. In: BENJAMIN, Antonio Heman de Vasconcellos; BESSA, Leonardo Roscoe; MARQUES, Cláudia Lima. *Manual de Direito do Consumidor*. 9. ed. São Paulo: Thomson Reuters, 2020, p. 195.

¹⁹⁸ MARQUES, Cláudia Lima. Campo de aplicação do CDC. In: BENJAMIN, Antonio Heman de Vasconcellos; BESSA, Leonardo Roscoe; MARQUES, Cláudia Lima. *Manual de Direito do Consumidor*. 9. ed. São Paulo: Thomson Reuters, 2020, p. 137.

ou serviço – resguardada, em qualquer caso, a necessidade de aferição dos demais elementos relevantes para a deflagração do dever de indenizar¹⁹⁹.

Na literatura estrangeira, Patrick Hubbard reivindica que o arcabouço legal de Responsabilidade Extracontratual dos Estados Unidos, no que diz respeito ao *product liability* (que, guardadas as peculiaridades de cada sistema jurídico, corresponderia à responsabilidade pelo fato do produto no Brasil), é suficiente para garantir um justo balanceamento entre a compensação de danos sofridos por humanos e o incentivo ao desenvolvimento da inteligência artificial sofisticada. Nesse sentido, Hubbard afirma que:

The problem of emergent behavior would be addressed under current law by requiring plaintiffs to provide reliable expert testimony: (1) that the use of emergent behavior (or a particular approach to emergent behavior) did not satisfy the cost–benefit standard for design and for warnings and instructions, and (2) that this failure to satisfy the standard caused the plaintiff’s injury²⁰⁰.

Acredita-se que, de fato, muitas hipóteses de uso de sistemas de inteligência artificial estarão abrangidas pela incidência do Código de Defesa do Consumidor, mas, ao contrário do que possa parecer, a questão da responsabilidade pelo defeito do produto ou do serviço em casos envolvendo inteligência artificial também não é facilmente resolvida.

Conforme já exposto por esta autora no artigo “Inteligência artificial e responsabilidade civil sob a perspectiva do direito do consumidor”:

Admite-se que, em relação aos sistemas autônomos, há dificuldade de aplicação do art. 12 do CDC e de suas bases teóricas. Isso porque a conduta emergente não pode ser confundida com defeito do produto, haja vista que os sistemas de inteligência artificial têm alta porosidade ao ambiente em que se encontram, de modo que o defeito pode surgir ao longo do uso deste pelo próprio consumidor e, conseqüentemente, esses defeitos seriam a ele imputáveis. Se assim for, responsabilizar o fornecedor pelo defeito emergente seria ônus excessivo a este, o que poderia ter impacto negativo sobre o

¹⁹⁹ SILVA, Rodrigo da Guia; TEPEDINO, Gustavo. Inteligência Artificial e elementos da Responsabilidade Civil. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Thomsom Reuters Brasil, 2019, p. 319.

²⁰⁰ O problema do comportamento emergente seria endereçado conforme a lei mediante o testemunho de especialistas em favor do autor no sentido de que: (1) a abordagem do fabricante em relação ao comportamento emergente não foi suficiente comparado ao padrão de custo-benefício em termos de design e instruções ao consumidor e que (2) que falha no atendimento desses padrões causou dano ao autor da ação [tradução nossa]. HUBBARD, F. Patrick. Allocating the risk of physical injury from “sophisticated robots”: Efficiency, fairness, and innovation, p. 42. In: CALO, Ryan; FROMKIN, A. Michael; KERR, Ian. *Robot Law*. Northampton: Edward Elgar Pub, 2016, p. 42.)

impulsionamento tecnológico e o desenvolvimento de sistemas de inteligência artificial²⁰¹.

Em outras palavras, na medida em que o sistema de inteligência artificial é capaz de se transmutar ao longo do tempo, de acordo com os estímulos que recebe (que, inclusive, podem ser advindos do próprio consumidor, o seu usuário), caso a resposta jurídica para os casos de danos envolvendo inteligência artificial, sob a égide das relações consumeristas, fosse a responsabilidade objetiva do fornecedor pelo defeito do produto ou serviço, estar-se-ia diante de uma regra de patente injusta, já que a conduta do fornecedor poderia em nada ter colaborado para a ocorrência da conduta lesiva do sistema.

Outros autores também se filiam a este entendimento, como Marcelo Milagres:

Parece que a teoria dos vícios e defeitos dos produtos não mais alcançaria situações possíveis em que não há evidente ou comprovado erro na programação ou na confecção dessas máquinas, mas subsistem danos decorrentes de escolhas provenientes de um processo de autoaprendizagem²⁰².

Nesse mesmo sentido, Eduardo Magrani afirma:

[...] na medida em que os robôs atingirem independência com relação aos humanos, [...] a mencionada lógica de responsabilização de cadeia consumerista não se faz mais possível, surgindo a necessidade de atribuir direitos e até eventualmente uma personalidade a robôs com nível de autonomia elevado além da possibilidade de criação de um seguro em casos de acidentes causados por robôs²⁰³.

Ressalvadas as diferenças entre os sistemas de *Common Law* e *Civil Law*, Čerka Grigienė e Sirbikytė também sustentam a impossibilidade da responsabilidade pelo fato do produto:

Also in some cases it would be difficult to apply the product liability case, because AI is a self-learning system that learns from its experience and can take autonomous decisions. Thus, for the plaintiff it would be difficult to prove an AI product defect and especially that the defect existed when AI left its manufacturer's or developer's hands. It is hard to believe that it is possible to

²⁰¹ RIBEIRO, Júlia Melo Carvalho. Inteligência artificial e responsabilidade civil sob a perspectiva do direito do consumidor. In: CHAVES, Natália Cristina; COLOMBINI, Henry (Org.). *Direito e Tecnologia: Novos modelos e tendências*. Porto Alegre: Editora Fi, 2021, p. 326.

²⁰² MILAGRES, Marcelo de Oliveira. A robótica e as discussões sobre a personalidade eletrônica. In: JÚNIOR, Marcos Ehrhardt Júnior; CATALAN, Marcos; MALHEIROS, Pablo (Coord.) *Direito Civil e Tecnologia*. Belo Horizonte: Forum, 2020. p. 517.

²⁰³ MAGRANI, Eduardo; SILVA, Priscilla; VIOLA, Rafael. Novas perspectivas sobre ética e responsabilidade de inteligência artificial. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 137.

draw the line between damages resulting from the AI will, i.e. derived from self-decision, and damages resulting from product defect; unless we would equate the independent decision-making (which is a distinctive AI feature) with a defect²⁰⁴.

Além desses exemplos, poderiam ser citadas outras tantas doutrinas contrárias à aplicação da responsabilidade objetiva pelo defeito do produto ou do serviço nos casos de inteligência artificial²⁰⁵. O que se tem, então, é que se engana quem assume que, para as hipóteses de defeito do produto ou do serviço dotado de inteligência artificial, o ordenamento jurídico pátrio apresenta uma resposta pronta. Ainda que se tenha o mecanismo de responsabilidade objetiva para estes casos de defeitos, isso não significa que esta resposta é justa. Parece que, ao contrário, a aplicação dos referidos dispositivos nos casos de danos gerados por inteligência artificial poderia acarretar ônus excessivo e injusto ao fornecedor, uma vez que o defeito poderia advir da conduta do próprio consumidor.

A outra faceta da aplicação do CDC para esses casos, então, seria a demonstração de culpa exclusiva da vítima para exclusão da responsabilidade do fornecedor. Trata-se de solução que, igualmente, poderia levar a injustiças, haja vista a dificuldade da prova, que deveria identificar precisamente que a exposição do sistema a algum ambiente específico, proporcionado pelo próprio consumidor, teria levado a um determinado dano.

Como uma anotação final sobre a aplicação da responsabilidade objetiva sob a lógica consumerista, menciona-se brevemente a questão do risco do desenvolvimento, que é levantada por alguns pesquisadores como possível solução para os casos envolvendo inteligência artificial, considerando que a dificuldade em se aferir a responsabilidade objetiva do fornecedor relaciona-se, justamente, ao fato de que eventual

²⁰⁴ ČERKA, Paulius; GRIGIENĖ, Jurgita; SIRBIKYTĖ, Gintarė. Liability for damages caused by Artificial Intelligence. *Computer Law & Security Review*, v. 31, n. 3, p. 376-389, 2015. Disponível em: https://is.muni.cz/el/1422/podzim2017/MV735K/um/ai/Cerka_Grigiene_Sirbikyte_Liability_for_Damages_caused_by_AI.pdf. Acesso em: 23 maio 2022.

²⁰⁵ No mesmo sentido, Mafalda Miranda Barbosa afirma sobre o defeito do produto: “Em primeiro lugar, os danos causados por robots, software ou algoritmos em geral podem não resultar de um defeito de conceção ou de um defeito de fabrico. Dito de outro modo, a idealização do robot (programação do software) pode não apresentar qualquer defeito, do mesmo modo que, na fase do fabrico do mecanismo no qual se integra a inteligência artificial, pode não ocorrer qualquer desconformidade entre o resultado final e o que era esperado pelo produtor. Os danos causados pelo robot dito inteligente são gerados pela sua atuação autónoma que, longe de ser uma marca de defeituosidade, se traduz numa sua característica intrínseca”. (BARBOSA, Mafalda Miranda. O futuro da responsabilidade civil desafiada pela inteligência artificial: as dificuldades dos modelos tradicionais e caminhos de solução. *Revista de Direito da Responsabilidade*, p. 288, 2020. Disponível em: <https://revistadireitoresponsabilidade.pt/2020/o-futuro-da-responsabilidade-civil-desafiada-pela-inteligencia-artificial-as-dificuldades-dos-modelos-tradicionais-e-caminhos-de-solucao-mafalda-miranda-barbosa/>. Acesso em: 23 maio 2022.)

conduta danosa do sistema pode vir à tona apenas em momento posterior ao lançamento do produto no mercado de consumo.

Segundo Herman Benjamin, “define-se o risco de desenvolvimento como aquele que não pode ser cientificamente conhecido ao momento do lançamento do produto no mercado, vindo a ser descoberto somente após um certo período de uso do produto e do serviço”²⁰⁶. À parte o debate acerca do risco de o desenvolvimento ser considerado como hipótese de exoneração de responsabilidade do fornecedor ou não²⁰⁷, alguns autores defendem que a conduta emergente do sistema de inteligência artificial se enquadra nos requisitos do risco do desenvolvimento. Caitlin Mulholland é uma das autoras que vincula o comportamento emergente da tecnologia aos elementos do risco do desenvolvimento da seguinte forma:

(i) O dano é casualmente ligado a um sistema de IA; (ii) é virtualmente impossível identificar, no momento inicial de programação de IA; e (iii) o desenvolvimento da aprendizagem autônoma pela IA, que independe de interferência humana, causa efetivamente o dano a uma pessoa. A esses três elementos soma-se um quarto, que é (iv) a inviabilidade de explicação por um humano do processo que levou a IA a uma decisão autônoma geradora do resultado danoso, na medida em que esses processos de aprendizagem e decisão independem da atuação e da racionalidade humana²⁰⁸.

Apesar de, realmente, os danos decorrentes de comportamento autônomo da inteligência artificial terem seus pontos de contato com a teoria do risco do desenvolvimento, despontam-se algumas diferenças que devem ser levadas em conta. Avalia-se, por exemplo, um caso prático em que determinada companhia coloca em circulação um robô capaz de desenvolver diálogos com seres humanos, os chamados

²⁰⁶ BENJAMIN, Antonio Heman de Vasconcellos. Fato do produto ou serviço. In: BENJAMIN, Antonio Heman de Vasconcellos; BESSA, Leonardo Roscoe; MARQUES, Cláudia Lima. *Manual de Direito do Consumidor*. 9. ed. São Paulo: Thomson Reuters, 2020, p. 204.

²⁰⁷ Segundo Heman Benjamin, “há argumentos contundentes – e prevalentes – no sentido do não reconhecimento da excludente”. (BENJAMIN, Antonio Heman de Vasconcellos. Fato do produto ou serviço. In: BENJAMIN, Antonio Heman de Vasconcellos; BESSA, Leonardo Roscoe; MARQUES, Cláudia Lima. *Manual de Direito do Consumidor*. 9. ed. São Paulo: Thomson Reuters, 2020, p. 205.) Para Tula Wesendonck, no entanto, “a exclusão da responsabilidade do produtor pelos riscos do desenvolvimento só tem lugar num ordenamento que adota apenas o sistema de responsabilidade civil subjetiva, baseado na culpa e não num sistema de responsabilidade baseado no risco, no qual aquele que cria o risco deve responder pelo perigo colocado por si, não sendo plausível que os riscos do desenvolvimento sejam suportados pela vítima”. (WESENDONCK, Tula. A responsabilidade civil pelos danos decorrentes dos riscos do desenvolvimento dos produtos postos em circulação. In: MILAGRES, Marcelo; ROSENVALD, Nelson (Coord.). *Responsabilidade Civil: novas tendências*. 2. ed. Indaiatuba: Editora Foco, 2018, p. 425.)

²⁰⁸ MULHOLLAND, Caitlin. Responsabilidade civil e processos decisórios autônomos em sistemas de inteligência artificial (IA): autonomia, imputabilidade e responsabilidade. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019, p. 335.

chatbots. O robô sai de fábrica sendo capaz de apreender as falas que capta no ambiente e de replicar frases, fazer novas inferências e “pronunciar” novas sentenças a partir do conteúdo que adquiriu. Caso esse robô seja posto em um ambiente pernicioso, em que há, por exemplo, falas discriminatórias constantes, é possível que o *chatbot* também passe a ter esse tipo de comportamento, o que pode levar à ofensa de algum interlocutor, em eventual diálogo entre este e a máquina. Nesse caso, não se pode dizer que o problema foi decorrente de risco do desenvolvimento, pois o dano nunca teria ocorrido se não houvesse a influência do meio em que o próprio usuário imergiu o sistema.

Feitas essas breves análises sobre as possíveis respostas aos desafios trazidos pela inteligência artificial ao instituto da responsabilidade civil, é importante lembrar que essas possíveis soluções não se encerram nas proposições destacadas acima. Estas são apenas algumas, encontradas com grande recorrência na literatura especializada. Apesar de a compaginação de todas as propostas, já estudadas na literatura nacional e internacional, não ser apropriada, ou mesmo viável, apenas para não deixar sem menção algumas outras interessantes respostas citam-se: a ideia de responsabilidade civil resolúvel, em que, com o passar do tempo, a responsabilidade pelos danos causados pelo sistema aos poucos deixaria de repousar sobre o seu desenvolvedor/fornecedor para passar ao próprio usuário²⁰⁹; a criação de patrimônio de afetação, para dotar o próprio algoritmo de capacidade patrimonial, sem a necessidade de criação de personalidade jurídica²¹⁰; e a criação de seguros ou fundos obrigatórios para socialização dos danos causados pela inteligência artificial²¹¹.

2.4 INVIABILIDADE DA PREVISÃO DE UM REGIME DE RESPONSABILIDADE CIVIL ÚNICO PARA OS CASOS DE DANOS GERADOS POR IA

²⁰⁹ FREITAS, Wallace Almeida de. Artificial Intelligence and Solvable Civil Responsibility. In: CARDOSO, Renato César; PARENTONI, Leonardo Netto (Org.). *Law, Technology and Innovation – v. II: Insights on Artificial Intelligence and the Law*. Belo Horizonte: Editora Expert, 2021, p. 291-308. Disponível em: <https://www.dtibr.com/livros-e-artigos>. Acesso em: 23 maio 2022.

²¹⁰ Para uma leitura sobre este tema, vide: SPALER, Mayara Guibor; XAVIER, Luciana Pedroso. Patrimônio de afetação: uma possível solução para os danos causados por sistemas de inteligência artificial. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2019.

²¹¹ Para uma leitura sobre este tema, vide: DIAS, Daniel. Implementação de seguro obrigatório de responsabilidade civil no contexto da inteligência artificial. In: SILVA, Rodrigo da Guia; TEPEDINO, Gustavo (Coord.). *O Direito Civil na era da inteligência artificial*. São Paulo: Thomson Reuters Brasil, 2020.

Pelo exposto no tópico anterior, resta claro que há um embate entre a inteligência artificial autônoma e os atuais mecanismos para a responsabilização civil. Apesar dessa clara constatação, entende-se não ser possível resolver o problema posto, qual seja, quem e como responde pelos danos gerados por inteligência artificial, de forma genérica.

Isso porque, conforme se demonstrou nas linhas iniciais deste trabalho, os tipos de inteligência artificial são muitas, bem como são diversos os mercados em que estão inseridas, sem se mencionar a sua rápida evolução. Como sintetiza Felipe Medon:

Quem responde pelos danos causados pela Inteligência Artificial? Diante de tudo que foi visto no primeiro capítulo, não se pode pretender dar uma resposta única para essa indagação. Afinal, a Inteligência Artificial abrange diversas espécies e, mesmo dentro destas, há que se atentar para os diferentes graus de autonomia e de delegação de atividades humanas para a máquina. Isso impactaria de maneira decisiva nos diferentes danos a serem causados pela máquina e, conseqüentemente, na responsabilidade civil²¹².

Anderson Schreiber também se posiciona nesse sentido:

E, aqui, vem à tona um aspecto que merece reflexão desapressada: o que se designa, hoje, sob o rótulo único da Inteligência Artificial corresponde, na verdade, a um conjunto relativamente amplo de tecnologias, que se prestam aos mais diferentes usos na realidade social e econômica. Cada uma destas aplicações da Inteligência Artificial é dotada de peculiaridades e especificidades que podem e devem se refletir sobre a construção de um modelo normativo ideal de responsabilização por danos. Nesse contexto, a criação de um regime único de responsabilidade civil pode acabar por resultar em uma generalização indevida. Seria, mal comparando, como criar um regime único de responsabilidade civil para todos os usos da Medicina²¹³.

Por fim, para corroborar a tese, Mafalda Miranda Barbosa afirma:

Apesar de nos referirmos ao problema dos danos causados por entes dotados de inteligência artificial de modo uniforme, importa salientar que não estamos a falar de uma realidade única e simples. Na verdade, existem diversos mecanismos dotados de inteligência artificial, com potenciais lesivos diversos, quer no que respeita ao tipo de danos que podem gerar, quer no que respeita à sua magnitude²¹⁴.

²¹² MEDON, Filipe. *Inteligência Artificial e Responsabilidade Civil: autonomia, riscos e solidariedade*. Editora JusPodivm: São Paulo, 2022, p. 157.

²¹³ SCHREIBER, Anderson. *PL da Inteligência Artificial cria fratura no ordenamento jurídico brasileiro*. Jota, 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/coluna-do-anderson-schreiber/pl-inteligencia-artificial-cria-fratura-no-ordenamento-juridico-02112021>. Acesso em: 10 fev. 2022.

²¹⁴ BARBOSA, Mafalda Miranda. O futuro da responsabilidade civil desafiada pela inteligência artificial: as dificuldades dos modelos tradicionais e caminhos de solução. *Revista de Direito da Responsabilidade*, p. 318, 2020. Disponível em: <https://revistadireitoresponsabilidade.pt/2020/o-futuro-da-responsabilidade-civil-desafiada-pela-inteligencia-artificial-as-dificuldades-dos-modelos-tradicionais-e-caminhos-de-solucao-mafalda-miranda-barbosa/>. Acesso em: 23 maio 2022.

Como se viu nos tópicos anteriores, são diversas as soluções propostas para solucionar a questão da responsabilidade civil e o comportamento emergente, algumas mais, outras menos viáveis, mas fato é que, provavelmente, tantos serão os caminhos a serem adotados juridicamente para a responsabilização, quantas são as utilidades da inteligência artificial na vida contemporânea. Não se espera, assim, ter uma resposta única em termos de sujeito responsável e de fundamento da responsabilidade civil para toda a polivalência da tecnologia em análise.

Em que pese a impossibilidade de se resolver o problema colocado neste trabalho de forma genérica, não foi opção desta autora identificar uma funcionalidade específica da inteligência artificial para tentar resolver o problema na concretude do caso, por entender que o Direito carece de determinada sistematização. Portanto, tem-se, assim, o embate de uma tecnologia, que oferece vários matizes e nuances em termos de previsibilidade e risco, e, por outro lado, uma matéria que melhor funciona sob espectros generalistas, sistemáticos e por raciocínio dedutivo.

Não obstante isso, a dificuldade em se determinar genericamente um regime único de responsabilidade civil não impede a fixação de diretrizes que conduzirão à resposta quanto à forma de responsabilização a ser aplicada no caso concreto. Tendo isso em vista, no próximo capítulo serão delimitadas essas diretrizes, a partir da análise da melhor forma de se regular a inteligência artificial do ponto de vista da responsabilidade civil.

3 INTELIGÊNCIA ARTIFICIAL E RESPONSABILIDADE CIVIL SOB A ÓTICA DA REGULAÇÃO PRINCIPIOLÓGICA

Pelo que foi exposto no capítulo anterior, conclui-se que, apesar de haver evidente conflito entre algumas qualidades da inteligência artificial e as atuais regras de responsabilidade civil positivadas no Brasil, entende-se não ser conveniente tentar resolver este conflito de forma *ante facto* e genérica. Diante disso, será investigado neste capítulo como se poderia regular a inteligência artificial sob o ponto de vista da responsabilidade civil da forma mais eficiente, haja vista a necessidade de potencialização do desenvolvimento e o uso da tecnologia de maneira benéfica para a sociedade, sem que seja deixada de lado a proteção dos indivíduos aos potenciais riscos apresentados pela inteligência artificial.

Em primeiro lugar, é preciso identificar o que será chamado de regulação nas próximas linhas deste trabalho, considerando que este termo alberga vários possíveis significados. Vale-se aqui do importante artigo de Christel Koop e Martin Lodge, denominado “What is regulation? An interdisciplinary concept analysis”, que exalta o dissenso a respeito da definição do termo “regulação”, principalmente por ser um termo utilizado em várias disciplinas, tais como Ciências Políticas, Economia, Sociologia, Administração, além do Direito. A partir do estudo dos 109 artigos mais citados dessas disciplinas que contêm a palavra regulação, eles encontram os elementos centrais para a definição do termo, que são apresentados por meio dos questionamentos abaixo:

(A) Is regulation by definition intentional or can it also be non-intentional? (B) Is regulation a distinct type of intervention, different from, for instance, taxation and subsidisation? (C) Can regulation be carried out by state as well as non-state actors? (D) Is regulation by definition targeted at economic activities or may other activities be subject to it too? (E) Do regulator and regulatee need to be separate actors²¹⁵?

Os autores chegam à conclusão de que a regulação é, na maior parte das vezes, tratada como uma forma intencional de intervenção nas atividades econômicas, sendo dado maior destaque à regulação feita pelo setor público direcionada ao setor privado, em

²¹⁵ “(A) A regulação é, por definição, intencional ou também pode ser não intencional? (B) A regulação é um tipo distinto de intervenção, diferente de, por exemplo, tributação e subsídio? (C) A regulação pode ser realizada por atores estatais e não estatais? (D) A regulação é, por definição, direcionada às atividades econômicas ou outras atividades também podem estar sujeitas a ela? (E) O regulador e o regulado precisam ser atores separados?” (LODGE, Martin; KOOP, Christel. What is regulation? An interdisciplinary concept analysis. *Regulation and Governance*, v. 11, p. 5, 14 Jul. 2015, tradução nossa.)

detrimento da regulação pelo setor privado (tal como a autorregulação), ou da possibilidade de o Governo ser o próprio ente regulado²¹⁶.

Adotando as conclusões obtidas por Martin e Christel, para fins deste capítulo, compreender-se-á a regulação como uma atividade intencional e direta estabelecida pelo Governo, ou por seus delegados. Não serão consideradas, portanto, as atividades espontâneas que, de forma não intencional, moldam um determinado comportamento, tampouco diretrizes, regras, ou normas concebidas pelos agentes privados, visando conformar uma certa atividade. Embora se reconheça o poder e a importância desse tipo de regulação, considerada no seu sentido amplo, para melhor sistematização e delimitação deste trabalho, optou-se por focar na regulação governamental intencional e direta, nesse caso relacionada à inteligência artificial. Observa-se que estão incluídas nesta definição de regulação não apenas as típicas leis positivadas, vinculantes e cujos comandos correspondem, necessariamente, a uma sanção em caso do seu descumprimento, mas também diretrizes deontológicas e recomendações governamentais, por se entender que esse tipo de regulação é instrumental no estágio, ainda incipiente, de normatização das questões jurídicas trazidas pela inteligência artificial.

A partir dessa definição, passa-se à reflexão de como deve ser a melhor regulação para a inteligência artificial sob os aspectos da responsabilidade civil, no atual estágio de desenvolvimento e emprego da tecnologia. É preciso lembrar que o conflito entre as regras positivadas e as inovações tecnológicas não é uma novidade. Ao menos desde o capitalismo industrial, marcado pela destruição criativa, como teorizou Schumpeter²¹⁷, observa-se o embate entre o Direito e as novas demandas da sociedade. Atualmente, esse fenômeno é apenas potencializado, haja vista a maior velocidade em que as novas tecnologias se reciclam e rompem com o *status quo* anterior a elas.

No contexto brasileiro contemporâneo, remontando à ocasião em que o Código Civil de 2002 foi publicado, é conhecido que este estatuto recebeu severas críticas quanto à sua inadequação aos avanços tecnológicos já havidos àquele tempo. Quanto a isso, Rui Stoco afirma:

Crítica tão contundente não merece o novo estatuto civil que, se por um lado, rege e deve disciplinar a vida civil das pessoas e o seu relacionamento, por outro não pode ser esperado como panaceia para todos os males ou repositório dos avanços tecnológicos e científicos, posto dinâmicos e muito mais velozes

²¹⁶ LODGE, Martin; KOOP, Christel. What is regulation? An interdisciplinary concept analysis. *Regulation and Governance*, v. 11, p. 17, 14 Jul. 2015.

²¹⁷ SCHUMPETER, Joseph Alois. *Capitalismo, Socialismo e Democracia*. São Paulo: Editora Unesp, 2017.

do que o próprio legislador, sendo de necessária prudência que leis que tendem a se tornar obsoletas e superadas e possam ter vida efêmera sejam editadas em separado²¹⁸.

A lição de Rui Stoco é válida também para a análise da regulação da inteligência artificial. Caso se proponha uma solução fechada para identificar o responsável pela reparação de eventuais danos causados pelo sistema de inteligência artificial, é possível que em poucos anos (ou até meses) essa solução se torne defasada, ou pouco aderente aos novos contornos da tecnologia, tendo em vista o seu rápido avanço e obsolescência. É preciso lembrar que, necessariamente, o Direito virá sempre a reboque do desenvolvimento tecnológico, o que serve como um alerta de cautela para os reguladores, que devem se voltar a normas mais flexíveis e revisitáveis de tempos em tempos.

Ainda sobre os principais desafios de regulação na sociedade atual, marcada pela disrupção tecnológica, pela globalização e pelo intenso fluxo informacional, Pamela Samuelson²¹⁹ afirma que, em primeiro lugar, deve-se avaliar se as normas atuais são suficientes, ou se é necessária a criação de novas regras para regular os desafios trazidos pelas novidades tecnológicas. Nesse segundo caso, de necessidade de estabelecimento de novas regras, estas devem ser flexíveis, proporcionais e razoáveis quando confrontadas com os riscos e benefícios trazidos pela tecnologia objeto da regulação, equilibrando o avanço tecnológico e os direitos fundamentais dos usuários. Por fim, tendo em vista o fato de essas novas tecnologias serem usualmente utilizadas em escala global, com compartilhamento de dados e informações provenientes de todo o mundo, deve haver um esforço para que essas regras sejam padronizadas globalmente²²⁰.

Nesse sentido, os Estados Unidos da América, país que tem tradição de menor intervenção estatal, em uma postura de incentivo à livre iniciativa, editaram, em 2020, o “Guidance for Regulation of Artificial Intelligence Applications”, documento que contém o guia para se identificar a necessidade ou não de regular a inteligência artificial e alguns requisitos necessários para a regulação, caso opte-se por tê-la. É mencionado que eventual

²¹⁸ STOCO, Rui. *Tratado de responsabilidade civil: doutrina e jurisprudência*. 7. ed. São Paulo: Revista dos Tribunais, 2007, p. 111

²¹⁹ SAMUELSON, Pamela. Five Challenges for Regulating the Global Information Society. *Social Science Research Network*, California, 9 Aug. 2000. Disponível em: <https://ssrn.com/abstract=234743>. Acesso em: 09 fev. 2022.

²²⁰ RIBEIRO, Júlia Melo Carvalho; PENNA, Thomaz Murta. A (des)regulação da tecnologia blockchain: uma análise da experiência regulatória brasileira. In: PARENTONI, Leonardo; MILAGRES, Marcelo de Oliveira; VAN DE GRAAF, Jeroen (Coord.); MOREIRA, Arthur Salles de Paula; CHAGAS, Ciro Costa; SANTANA, Mariana Damiani (Org.). *Direito, Tecnologia e Inovação – V. III: aplicações jurídicas de blockchain*. Belo Horizonte: Editora Expert, 2021, p. 303.

regulamento deve ser flexível, no sentido de ser tecnologicamente neutro, haja vista que a tecnologia sempre se aprimorará de forma mais rápida que o desenvolvimento regulatório. Além disso, a escolha pela regulação apenas deve ocorrer quando ela aumentar significativamente o nível de confiança da sociedade na inteligência artificial, sendo que seus prós e os contras devem ser levantados com forte embasamento técnico e científico, por meio da elaboração de relatórios de análise e mitigação de riscos. Por fim, é mencionada a necessidade de análise custo-benefício da regulação, quando se cita o exemplo da responsabilidade civil:

For example, while the broader legal environment already applies to AI applications, the application of existing law to questions of responsibility and liability for decisions made by AI could be unclear in some instances, leading to the need for agencies, consistent with their authorities, to evaluate the benefits, costs, and distributional effects associated with any identified or expected method for accountability²²¹.

Como uma recomendação a mais para regular novas tecnologias como um todo e, no caso estudado neste trabalho, a inteligência artificial, citam-se as ásperas críticas de Frank Easterbrook, feitas ainda em 1996, no famoso artigo “Cyberspace and the Law of the Horse”²²². Nesse artigo, publicado pela Chicago Law School, o autor reflete sobre o cenário acadêmico da época e a extrema especialização das disciplinas ofertadas, em especial, sobre o estudo do direito de propriedade intelectual no ciberespaço. De forma retórica, ele se refere ao exemplo de um suposto estudo sobre a “lei do cavalo”: vários casos lidam com a venda de cavalos, outros com pessoas que foram lesionadas por um cavalo, outros com o cuidado (ou falta dele) no tratamento dos cavalos e, ainda, casos em que se avaliam o licenciamento e os prêmios em corridas de cavalos. Segundo ele, se todos esses casos fossem tratados sob a insígnia de “lei dos cavalos”, esta estaria fadada à superficialidade e à falta de princípios unificadores e sistematizados²²³.

²²¹ “Por exemplo, embora algumas regras jurídicas sejam aplicáveis à inteligência artificial e seus variados usos, a aplicação da lei existente a questões de responsabilidade por decisões tomadas por IA pode não ser clara em alguns casos, levando à necessidade de autoridades avaliarem os custos, os benefícios e os efeitos distributivos associados a qualquer método de prestação de contas que seja imposto”. (ESTADOS UNIDOS DA AMÉRICA. *Guidance for Regulation of Artificial Intelligence Applications*. Memorandum for the Heads of Executive Departments and Agencies, p. 4, tradução nossa. Disponível em: <https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf>. Acesso em: 10 fev. 2022)

²²² EASTERBROOK, Frank Hoover. *Cyberspace and the Law of the Horse*. Chicago: University of Chicago Legal Forum, 1996. Disponível em https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=2147&context=journal_articles. Acesso em 10 fev. 2022.

²²³ EASTERBROOK, Frank Hoover. *Cyberspace and the Law of the Horse*. Chicago: University of Chicago Legal Forum, 1996, p. 207.

Sob essas lições a respeito da regulação de novas tecnologias é que se conclui que a melhor forma de regular a inteligência artificial no que diz respeito à responsabilidade civil, ao menos no seu atual estágio de desenvolvimento, é a partir de um colchão principiológico genérico, válido para diversas aplicabilidades da inteligência artificial, que futuramente poderá ser acompanhado por regras específicas, de acordo com o mercado em que a tecnologia é inserida (caso este tenha suas regras próprias) que orientará melhor o uso da inteligência artificial, considerando os diferentes tipos e graus de risco aos quais ela pode expor o indivíduo ou a coletividade. Ademais, optou-se pela análise da regulação principiológica da inteligência artificial, apostando que, no atual estado da técnica da IA, esse tipo de regulação seria mais favorável ao equilíbrio entre riscos e benefícios da tecnologia, sem que se arrisque a uma alteração substancial das regras positivadas de responsabilidade civil, o que poderia levar a um diletantismo equivalente à “lei dos cavalos”.

Vale a pena, aqui, lembrar das ponderações de Calixto Salomão Filho sobre a análise regulatória da atividade empresarial sob a lógica do desenvolvimento:

Assim, é preferível uma regulamentação que meramente enumere interesses envolvidos (através de dispositivos declaratórios) e os princípios a eles relacionados. [...] Referidos dispositivos declaratórios e princípios são, então, guias de interpretação para o restante da legislação, que deve ser mantida em diversos microssistemas, cada um deles iluminado por seu específico dispositivo declaratório e princípio²²⁴.

E ele arremata com as seguintes palavras:

O método de desenvolvimento e enumeração apenas de dispositivos declaratórios e princípios em um regulamento geral, com manutenção dos microssistemas, é, evidentemente, muito mais apto tanto ao reconhecimento de interesses quanto a permitir rápidas mudanças de entendimento legislativo e jurisprudencial – o que é fundamental para todo o Direito, e em particular para o direito empresarial²²⁵.

Embora se trate de outra matéria, as críticas do autor são bem-vindas também para o objeto de estudo deste trabalho, por se defender, no caso da inteligência artificial, justamente que aplicações específicas da tecnologia, que reclamem regras próprias, sejam

²²⁴ FILHO, Calixto Salomão. Regulamentação da atividade empresarial para o desenvolvimento. In: FILHO, Calixto Salomão (Org.). *Regulação e desenvolvimento: novos temas*. São Paulo: Editora Malheiros, 2012, p. 64.

²²⁵ FILHO, Calixto Salomão. Regulamentação da atividade empresarial para o desenvolvimento. In: FILHO, Calixto Salomão (Org.). *Regulação e desenvolvimento: novos temas*. São Paulo: Editora Malheiros, 2012, p. 64.

reguladas por microssistemas setoriais – por exemplo, que regulem carros autônomos, drones, diagnósticos médicos, entre outros, vinculados aos ramos de estudo já hoje existentes de automobilística, aeronáutica e medicina, respectivamente – que girem em torno dos ditames principiologicos válidos para todos esses microssistemas.

Voltando ao contexto da inteligência artificial, segundo Virgílio Almeida e Urs Gasser, uma regulação eficaz de sistemas de inteligência artificial deve incorporar mecanismos destinados a aprimorar a compreensão da sociedade sobre o fenômeno algorítmico, suas diferentes manifestações e contextos de aplicação e, para isso, é necessário ir além da clássica regulação de comando e controle e propor medidas mais flexíveis e adaptáveis, por meio de métodos não tradicionais²²⁶. Eles propõem, então, uma regulação “em camadas”, iniciando-se pela camada tecnológica, passando pela camada ética e chegando, ao final, na camada social e jurídica, a serem aplicadas respectivamente em curto, médio e longo prazo. A primeira camada diz respeito aos parâmetros técnicos de criação dos algoritmos. A ideia é que alguns princípios sejam incorporados *by desing*, ou seja, no momento inicial de criação do modelo algorítmico para que o próprio código seja imbuído de características que obedeçam a estes valores. Na sequência, a regulação se voltaria para zelar o uso ético do algoritmo, evitando que seus operadores e usuários o utilizassem de modo a violar os direitos humanos, como a igualdade e a não discriminação. Por fim, a camada social e jurídica cuidaria de criar instituições reguladoras e mecanismos de responsabilização relacionados à implementação e ao uso da inteligência artificial²²⁷.

Nota-se que o mencionado artigo segue a linha regulatória que ora se propõe, com atenção inicial a princípios gerais regentes da criação e desenvolvimento da inteligência artificial, deixando-se para um segundo momento a sistematização de regras específicas quanto à responsabilidade civil.

Nas reflexões de Erik Vermeulen *et al.* sobre como regular o “amanhã”, em referência à regulação de novas tecnologias, também são destacados os benefícios de regras principiologicas. De acordo com os autores:

²²⁶ GASSER, Urs; ALMEIDA, Virgílio Augusto Fernandes. A Layered Model for AI Governance. *IEEE Internet Computing*, v. 21, n. 6, p. 3-4, 2017. Disponível em: <https://dash.harvard.edu/handle/1/34390353>. Acesso em: 08 maio 2021.

²²⁷ GASSER, Urs; ALMEIDA, Virgílio Augusto Fernandes. A Layered Model for AI Governance. *IEEE Internet Computing*, v. 21, n. 6, p. 5-6, 2017. Disponível em: <https://dash.harvard.edu/handle/1/34390353>. Acesso em: 08 maio 2021.

Regulators need to abandon a fixation on finality and legal certainty and embrace contingency, flexibility and an openness to the new. This shift in perspective also affects *how* we regulate disruptive technologies. A key element of a more open approach involves a shift from rules to principles. Certainly, a principle-based approach facilitates a greater degree of openness and flexibility on the part of regulators and prevents innovative technologies (and the companies that have developed them) from becoming bogged down in the regulatory thicket that often results from a rules-focused approach. Re-framing regulation in this way and adopting a principle-based approach facilitates action, whilst also allowing future revisions in the regulatory regime to be based on the incorporation of new knowledge or subsequent discoveries²²⁸.

De forma similar, na proposta de Matthew Scherer, o primeiro passo para a regulação algorítmica seria a criação do *Artificial Intelligence Development Act* cujo propósito seria garantir que a inteligência artificial seja segura e suscetível ao controle humano, estabelecendo os princípios gerais para se alcançar tais propósitos²²⁹.

Pelo exposto acima, reconhece-se que a opção trazida neste trabalho de uma regulação principiológica da inteligência artificial encontra ressonância na literatura nacional e internacional. Não se limitando a isso, atuais discussões a nível Estatal sobre a regulação da inteligência artificial e os seus possíveis métodos transparecem, muitas vezes, a opção pela abordagem principiológica. No momento atual, observa-se uma resistência pela aplicação de leis vinculantes nos esquemas legislativos de cada país – o que, diante de todas as dificuldades que a inteligência artificial apresenta, parece ser uma opção prudente –, fazendo emergir diretrizes deontológicas para a regulação da inteligência artificial tanto no cenário doméstico quanto no internacional. Segundo

²²⁸ “Os reguladores precisam abandonar a fixação na finalidade e na segurança jurídica e adotar a contingência, a flexibilidade e a abertura ao novo. Essa mudança de perspectiva também afeta a forma como regulamos as tecnologias disruptivas. Um elemento-chave de uma abordagem mais aberta envolve uma mudança de paradigma de regras para princípios. Certamente, uma abordagem baseada em princípios facilitaria o maior grau de abertura e flexibilidade por parte dos reguladores e evitaria que tecnologias inovadoras (e as empresas que as desenvolveram) fiquem atoladas no emaranhado regulatório constituído por regras. Reenquadrar a regulação desta forma e adotar uma abordagem baseada em princípios facilita a ação, ao mesmo tempo que permite que futuras revisões no regime regulatório sejam baseadas na incorporação de novos conhecimentos ou descobertas posteriores à regulação original.” (FENWICK, Mark; KAAL, Wulf A.; VERMEULEN, Erik. Regulation Tomorrow: What Happens When Technology is Faster than the Law? *American University Business Law Review*, v. 6, n. 3, p. 24, 2017, tradução nossa. Disponível em: <https://ssrn.com/abstract=2834531>. Acesso em: 09 fev. 2022.)

²²⁹ “The starting point for regulating AI should be a statute that establishes the general principles for AI regulation [...] AIDA should begin, as do most modern statutes, with a statement of purpose. The purpose of AIDA would be to ensure that AI is safe, secure, susceptible to human control, and aligned with human interests.” SCHERER, Matthew U. Regulating Artificial Intelligence Systems. *Harvard Journal of Law & Technology*, v. 29, n. 2, p. 394, 2016. Disponível em: <https://ssrn.com/abstract=2609777>. Acesso em: 09 fev. 2022.

Fabrício Bertini Pasquot Polido, quanto às iniciativas transnacionais sobre inteligência artificial:

No entanto, uma opção pela negativa de não regular, harmonizar, codificar pelos instrumentos normativos tradicionais o campo da IA abre espaço para discussão da base em torno da formulação de presentes e futuros princípios e normas a serem adotados no campo da IA, envolvendo, por exemplo, aspectos de segurança, responsabilidade e ética e que são capturados pelos instrumentos não vinculantes ou normas narrativas (*soft law*)²³⁰.

O Reino Unido é exemplo de país que se debruçou sobre a análise da melhor forma de regular a tecnologia. Em abril de 2018, foi publicado, pelo Comitê de Inteligência Artificial da Authority of the House of Lords, o documento intitulado “AI in UK”, elaborado para direcionar o governo a respeito dos próximos passos em investimento, governança e desenvolvimento da inteligência artificial. Para a sua elaboração, foram consultados vários *players* do Governo, da academia e do setor privado, entre os quais citam-se as ponderações do escritório de advocacia Baker McKenzie sobre a regulação da inteligência artificial, conforme se transcreve abaixo:

Baker McKenzie, an international law firm, recommended a “proactive, principles-led intervention, based on a sound understanding of the issues and technology, careful consideration and planning’ rather than reactive regulation, put in place after something goes wrong. They recommended that “the right regulatory approach... is staged and considered” and the Government should “facilitate ethical (as opposed to legal) frameworks for the development of AI technologies” to support self-regulation in industry²³¹.

Mais adiante no mesmo documento, menciona-se que foram ouvidos muitos especialistas com argumentos contrários à regulação específica para a inteligência artificial. Em entrevista com Chris Reed, professor de Direito do Comércio Eletrônico na Queen Mary University of London, ele afirma que não há uma medida única para todos os casos de inteligência artificial (“one size fits all answer”) e que seria inapropriado, ou

²³⁰ POLIDO, Fabrício Bertini Pasquot. Novas perspectivas para a regulação da Inteligência Artificial: diálogos entre as políticas domésticas e os processos legais transnacionais. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Revista dos Tribunais, 2019, p. 185.

²³¹ “Baker McKenzie, escritório de advocacia internacional, recomendou uma “intervenção proativa e baseada em princípios, com base em uma sólida compreensão dos problemas e da tecnologia, além de planejamento cuidadoso”, em vez de regulamentação reativa, implementada depois que algo der errado. Eles recomendaram que “a abordagem regulatória correta... seja feita em estágios e com moderação” e que o governo deve “facilitar estruturas éticas (em oposição a legais) para o desenvolvimento de tecnologias de IA” para apoiar a autorregulação na indústria.” (REINO UNIDO. *AI in the UK: ready, willing and able?* House of Lords, Select Committee on Artificial Intelligence, p. 113, 16 Apr. 2018. Disponível em: <https://www.parliament.uk/ai-committee/>. Acesso em: 09 fev. 2022.)

mesmo impossível, produzir um regime regulatório que se aplique a todos os tipos de inteligência artificial²³².

Colhidos todos os argumentos pró e contrarregulação, o relatório concluiu que, naquele momento, uma regulação específica para a inteligência artificial cobrindo todo o seu espectro seria inapropriada, considerando que as regulações setoriais estariam mais bem posicionadas para identificar o impacto da inteligência artificial nos seus respectivos mercados²³³.

Dois anos depois da publicação do “AI in UK”, a “National AI Strategy” vem confirmar que o Reino Unido de fato havia, até aquele momento, apostado as suas fichas nessa abordagem regulatória²³⁴. Contudo, com maior conhecimento sobre o funcionamento e os riscos da inteligência artificial e, conseqüentemente, maior maturidade para se identificar como regulá-la, alguns problemas dessa abordagem também foram expostos na estratégia nacional. Esses problemas podem ser resumidos nas possíveis inconsistências e contradições entre as tratativas de casos envolvendo a inteligência artificial em cada setor, o que acabaria por gerar incertezas e insegurança jurídica. Percebeu-se também que iniciativas internacionais multisetoriais começaram a tomar lugar de destaque no cenário global, iniciando-se um diálogo a respeito de parâmetros para a regulação da inteligência artificial e padrões técnicos internacionais, o que exige uma abordagem mais generalista do tema. Dados todos esses aspectos, no próprio documento (“National AI Strategy”) é questionado se a abordagem regulatória exclusivamente setorial até então adotada pelo Reino Unido face à inteligência artificial seria, de fato, adequada e até que ponto regras transversais, ou seja, que abrangeriam vários tipos de aplicabilidade da tecnologia, não seriam benéficas para gerar mais coerência regulatória entre setores específicos:

²³² REINO UNIDO. *AI in the UK: ready, willing and able?* House of Lords, Select Committee on Artificial Intelligence, p. 114, 16 Apr. 2018. Disponível em: <https://www.parliament.uk/ai-committee/>. Acesso em: 09 fev. 2022

²³³ REINO UNIDO. *AI in the UK: ready, willing and able?* House of Lords, Select Committee on Artificial Intelligence, p. 116, 16 Apr. 2018. Disponível em: <https://www.parliament.uk/ai-committee/>. Acesso em: 09 fev. 2022

²³⁴ REINO UNIDO. *National AI Strategy*. Office for Artificial Intelligence, p. 52, Sept. 2021. Disponível em: <https://www.gov.uk/government/publications/national-ai-strategy>. Acesso em: 10 fev. 2022.

These challenges raise the question of whether the UK's current approach is adequate, and whether there is a case for greater cross-cutting AI regulation or greater consistency across regulated sectors²³⁵.

Aventados, então, os benefícios e os desafios de uma regulação estritamente setorial, algumas alternativas são levantadas pelo Conselho de IA do Reino Unido, entre elas: “introducing additional cross-sector principles or rules, specific to AI, to supplement the role of individual regulators to enable more consistency across existing regimes”²³⁶. Ou seja, passa-se a considerar o estabelecimento de princípios norteadores que se sobreporão às regras setoriais.

Tal conclusão parece estar alinhada com a orientação da União Europeia sobre inteligência artificial e responsabilidade civil, em específico, à Resolução 2020/2014 (INL), de outubro de 2020, “que contém recomendações à Comissão sobre o regime de responsabilidade civil aplicável à inteligência artificial”²³⁷. Nessa resolução encontra-se, de forma bem direta, a recomendação de estabelecimento de um regime principiológico de regulação de inteligência artificial sob a perspectiva da responsabilidade civil, a qual vale transcrever na íntegra:

Considerando que a diversidade dos sistemas de IA e a gama diversificada de riscos que a tecnologia representa dificulta os esforços destinados a encontrar uma solução única, adequada a todos os riscos possíveis; considerando que, a esse respeito, deve ser adotada uma abordagem em que são utilizadas experiências, projetos-piloto e ambientes de teste da regulamentação para encontrar soluções proporcionais e baseadas em dados concretos que se adaptem a situações e setores específicos, sempre que necessário²³⁸,

²³⁵ “Esses desafios colocam em xeque a atual abordagem regulatória do Reino Unido e traz a reflexão sobre uma possível regulação transversal de inteligência artificial, ou outro tipo de regulação capaz de trazer maior consistência jurídica entre os setores do mercado em que há aplicação da IA.” (REINO UNIDO. *National AI Strategy*. Office for Artificial Intelligence, p. 53, Sept. 2021, tradução nossa. Disponível em: <https://www.gov.uk/government/publications/national-ai-strategy>. Acesso em: 10 fev. 2022.)

²³⁶ “[...] a introdução de princípios ou regras intersetoriais, específicas para a inteligência artificial, de modo a complementar o papel de reguladores setoriais, permitindo, assim, maior consistência jurídica entre os regimes existentes.” (REINO UNIDO. *National AI Strategy*. Office for Artificial Intelligence, p. 55, Sept. 2021, tradução nossa. Disponível em: <https://www.gov.uk/government/publications/national-ai-strategy>. Acesso em: 10 fev. 2022.)

²³⁷ UNIÃO EUROPEIA. Parlamento Europeu. *Regime de responsabilidade civil aplicável à inteligência artificial*. Resolução do Parlamento Europeu, de 20 de outubro de 2020, que contém recomendações à Comissão sobre o regime de responsabilidade civil aplicável à inteligência artificial (2020/2014 (INL)). Bruxelas, 2017. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_PT.html. Acesso em: 10 fev. 2022.

²³⁸ UNIÃO EUROPEIA. Parlamento Europeu. *Regime de responsabilidade civil aplicável à inteligência artificial*. Resolução do Parlamento Europeu, de 20 de outubro de 2020, que contém recomendações à Comissão sobre o regime de responsabilidade civil aplicável à inteligência artificial (2020/2014 (INL)). Considerando “L”. Bruxelas, 2017. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_PT.html. Acesso em: 10 fev. 2022.

[...] Está firmemente convicto de que, para tirar devidamente proveito das vantagens e prevenir potenciais abusos dos sistemas de IA e evitar a fragmentação regulamentar na União, é fundamental que a União se dote, para todos os sistemas de IA, de uma legislação uniforme, baseada em princípios e virada para o futuro; considera que, embora seja preferível uma regulamentação setorial para uma gama vasta de aplicações possíveis, se afigura necessário um quadro jurídico horizontal e harmonizado baseado em princípios comuns, para garantir segurança jurídica, estabelecer normas iguais em toda a União e proteger eficazmente os nossos valores europeus e os direitos dos cidadãos²³⁹ [...]

Apenas para lembrar mais alguns exemplos, cita-se que a estratégia alemã para a inteligência artificial também menciona que o Governo Federal se apoiará em uma regulação baseada em princípios, que devem harmonizar com as demais regulamentações pela União Europeia, para o futuro de uma inteligência artificial de confiança no país²⁴⁰. Da mesma forma, o Japão também já propôs os seus princípios sobre a IA, afirmando que o Governo reconhece que, para se promover a implementação proativa da tecnologia na sociedade, será importante estabelecer os princípios básicos pelos quais os interessados no desenvolvimento e uso da inteligência artificial devem se fiar²⁴¹.

Por fim, passando das realidades domésticas para o ambiente internacional, em que a aplicação da chamada *soft law* sempre foi mais difundida, observam-se também algumas diretrizes principiológicas sobre inteligência artificial, tais como a *OECD AI principles*²⁴² e o *G20 AI Principles*²⁴³.

²³⁹ UNIÃO EUROPEIA. Parlamento Europeu. *Regime de responsabilidade civil aplicável à inteligência artificial*. Resolução do Parlamento Europeu, de 20 de outubro de 2020, que contém recomendações à Comissão sobre o regime de responsabilidade civil aplicável à inteligência artificial (2020/2014 (INL)). Introdução “2”. Bruxelas, 2017. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_PT.html. Acesso em: 10 fev. 2022.

²⁴⁰ “In particular, the Federal Government believes it makes sense in the interest of principle-based regulation to draft central principles for trustworthy AI which are harmonised across the EU”. / “Em particular, o governo federal acredita fazer sentido elaborar princípios centrais, harmonizados em toda a EU, para o desenvolvimento de sistema de IA confiáveis.” (ALEMANHA. *Artificial Intelligence Strategy of the German Federal Government*. Governo Federal Alemão, p. 23, 2020, tradução nossa. Disponível em: https://knowledge4policy.ec.europa.eu/ai-watch/germany-ai-strategy-report_en. Acesso em: 10 fev. 2022.)

²⁴¹ “We recognize that in order to realize an ‘AI-Ready Society’ and to promote an appropriate and proactive social implementation of AI, it is important to establish basic principles that each stakeholder should keep in mind”. / “Reconhecemos que, para alcançar uma ‘sociedade pronta para a IA’ e promover a implementação social apropriada e proativa da IA, é importante estabelecer princípios básicos que cada parte interessada deve ter em mente.” (JAPÃO. *Social Principles of Human-Centric Artificial Intelligence*. Conselho para Ciência, Tecnologia e Inovação, p. 7, 2019, tradução nossa. Disponível em: <https://www8.cao.go.jp/cstp/english/humancentricai.pdf>. Acesso em: 10 fev. 2022.)

²⁴² OCDE. *Recommendation of the Council on Artificial Intelligence*. Secretary-General of the OECD, 2019. Disponível em: <https://oecd.ai/en/ai-principles>. Acesso em: 10 fev. 2022.

²⁴³ CASALONE, Carlo *et al.* *Human-centric AI: from principles to actionable and shared policies*. G20 AI Principles. Sept. 30, 2021. Disponível em: https://www.g20-insights.org/policy_briefs/human-centric-ai-from-principles-to-actionable-and-shared-policies/. Acesso em: 10 fev. 2022.

Assim, tem-se que a literatura especializada e as iniciativas governamentais têm autorizado e, até mesmo, incentivado a abordagem regulatória principiológica da inteligência artificial. Questiona-se, pois, quais princípios têm sido estabelecidos e quais são recomendáveis visando o direcionamento legal dos casos envolvendo danos causados por sistemas de inteligência artificial. Para responder a tal questionamento, optou-se pelo estudo de casos, pinçando as principais recomendações governamentais já existentes e os princípios que elas dispõem, de modo a limitar o espectro de análise de forma eficiente e embasada em discussões já em voga mundialmente. Após a coleta de dados dessas diretrizes governamentais, analisar-se-á como os princípios se relacionam com a responsabilidade civil e a (in)conveniência da sua adoção.

3 ESTUDO DE CASO: REGULAMENTOS DOS ESTADOS UNIDOS, DA CHINA E DO REINO UNIDO

Neste tópico, buscou-se estabelecer uma metodologia para pesquisar quais os princípios gerais sobre o tema responsabilidade civil e inteligência artificial mais têm sido valorizados e aplicados globalmente, para que, então, se passasse à análise dos seus prós e contras, no intuito de colaborar com um arcabouço nacional de governança da inteligência artificial.

Essa metodologia consiste em (i) pesquisar os países mais avançados em termos de estudo e emprego da inteligência artificial; (ii) identificar a existência de normas (vinculantes ou não) editadas pelos respectivos Estados que lideram o ranking de desenvolvimento da tecnologia; (iii) identificar se entre essas normas, caso existentes, se poderia identificar princípios comuns entre elas que têm reflexos na matéria de responsabilidade civil²⁴⁴.

Assim, inicia-se pela pesquisa dos países líderes em desenvolvimento da inteligência artificial. Para viabilizar a pesquisa dentro dos limites impostos por esta espécie de trabalho, concentrar-se-á no tríplice pódio dos países que estão à frente em termos de desenvolvimento e emprego da inteligência artificial e, na sequência, será feita a análise do cenário brasileiro.

O *Global AI Index* editado pela Tortoise Media, em 2021, pesquisou 62 países em relação ao desenvolvimento da inteligência artificial sob três temas: investimento, inovação e implementação. Desses pilares, foram avaliados uma outra camada de categorias, dividida em talento, infraestrutura, operação, ambiente, pesquisa, desenvolvimento, estratégia governamental e comercial, analisando-se, ao todo, 143 indicadores nessas categorias. A conclusão da pesquisa foi que os Estados Unidos da América lideram o ranking mundial de inteligência artificial, seguidos da China, que, por sua vez, é seguida pelo Reino Unido²⁴⁵.

Do ponto de vista do emprego e do desenvolvimento da tecnologia no âmbito privado (em companhias privadas), a pesquisa realizada pela Deloitte, em maio de 2019,

²⁴⁴ Importante destacar que cada um dos princípios escolhidos para análise poderia ser objeto de trabalhos individuais e de fôlego. O que se pretende, neste estudo, é apenas um voo panorâmico para levar luz à exploração da ideia de uma regulação principiológica genérica no Brasil.

²⁴⁵ Tortoise Media. *The Global AI Index*. 2021. Disponível em: <https://www.tortoisemedia.com/intelligence/global-ai/>. Acesso em: 10 fev. 2022.

aponta para o mesmo sentido. A pesquisa indica que, entre sete países, os Estados Unidos, a China e o Reino Unido são os mais maduros na compreensão da importância de uma estratégia articulada de implementação de inteligência artificial nas companhias²⁴⁶.

Por sua vez, a plataforma de conteúdo Analytics Insight, que coleta dados sobre inteligência artificial de fontes oficiais em todo o mundo, também listou, em janeiro de 2021, os 10 países que mais investem em inteligência artificial, com destaque para a China, os Estados Unidos e o Reino Unido nos três primeiros lugares, sem fazer distinção de qual estaria mais avançado que o outro²⁴⁷.

Além das mencionadas pesquisas, vários outros artigos e reportagens apontam essas três potências mundiais como líderes em inteligência artificial²⁴⁸. Diante disso, apesar de o nível de desenvolvimento da inteligência artificial em cada país poder ser um dado relativo, com algumas subjetividades na pesquisa, parece haver consenso de que os três países citados acima lideram esse ranking sob diferentes aspectos.

Com esses dados em mãos, passa-se a avaliar se essas nações possuem regulamentos sobre a inteligência artificial que trazem princípios capazes de lançar luz aos problemas de responsabilidade civil e IA. Nesse ponto, é necessário mencionar que cada um desses países tem seu próprio escopo legislativo de responsabilidade civil, que não será mencionado neste estudo, exatamente porque o que se deseja é encontrar princípios gerais que sejam aplicáveis ao instituto da responsabilidade civil em relação à inteligência artificial, independente da sua lógica legal.

²⁴⁶ JARVIS, David; LOUCKS, Jeff; HUPFER, Susanne; MURPHY, Tomothy. *Future in the balance? How countries are pursuing an AI advantage*. Deloitte Insights, 2019. Disponível em: <https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/ai-investment-by-country.html>. Acesso em: 10 fev. 2022.

²⁴⁷ ANALYTICS INSIGHT. *Artificial Intelligence Investment by Top 10 Countries*. 2021. Disponível em: <https://www.analyticsinsight.net/artificial-intelligence-investment-by-top-10-countries/>. Acesso em: 10 fev. 2022.

²⁴⁸ Alguns exemplos de outras publicações que mencionam que os Estados Unidos, a China e o Reino Unido figuram como os principais líderes em IA globalmente são: GOPANI, Avi. *Meet The New Player In The US-China AI Arms Race: United Kingdom*. *Analytics India Magazine*, 2021. Disponível em: <https://analyticsindiamag.com/meet-the-new-player-in-the-us-china-ai-arms-race-united-kingdom/>. Acesso em: 10 fev. 2022. / NATURE INDEX. *Top 25 countries/territories in artificial intelligence*. Nature Index 2020 Artificial Intelligence. 2020. Disponível em: <https://www.natureindex.com/supplements/nature-index-2020-ai/tables/countries>. Acesso em: 10 fev. 2022.

3.1 ESTADOS UNIDOS DA AMÉRICA

Apesar de toda a tensão a respeito da alavancagem chinesa nos investimentos em inteligência artificial, de acordo com as pesquisas analisadas neste estudo, os EUA continuam sendo os líderes mundiais em inteligência artificial. Segundo pesquisa realizada pela Data Innovation, que compara os Estados Unidos da América e a China em termos de desenvolvimento da tecnologia, os EUA largam à frente por possuírem mais profissionais capacitados, mais recursos em pesquisa e desenvolvimento e uma maior tecnologia de hardwares preparados para o uso de inteligência artificial, quando contrapostos com a China, ou com qualquer país no mundo²⁴⁹.

A importância estratégica da inteligência artificial para a nação se traduz em altos investimentos no desenvolvimento e no emprego da tecnologia atualmente. Entre 2018 e 2020, as agências governamentais dos EUA gastaram um total de US\$ 1,9 bilhão em serviços relacionados à inteligência artificial, representando um aumento de investimento de 70% no período de três anos. Os gastos com pesquisa e desenvolvimento alcançaram o patamar de US\$ 1,2 bilhão no mesmo período – de 2018 a 2020. Para os próximos anos, o governo de Joe Biden se propôs a gastar 2% do PIB americano de 2022 em ciência e tecnologia, sendo que grande fatia será dedicada à inteligência artificial²⁵⁰.

Somando-se a importância estratégica da inteligência artificial para o futuro do país à maior competitividade de países como a China, os últimos anos do governo americano foram marcados pela edição de várias Ordens Executivas e diretrizes relacionadas à inteligência artificial. Ainda em 2016, foi publicado o Relatório Americano para a Inteligência Artificial, denominado “Preparing for the Future of Artificial Intelligence”, considerando como o primeiro documento nacional de maior relevância sobre o tema. Conforme exposto por Corinne Cath *et al.*, o relatório tem um tom confiante e reflete uma visão positiva da tecnologia, advinda do Vale do Silício²⁵¹. Em vista disso, o relatório traz recomendações sobre o desenvolvimento da inteligência

²⁴⁹ CASTRO, Daniel; CHIVOT, Eline; MCLAUGHLIN, Michael. *Who is Winning the AI Race: China, EU or the United States?* Center for Data Innovation. 2019. Disponível em: <https://datainnovation.org/2019/08/who-is-winning-the-ai-race-china-the-eu-or-the-united-states/>. Acesso em: 10 fev. 2022.

²⁵⁰ DELTEK. *Federal Artificial Intelligence Landscape*. GovWin IQ. 2022. Disponível em: <https://info.deltek.com/Federal-Artificial-Intelligence-Landscape-2022-Summary-GovWin-Deltek>. Acesso em: 10 fev. 2022.

²⁵¹ CATH, Corinne *et al.* *Artificial Intelligence and the “Good Society”*: the US, EU, and UK approach. *Sci Eng Ethics*, v. 24, p. 505–528, 2018, p. 509.

artificial no país, considerando a tecnologia como um dos grandes propulsores econômicos dos Estados Unidos e do bem-estar coletivo. Algumas dessas recomendações são (i) investir em pesquisa em desenvolvimento na área; (ii) servir como “primeiros clientes” (“early customers”) dos novos produtos de inteligência artificial; (iv) apoiar projetos-piloto e testes no ambiente real, ou seja, fora de laboratórios; (v) disponibilizar bancos de dados ao público para propulsão a pesquisa e desenvolvimento em inteligência artificial. Tais recomendações deixam clara a abertura do país para o experimentalismo tecnológico, sendo reforçada pela postura menos intervencionista em termos de regulamentação da atividade, que é também acentuada nesse relatório²⁵².

Não obstante isso, sob a justificativa de que a falta de segurança e de confiança em um sistema de inteligência artificial seria o principal motivo de hesitação do seu uso fora dos laboratórios, o relatório contém um capítulo dedicado à justiça, à segurança e à governança algorítmica, lançando mão da transparência, da prestação de contas e da possibilidade de interpretação (*explainability*) como aspectos fundamentais para a disseminação, em ainda maior escala, da inteligência artificial em suas aplicações mais críticas do ponto de vista de segurança e de justiça.

O relatório cita que “as preocupações com a transparência se concentraram não apenas nos dados e códigos usados, mas também na necessidade de haver alguma forma de explicação para qualquer determinação baseada em IA” (tradução)²⁵³. Quanto à prestação de contas (*accountability*)²⁵⁴, são mencionadas algumas possíveis técnicas já

²⁵² “The general consensus of the RFI commenters was that broad regulation of AI research or practice would be inadvisable at this time. Instead, commenters said that the goals and structure of existing regulations were sufficient, and commenters called for existing regulation to be adapted as necessary to account for the effects of AI”. / “O consenso entre os comentadores é que uma regulação ampla da pesquisa e aplicação da inteligência artificial neste momento seria desaconselhável. Ao contrário, eles entendem que os objetivos e a estrutura da regulação atualmente existentes são suficientes e que, quando necessário, devem ser adaptadas para responderem aos efeitos da IA.” (ESTADOS UNIDOS DA AMÉRICA. *Preparing for the Future of Artificial Intelligence*. National Science and Technology Council. p. 17, 2016, tradução nossa. Disponível em: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf. Acesso em: 10 fev. 2022.)

²⁵³ “Transparency concerns focused not only on the data and algorithms used, but also on the potential to have some form of explanation for any AI-based determination”. (ESTADOS UNIDOS DA AMÉRICA. *Preparing for the Future of Artificial Intelligence*. National Science and Technology Council. p. 31, 2016, tradução nossa. Disponível em: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf. Acesso em: 10 fev. 2022.)

²⁵⁴ No Brasil, o termo “accountability” vem sendo entendido como a obrigatoriedade de se prestar contas a respeito de um determinado resultado. Muitas vezes, o termo é traduzido como “prestação de contas” e/ou “responsabilidade”. Neste trabalho, valer-se-á dessa tradução, mas também será admitido o uso da palavra “accountability” na tentativa de que a análise sobre os princípios trazidos nos regulamentos estrangeiros seja a mais fiel possível, considerando a sua redação original. Mais adiante, será realizada uma análise mais aprofundada do sentido do termo.

disponíveis para alcançá-la, tais como a testagem do sistema ou a publicação de informações sobre o algoritmo e/ou seus outputs:

There are several technical approaches to enhancing the accountability and robustness of complex algorithmic decisions. A system can be tested “in the wild” by presenting it with situations and observing its behavior. A system can be subjected to black-box testing, in which it is presented with synthetic inputs and its behavior is observed, enabling behavior to be tested in scenarios that might not occur naturally. Some or all of the technical details of a system’s design can be published, enabling analysts to replicate it and analyze aspects of its internal behavior that might be difficult to characterize with testing alone. In some cases it is possible to publish information that helps the public evaluate a system’s risk of bias, while withholding other information about the system as proprietary or private²⁵⁵.

Por fim, concluindo a análise sobre os princípios, são expostas as recomendações governamentais em suporte a eles, que passam por tornar obrigatórias disciplinas de segurança em inteligência artificial para os estudantes de Ciência da Computação e, no campo profissional, por incentivar o amadurecimento da segurança algorítmica por meio do investimento em novas técnicas de codificação. Além disso, há orientação explícita de que todas as agências federais que utilizam sistemas baseados em inteligência artificial para a tomada de decisões de impacto devem ter mecanismos para assegurar a prestação de contas e a transparência, fundadas em testes e evidências fáticas²⁵⁶.

Demonstra-se, assim, que, para estimular o desenvolvimento da inteligência artificial e a sua adoção em larga escala, compreendeu-se ser necessário dispor de

²⁵⁵ “Existem várias abordagens técnicas para melhorar a prestação de contas e a robustez de decisões algorítmicas complexas. Um sistema pode ser testado fora de laboratórios, sendo-lhe apresentadas novas situações, em que se observará seu comportamento. Um sistema pode ser submetido a testes de caixa-preta, por meio dos quais são apresentados *inputs* sintéticos e seu comportamento é observado, permitindo que o comportamento seja testado em cenários que podem não ocorrer naturalmente. Alternativamente, alguns ou todos os detalhes técnicos do design de um sistema podem ser publicados, permitindo que experts o repliquem e analisem aspectos de seu comportamento interno que podem ser de difícil caracterização apenas com testes. Em outros casos, é possível publicar informações que ajudem o público a avaliar o risco de possíveis vieses de um sistema, mantendo-se sob sigilo informações protegidas por direitos de propriedade intelectual.” (ESTADOS UNIDOS DA AMÉRICA. *Preparing for the Future of Artificial Intelligence*. National Science and Technology Council, p. 32, 2016, tradução nossa. Disponível em:

https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf. Acesso em: 10 fev. 2022.)

²⁵⁶ ESTADOS UNIDOS DA AMÉRICA. *Preparing for the Future of Artificial Intelligence*. National Science and Technology Council, p. 34, 2016. Disponível em: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf. Acesso em: 10 fev. 2022.

algumas balizas para garantir padrões mínimos de segurança, justiça e controle da tecnologia, o que poderá ser impulsionado por sua transparência e explicabilidade²⁵⁷.

Seguindo essa direção, mais recentemente, já na administração Trump, em 2019, a publicação denominada “National Artificial Intelligence Research and Development Strategic Plan”²⁵⁸ deu grande notoriedade à necessidade de se aprimorar a transparência, a explicabilidade e a *accountability by design* (ou seja, a prestação de contas intrínseca ao próprio design) dos sistemas de inteligência artificial. Recomendou-se que pesquisadores se debrucem sobre as técnicas de engenharia de softwares interpretáveis, de modo que possa ser verificado, inclusive por não experts, se uma determinada decisão automatizada foi justa ou não²⁵⁹⁻²⁶⁰. Para isso, são previstos maiores investimentos em pesquisa e desenvolvimento para produção de sistemas de inteligência artificial que incorporem parâmetros éticos e legais, principalmente por meio de mecanismos de transparência e explicabilidade. Ainda segundo o Plano Estratégico, uma vez que esses parâmetros éticos e legais podem se alterar dependendo do contexto ou da aplicação do sistema, o desenvolvimento de pesquisa na área dependerá não somente do apoio de

²⁵⁷ Neste trabalho permitir-se-á o emprego dos neologismos “explicabilidade” e “interpretabilidade” como tradução de *explainability* e *interpretability*, respectivamente. Entende-se que, nesse caso, lançar mão do neologismo é adequado para se exprimir, de forma mais direta, a ideia que se deseja, referente à qualidade de o sistema de IA ser explicável ou interpretável.

²⁵⁸ Lembrando que este é o segundo plano estratégico para a IA editado pelo governo americano, sendo o primeiro de 2016, o que não foi citado justamente por já haver um mais recente, o de 2019.

²⁵⁹ “Researchers must learn how to design these systems so that their actions and decision-making are transparent and easily interpretable by humans, and thus can be examined for any bias they may contain, rather than just learning and repeating these biases. There are serious intellectual issues about how to represent and ‘encode’ value and belief systems. Scientists must also study to what extent justice and fairness considerations can be designed into the system, and how to accomplish this within the bounds of current engineering techniques [...] Thus, researchers must develop systems that are transparent, and intrinsically capable of explaining the reasons for their results to users”. “Pesquisadores devem aprender como projetar esses sistemas, para que suas ações e tomadas de decisão sejam transparentes e facilmente interpretáveis por humanos e, portanto, que eventuais vieses apresentados pelo sistema possam ser examinados. Existem sérias questões intelectuais sobre como representar e ‘codificar’ valores em sistemas. Os cientistas também devem estudar até que ponto as considerações de justiça e equidade podem ser projetadas no sistema e como fazer isso dentro dos limites das técnicas atuais de engenharia [...] Assim, os pesquisadores devem desenvolver sistemas que sejam transparentes e intrinsecamente capazes de explicar os motivos de seus resultados aos usuários.” (ESTADOS UNIDOS DA AMÉRICA. *National Artificial Intelligence Research and Development Strategic Plan*. National Science and Technology Council. p. 21, 2019, tradução nossa. Disponível em: <https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf>. Acesso em: 10 fev. 2022.)

²⁶⁰ “Apesar do apelo pela prestação de contas *by design*, no próprio documento se reconhece a existência do *trade-off* entre transparência (ou possibilidade de interpretação) e a precisão do sistema: ‘Algumas técnicas de IA, como a árvore de decisão, fornecem explicações intrínsecas, mas que, geralmente, são menos precisas’.” (ESTADOS UNIDOS DA AMÉRICA. *National Artificial Intelligence Research and Development Strategic Plan*. National Science and Technology Council. p. 25, 2019, tradução nossa. Disponível em: <https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf>. Acesso em: 10 fev. 2022.)

profissionais do Direito, da Filosofia e das Ciências Sociais, mas também de especialistas em cada área de mercado, para melhor adequação do sistema ao respectivo setor²⁶¹.

Mais adiante, o documento “Guidance for Regulation of Artificial Intelligence Applications”, já mencionado neste trabalho, que discorre sobre a estratégia para regular a inteligência artificial, menciona que qualquer abordagem regulatória que seja tomada deve se basear em alguns princípios, sendo dois deles o de transparência e o de divulgação. Segundo o guia, esses pré-requisitos ampliam a confiança nos sistemas de inteligência artificial, permitindo que leigos no assunto compreendam como a aplicação de IA funciona e como chegou a determinado resultado²⁶². Ademais, de acordo com o documento, o nível de transparência considerado adequado para cada aplicação da inteligência artificial é específico para cada contexto, devendo-se levar em conta os riscos impostos pelo sistema²⁶³.

Apenas para se corroborar a importância dos princípios já mencionados acima, ou seja, a transparência e a divulgação, a explicabilidade, a controlabilidade e *accountability*, citam-se brevemente outras medidas editadas pelo governo americano que também dispõem dos seus conjuntos de ditames para o uso seguro da inteligência artificial. A Ordem Executiva 13960 – “Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government” – demonstra a preocupação governamental de que o público tenha confiança na adoção da inteligência artificial e, para tanto, informa que esta deve ser “preciso, confiável e eficaz; seguro e resiliente; compreensível, responsável e rastreável; transparente e passível de prestação de contas”²⁶⁴. Por sua vez, o Department of Defense publicou o “AI Ethical Principals” relacionados ao uso militar da inteligência artificial, que também está alinhado com as demais medidas governamentais, ao prever que a inteligência artificial deve ser confiável e, portanto, controlável, responsável e rastreável. O National Institute of Standards and Technology – NIST vinculado ao U.S. Department of Commerce também editou documento enunciando quatro princípios para

²⁶¹ ESTADOS UNIDOS DA AMÉRICA. *National Artificial Intelligence Research and Development Strategic Plan*. National Science and Technology Council. p. 20, 2019. Disponível em: <https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf>. Acesso em: 10 fev. 2022.

²⁶² ESTADOS UNIDOS DA AMÉRICA. *Guidance for Regulation of Artificial Intelligence Applications*. Memorandum for the Heads of Executive Departments and Agencies. p. 6, 2020. Disponível em: <https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf>. Acesso em: 10 fev. 2022.

²⁶³ ESTADOS UNIDOS DA AMÉRICA. *Guidance for Regulation of Artificial Intelligence Applications*. Memorandum for the Heads of Executive Departments and Agencies. p. 6, 2020. Disponível em: <https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf>. Acesso em: 10 fev. 2022.

²⁶⁴ “Accurate, reliable, and effective; safe, secure, and resiliente; understandable; responsible and traceable; transparent; and accountable.” (ESTADOS UNIDOS DA AMÉRICA. *Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government*, Seção 3, 2020.)

a explicabilidade da inteligência artificial²⁶⁵. Por fim, em trâmite desde 2019, cita-se o “Algorithmic Accountability Act”, que prevê a necessidade de instituições envolvidas no desenvolvimento de sistemas que tomam decisões automatizadas de alto risco disponibilizarem ao público relatórios de impacto²⁶⁶.

O que se conclui, então, é que apesar do histórico de resistência regulatória dos Estados Unidos, visando não impor maiores barreiras ao desenvolvimento (o que, de alguma forma, se repete no caso da inteligência artificial), nota-se que há uma consciência sobre a necessidade de adoção de regras mínimas para que a tecnologia seja utilizada em larga escala com maior confiança dos desenvolvedores e dos usuários da inteligência artificial. Entre essas regras, há claro destaque à transparência, à explicabilidade, à divulgação, à *accountability* e à rastreabilidade.

3.2 CHINA

Para se iniciar uma breve exposição a respeito do contexto chinês em torno da inteligência artificial, lembra-se do que ficou conhecido como o “momento Spunik da China”, quando o aplicativo AlphaGo, da empresa britânica DeepMind (adquirida pelo Google), venceu, em 2017, o jovem chinês Ke Jie, campeão de Go, nesse jogo de importância histórica para a sociedade chinesa. Para se compreender a relevância do momento para a China, vale-se da narração de Kai-Fu Lee sobre o acontecimento:

O campo de batalha era um tabuleiro de dezenove por dezenove linhas, cheio de pequenas pedras pretas e brancas – as matérias-primas do complexo jogo Go. Durante a partida, dois jogadores alternam-se, colocando pedras no tabuleiro e tentando cercar as peças do oponente. Nenhum humano na Terra poderia fazer isso melhor do que Ke Jie, mas, naquele momento, ele enfrentava um jogador de Go que estava em um nível que ninguém jamais havia visto. Inventado, acredita-se, há mais de 2.500 anos, o Go tem a história mais longa que qualquer jogo de tabuleiro jogado ainda hoje. Na antiga China, representava uma das quatro formas de arte que todo acadêmico chinês deveria dominar. Acredita-se que o jogo levava seus jogadores a um refinamento e a uma sabedoria intelectual. Enquanto jogos como o xadrez ocidental eram considerados grosseiramente táticos, o Go é baseado no posicionamento paciente e no lento cerco, o que o transformou em uma forma de arte, um

²⁶⁵ ESTADOS UNIDOS DA AMÉRICA. *Four Principles of Explainable Artificial Intelligence*. Instituto Nacional de Normas e Tecnologia, Departamento de Comércio do Governo Federal, 2020. Disponível em: <https://www.nist.gov/publications/four-principles-explainable-artificial-intelligence-draft>. Acesso em: 10 fev. 2022.

²⁶⁶ ESTADOS UNIDOS DA AMÉRICA. *House Resolution 2231*, de 10 de abril de 2019. House of Representatives. Algorithmic Accountability Act. Disponível em: <https://www.congress.gov/bill/116th-congress/housebill/2231>. Acesso em: 10 fev. 2022.

estado de espírito. [...] O AlphaGo não deu nenhuma abertura a Ke. Ao contrário, lentamente foi deixando-o sem saída²⁶⁷.

Esse momento, que foi visto com entusiasmo pelos espectadores ocidentais, teve um impacto ainda maior na sociedade chinesa. O chamado momento Sputnik chinês, fez com que, ao mesmo tempo, nascesse uma ansiedade no Governo chinês sobre a suposta superioridade tecnológica americana (*mutatis mutandi* o sentimento americano em relação à superioridade soviética quando do lançamento do foguete) e também gerasse um grande incentivo para o desenvolvimento da inteligência artificial que a China observou desde então.

No mesmo ano da derrota de Ke Jie, o presidente chinês, Xi Jinping, reiterou em pronunciamento público o seu sonho de que a China se tornasse uma grande potência da ciência e tecnologia, e é claro que a inteligência artificial tinha singular relevância na conquista desse objetivo²⁶⁸.

Aliado ao interesse governamental de promover o avanço em inteligência artificial, a conjuntura econômica e social atual da China também a conduz para ganhar ainda mais espaço entre as lideranças globais em inteligência artificial. Uma pesquisa recente da PricewaterhouseCoopers – PwC demonstra que a China, quando comparada com todos os continentes do mundo, promete apresentar maiores ganhos econômicos decorrentes do desenvolvimento doméstico da inteligência artificial. Isso se dá pelo fato de o PIB chinês ser majoritariamente composto por manufaturas cuja produção e venda podem ser potencializadas pela introdução de novas tecnologias, o que levaria, não só à maior eficiência da cadeia produtiva, mas também ao desenvolvimento de produtos mais sofisticados. Além disso, a China é um país que tem alta taxa de reinvestimento de capital na economia, principalmente no que diz respeito ao investimento em tecnologia, o que é demonstrado, por exemplo, pelo aumento no depósito de patentes chinesas (atrás apenas dos Estados Unidos)²⁶⁹. De fato, observando-se alguns dados, percebe-se que a China está buscando aumentar seu potencial econômico baseado no investimento na inteligência

²⁶⁷ LEE, Kai-Fu. *Inteligência Artificial: como os robôs estão mudando o mundo, a forma como amamos, nos relacionamos, trabalhamos e vivemos*. Trad. Marcelo Barbão. Rio de Janeiro: Globo Livros, 2019, p. 13-14.

²⁶⁸ DING, Jeffrey. *Deciphering China's AI Dream: The context, components, capabilities, and consequences of China's strategy to lead the world in AI*. Future of Humanity Institute: Universidade de Oxford, 2018, p. 7.

²⁶⁹ PRICEWATERHOUSECOOPERS GLOBAL. *Sizing the price: What's the real value of AI for your business and how can you capitalise?* p. 8, 2017. Disponível em: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. Acesso em: 10 fev. 2021.

artificial de forma concreta e proativa: desde 2012 houve um aumento de 500% nas instalações robóticas pelo país²⁷⁰; *startups* têm recebido benefícios fiscais e apoio financeiro do Governo, além da grande aposta nas *tech giants* chinesas, Baidu, Tencent e Alibaba²⁷¹.

Diante desse contexto, nos últimos anos houve um esforço regulatório do país, visando o melhor planejamento e organização para o futuro tecnológico chinês. Entre as publicações mais relevantes citam-se o “Internet +”, conjunto de diretrizes para integração e disseminação da internet em todos os elementos da economia e sociedade, que traz também a importância de se cultivar empresas no ramo de inteligência artificial e de se investir em pesquisa e desenvolvimento na área; o *Made in China 2025*, que é uma estratégia de 10 anos para que a China se torne um líder global em manufaturas de alta tecnologia, o que inclui o incentivo à inteligência artificial. Por fim, o *Communist Party of China’s 13th 5-year plan*, que elencou a inteligência artificial como um dos seis pilares para o desenvolvimento da indústria chinesa²⁷². Apesar de relevantes para desenhar o cenário regulatório de inteligência artificial na China, tal tecnologia era tratada em conjunto com outras no planejamento estratégico econômico do país. Apenas em 2017, foi editado o “New Generation of Artificial Intelligence Development Plan” (AIDP, como ficou conhecido no ocidente), publicação que dispõe sobre os objetivos da China em curto, médio e longo prazo, para que até 2030 o país seja o maior centro de inovação em inteligência artificial do mundo:

By 2030, we shall make artificial intelligence theory, technology and application achieve the world's leading level to be the major artificial intelligence innovation center of the world, intelligent economy, intelligent society achieve remarkable results, and lay an important foundation for

²⁷⁰ ROBERTS, Huw *et al.* The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. *AI & Society Review*, p. 64, 2 Mar. 2021. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3469784. Acesso em: 10 fev. 2022.

²⁷¹ “Young companies also receive tax breaks, government contracts and offices in AI clusters if they wish. In parallel, China’s government is working closely with established digital companies such as Baidu, Alibaba and Tencent. The strategic orientations are accompanied by the central government, the collection and exchange of data within the companies is made possible”. / “Empresas jovens também recebem incentivos fiscais, contratos governamentais e, caso assim desejem, escritórios em clusters de IA. Paralelamente, o governo da China está trabalhando em estreita colaboração com empresas digitais já estabelecidas, como Baidu, Alibaba e Tencent. As orientações estratégicas são acompanhadas pelo governo central, permitindo, inclusive, uma maior coleta e troca de dados entre as empresas.” (WESTRHEIDE, Fabian. China: The First Artificial Intelligence Superpower. *Forbes*, 2020, tradução nossa. Disponível em: <https://www.forbes.com/sites/cognitiveworld/2020/01/14/china-artificial-intelligence-superpower/?sh=7f4050152f05>. Acesso em: 10 fev. 2022.)

²⁷² ROBERTS, Huw *et al.* The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. *AI & Society Review*, p. 60, 2 Mar. 2021. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3469784. Acesso em: 10 fev. 2022.

China's entry into the forefront of the innovative countries and economic powers²⁷³.

Para o alcance dessa meta, os reflexos jurídicos da inteligência artificial não foram deixados para trás, tendo um capítulo no AIDP dedicado às diretrizes para construção de um arcabouço normativo capaz de mitigar os riscos trazidos pela nova tecnologia.

Entre as diretrizes regulatórias mencionadas no AIDP, as mais relevantes para este estudo, por apresentam impactos em responsabilidade civil, dizem respeito à necessidade de estabelecimento de sistemas rastreáveis, por meio da transparência e da explicabilidade, que permitam a *accountability*. Esses princípios são delineados de forma objetiva no documento, sem maior desenvolvimento de como eles serão alcançados. Lê-se:

Adhere to problem-specific and accurate interpretations of technology and industry trends. Enhance risk awareness; pay attention to risk assessment, prevention and control. Strengthen the prospective prevention and restraint guidance, the recent focus on the impact of employment, long-term focus on the impact of social ethics, to ensure that the development of AI is safe. Establish and improve the open and transparent AI supervision system. Implement design accountability and application of the supervision of the twotier regulatory structure. Achieve AI algorithm design, monitor product development, application of results, and the rest of the process [...] Around the AI design, product and system complexity, risk, uncertainty, interpretability, potential economic impact and other issues, develop systematic testing methods and indicators system. Construct cross-domain AI test platforms to promote AI safety certification, assessment of AI products and systems of key performance²⁷⁴.

Finalizando a análise sobre os regulamentos chineses sobre inteligência artificial, menciona-se, ainda, o “Artificial Intelligence Standardization White Paper”, publicado em 2018, pelo Chinese Electronics Standards Institute. O documento revela o esforço

²⁷³ CHINA. *New Generation of Artificial Intelligence Development Plan*. Conselho do Estado Chinês. p. 6, 2017. Disponível em: <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/>. Acesso em: 10 fev. 2022.

²⁷⁴ “Aderir à abordagem de interpretações precisas e específicas de problemas da tecnologia e às tendências do setor. Aumentar a consciência de risco; prestar atenção à avaliação, à prevenção e ao controle de riscos. Reforçar a prevenção e a contenção relacionadas ao impacto no trabalho e na mão de obra e em questões éticas e sociais, visando garantir o desenvolvimento seguro da IA. Estabelecer e incentivar a transparência dos sistemas. Implementar a prestação de contas *by design* e a supervisão de dois níveis. Monitorar o desenvolvimento dos produtos e seus resultados, além de todo o processo de emprego da IA [...] Desenvolver testagem sistemática em relação ao design do algoritmo, seus riscos, incertezas, interpretabilidade, potencial econômico e impacto social. Constituir testagens que se aplicam a várias aplicabilidades da IA, a fim de promover segurança, possibilidade de certificações e relatórios de riscos dos produtos de IA.” (CHINA. *New Generation of Artificial Intelligence Development Plan*. Conselho do Estado Chinês. p. 6, 2017, tradução nossa. Disponível em: <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/>. Acesso em: 10 fev. 2022.)

chinês em adotar padrões éticos e técnicos nacionais em inteligência artificial em todas as suas áreas de aplicação e, também, em fazer frente às iniciativas internacionais de estabelecimento de parâmetros de segurança e confiabilidade dos sistemas de IA. Segundo disposto no próprio documento, a sua importância se dá à medida que compartilha resultados de pesquisas e experiências práticas no campo da inteligência artificial com os envolvidos no setor. Além disso, o instrumento convocar diferentes segmentos da sociedade para fortalecer conjuntamente a pesquisa e o desenvolvimento de tecnologias baseadas em inteligência artificial²⁷⁵.

Entre os parâmetros éticos e legais para a inteligência artificial na China, encontram-se, mais uma vez, a prestação de contas e a transparência. Segundo o documento, a primeira será importante para o estabelecimento mais claro do racional de responsabilização e do sistema de compensação em casos de danos gerados por inteligência artificial. A transparência, por sua vez, permitirá que se compreenda o funcionamento subjacente de um sistema de inteligência artificial, o que é igualmente indispensável para aferição de responsabilidade. De acordo com o “White Paper”:

A second principle is that of accountability, that is, establishing a clearly defined accountability system for both the technology development and application aspects. This will facilitate, at the technical level, holding the personnel or departments that develop AI technology accountable, and at the application level, will facilitate the establishment of a rational liability and compensation system. Under the accountability principle, it is necessary to adhere to the principle of transparency in technology development aspects; [...] The principle of transparency, in particular, requires understanding a system's working principles and thereby predicting future development; that is, humans should know how and why AI makes specific decisions. This is critical for assigning responsibility²⁷⁶.

²⁷⁵ CHINA. *Artificial Intelligence Standardization White Paper*. Instituto de Padronização Eletrônica da China. Trad. Etcetera Language Group, Inc. p. 2, 2018. Disponível em: <https://cset.georgetown.edu/publication/artificial-intelligence-standardization-white-paper/>. Acesso em: 10 fev. 2022.

²⁷⁶ “Um segundo princípio é o da prestação de contas, ou seja, é necessário o estabelecimento de um sistema de prestação de contas claramente definido tanto para o desenvolvimento da tecnologia quanto para os aspectos de sua aplicação. Isso facilitará, em nível técnico, identificar a responsabilidade daqueles envolvidos no desenvolvimento da tecnologia de IA e facilitará, também, o estabelecimento de um sistema racional de responsabilidade e compensação. Sob o princípio da prestação de contas, é necessário aderir ao princípio da transparência nos aspectos de desenvolvimento de tecnologia [...] O princípio da transparência, em particular, requer a compreensão dos princípios de funcionamento de um sistema; ou seja, os humanos devem saber como e por que a IA toma certas decisões, o que é fundamental para a atribuição de responsabilidades.” (CHINA. *Artificial Intelligence Standardization White Paper*. Instituto de Padronização Eletrônica da China. Trad. Etcetera Language Group, Inc. p. 32, 2018, tradução nossa. Disponível em: <https://cset.georgetown.edu/publication/artificial-intelligence-standardization-white-paper/>. Acesso em: 10 fev. 2022.)

Observa-se, assim, que a transparência e a prestação de contas, princípios que levam à possibilidade de estabelecimento de regras de responsabilidade e compensação, são levados em conta para a parametrização do uso da inteligência artificial. Em suma, também na China, assim como nos Estados Unidos, observa-se uma preocupação regulatória em relação à prestação de contas (*accountability*) e à necessidade de o sistema ser rastreável, o que se traduz, em termos técnicos, em transparência, explicabilidade e métodos acessórios como relatórios de risco e divulgação de informação aos usuários, entre outros.

3.3 REINO UNIDO

Diante do embate, alimentado pela mídia, entre os EUA e a China, para a liderança mundial em desenvolvimento da inteligência artificial, o Reino Unido vem preparando a sua conjuntura nacional para fazer frente às duas potências, demonstrando possuir importantes recursos para tanto, que atualmente o posicionam como o terceiro líder em inteligência artificial no mundo.

O país, além de honrosamente carregar o legado de Alan Turing, sustenta a tradição acadêmica de universidades como Oxford e Cambridge, que atualmente têm seus centros de estudos focados em inteligência artificial, o Oxford Artificial Intelligence Society – OXAI e o Cambridge AI, respectivamente. Soma-se a isso o fato de o Reino Unido contar com importantes companhias e *startups* no ramo da inteligência artificial: a DeepMind – companhia britânica que foi adquirida pelo Google por 400 milhões de libras²⁷⁷; a Graphcore, especializada em semicondutores baseados em *machine learning*; a Darktrace, empresa líder global em cibersegurança; e, ainda, no ramo da saúde, as empresas Benevolent AI e Babylon²⁷⁸ engrossam a fila das *startups* do Reino Unido de grande relevância internacional, para se citar algumas²⁷⁹.

²⁷⁷ GIBBS, Samuel. Google buys UK artificial intelligence startup Deepmind for £400m. *The Guardian*, 2014. Disponível em: <https://www.theguardian.com/technology/2014/jan/27/google-acquires-uk-artificial-intelligence-startup-deepmind>. Acesso em: 10 fev. 2022.

²⁷⁸ REINO UNIDO. *National AI Strategy*. Office for Artificial Intelligence. p. 18, 2021. Disponível em: <https://www.gov.uk/government/publications/national-ai-strategy>. Acesso em 10 fev. 2022.

²⁷⁹ Trata-se do terceiro país com mais *startups* desenvolvedoras de inteligências artificial, perdendo apenas para os EUA e a China. (STATISTA. *Number of artificial intelligence (AI) companies worldwide as of June 2018, by country*. 2018. Disponível em: <https://www.statista.com/statistics/941054/number-of-ai-companies-worldwide-by-country/>. Acesso em: 10 fev. 2022.)

Aliando-se à iniciativa privada, o governo britânico realiza altos investimentos em inteligência artificial. De 2014 a 2019, o governo injetou mais de £372 milhões em companhias de inteligência artificial por meio do British Business Bank²⁸⁰. Além disso, iniciativas que mereceram especial apoio financeiro do governo foram a NHS AI Lab, para a aceleração da adoção segura da inteligência artificial no ramo da saúde, e o Centre for Connected and Autonomous Vehicles, para desenvolvimento do futuro da mobilidade no Reino Unido, que receberam, ao todo, aproximadamente £500 milhões em investimentos governamentais²⁸¹.

Passando-se, então, para a análise do ponto de vista regulatório, foi criado, em 2019, o AI Council, um comitê não estatutário e independente formado por especialistas da área cujo objetivo é promover o aconselhamento especializado para o governo e as lideranças de alto nível do ecossistema britânico de inteligência artificial. Em janeiro de 2021, esse conselho publicou o “AI Roadmap”, que contém 16 recomendações governamentais no sentido de traçar uma estratégia nacional para a inteligência artificial no país.

O documento, por sua vez, deu origem ao “UK National AI Strategy”, publicado em 2021, que pode ser considerado atualmente como o instrumento mais importante sobre inteligência artificial no Reino Unido, pela sua atualidade e por traçar a estratégia do país sob vários aspectos, em curto, médio e longo prazo, para o desenvolvimento da IA.

É importante salientar que a referida estratégia nacional é rodeada por outras publicações estatais, que colaboram na definição das diretrizes para o uso e desenvolvimento da inteligência artificial no país. Apenas para se citar algumas, o Plan for Growth e o Innovation Strategy preveem a necessidade de desenvolvimento de um ambiente de expansão da inteligência artificial diverso e inclusivo, visando fortalecer a capacidade de inovação no país; e o Plan for Digital Regulation, que busca a implementação de uma abordagem pró-inovação para regular as tecnologias digitais, de forma a impulsionar a prosperidade tecnológica e criar confiança em seu uso²⁸².

²⁸⁰ REINO UNIDO. *National AI Strategy*. Office for Artificial Intelligence. p. 18, 2021. Disponível em: <https://www.gov.uk/government/publications/national-ai-strategy>. Acesso em 10 fev. 2022.

²⁸¹ Foram aplicados £250 milhões pelo Governo em cada uma das iniciativas, conforme consta da “UK AI National Strategy”. (REINO UNIDO. *National AI Strategy*. Office for Artificial Intelligence. p. 18, 2021. Disponível em: <https://www.gov.uk/government/publications/national-ai-strategy>. Acesso em 10 fev. 2022.)

²⁸² REINO UNIDO. *National AI Strategy*. Office for Artificial Intelligence. p. 13, 2021. Disponível em: <https://www.gov.uk/government/publications/national-ai-strategy>. Acesso em 10 fev. 2022.

Uma característica comum a todos esses regulamentos é a clara intenção do Reino Unido de não engessar o progresso tecnológico em razão de regras muito rígidas e imutáveis. Dada a importância da tecnologia e da inovação para o desenvolvimento econômico de qualquer país na atualidade, as publicações do Reino Unido – que visam estabelecer a estratégia para o emprego da inteligência artificial e o avanço do meio digital no país, de modo geral –, preveem a necessidade de uma regulamentação pró-inovação. Na própria estratégia nacional para a IA, define-se o que é considerado uma regulamentação pró-inovação:

Effective, pro-innovation governance of AI means that (i) the UK has a clear, proportionate and effective framework for regulating AI that supports innovation while addressing actual risks and harms, (ii) UK regulators have the flexibility and capabilities to respond effectively to the challenges of AI, and (iii) organisations can confidently innovate and adopt AI technologies with the right tools and infrastructure to address AI risks and harms. The UK public sector will lead the way by setting an example for the safe and ethical deployment of AI through how it governs its own use of the technology²⁸³.

Conforme o que já foi relatado neste capítulo, o governo britânico apoiou-se na análise do setor privado e do público, além de acadêmicos e especialistas, para informar o entendimento estatal quanto ao tipo de regulação que melhor atenderia aos anseios do país em relação à inteligência artificial. Como se viu, ponderados os prós e os contras de alguns modelos regulatórios (inclusive o de total abstenção de imposição de normas), chegou-se à conclusão de que há espaço para o estabelecimento de regulações principiológicas no país. Inserida nesse contexto, a estratégia nacional dispõe sobre algumas diretrizes que devem ser amplamente observadas no emprego da inteligência artificial, com especial destaque ao que é denominado de “AI assurance”, que pode ser compreendido como um pacote de características dos próprios sistemas, além de meios alternativos para se garantir ao público que um determinado sistema algorítmico seja confiável. Assim, é disposto que um dos parâmetros técnicos que deve ser almejado não apenas para aplicação no Reino Unido, mas também explorado em conjunto com

²⁸³ “A governança eficaz e pró-inovação da IA pode ser atingida por meio de (i) uma estrutura clara, proporcional e eficaz para regular a IA que apoia a inovação, por um lado, e que endereça riscos e possibilidades efetivas de dano, por outro, (ii) flexibilidade dos reguladores e recursos para responder efetivamente a os desafios da IA e (iii) inovação e adoção de tecnologias de IA com o uso de ferramentas tecnológicas confiáveis para lidar com os riscos e eventuais danos causados pela IA. O setor público do Reino Unido liderará o caminho, dando o exemplo a outras noções sobre a implantação segura e ética da IA, pela forma de governança aplicada domesticamente.” (REINO UNIDO. *National AI Strategy*. Office for Artificial Intelligence. p. 50, 2021, tradução nossa. Disponível em: <https://www.gov.uk/government/publications/national-ai-strategy>. Acesso em 10 fev. 2022.)

parcerias internacionais, para que seja um parâmetro global, é precisamente o desenvolvimento de inteligência artificial segura e confiável, no sentido de prover accountability, transparência e explicabilidade:

The Integrated Review set out the role of technical standards in embedding transparency and accountability in the design and deployment of technologies. AI technical standards (e.g. for accuracy, explainability and reliability) should ensure that safety, trust and security are at the heart of AI products and services²⁸⁴.

O documento não entra em maiores detalhes sobre esses princípios, mas apenas dispõe que são atributos importantes para a parametrização de uma inteligência artificial segura. Por fim, recorda-se que a questão da transparência e da prestação de contas já é debatida pelo governo britânico há alguns anos, considerando-se inclusive que no documento “Robotics and artificial intelligence” (publicado em 2016 pela House of Commons) já são mencionadas, também de forma vaga, essas características que são tidas como necessárias para a ampla aceitação das formas mais complexas de inteligência artificial na sociedade²⁸⁵. Além disso, em relação à explicabilidade, menciona-se que o *General Data Protection Regulation* da União Europeia, ao qual o Reino Unido é aderente, dispõe sobre o direito do titular de dados à explicação de decisões automatizadas²⁸⁶.

Em conclusão, uma vez mais, encontra-se a preocupação em se obter sistemas de inteligência artificial que sejam transparentes, explicáveis e cujos agentes sejam capazes de prestar contas em relação aos efeitos indesejados do sistema, quando for o caso.

Ao final da análise do arcabouço regulatório dos três países acima, reitera-se que foi opção desta autora evidenciar que os princípios têm relação mais imediata com a responsabilidade civil, o que será melhor demonstrado mais adiante neste trabalho.

²⁸⁴ “A Revisão Integrada estabeleceu o papel dos padrões técnicos para desenvolver a transparência e a prestação de contas no *design* e aplicação de tecnologias. Os padrões técnicos de IA (por exemplo, acurácia, explicabilidade e confiabilidade) devem garantir que a segurança, a confiança e a proteção estejam no centro de atenções dos produtos e serviços de IA.” (REINO UNIDO. *National AI Strategy*. Office for Artificial Intelligence. p. 57, 2021, tradução nossa. Disponível em: <https://www.gov.uk/government/publications/national-ai-strategy>. Acesso em 10 fev. 2022.)

²⁸⁵ REINO UNIDO. *Robotics and artificial intelligence*. House of Commons. 2016. Disponível em: <https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/14501.htm>. Acesso em: 10 fev. 2022.

²⁸⁶ UNIÃO EUROPEIA. Regulamento (UE) n° 2016/679 do Parlamento Europeu e do Conselho, de 23 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). *Jornal Oficial da União Europeia*, Estrasburgo, 04/05/2016, Artigo 22°. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016R0679&from=PT>. Disponível em: 10 fev. 2022.

Contudo, é importante mencionar que vários outros princípios são mencionados nesses regulamentos, tais como privacidade e não discriminação, mas que, por fugirem ao escopo desta pesquisa, não foram mencionados²⁸⁷.

3.4 REFLEXOS NO BRASIL

Não se poderia deixar de demonstrar os reflexos das ponderações feitas acima no cenário brasileiro de regulação da inteligência artificial. Sabe-se que o país vem lentamente evoluindo nesse sentido desde a publicação do primeiro Projeto de Lei sobre inteligência artificial. Trata-se do Projeto de Lei do Senado nº 5.051/2019 (PLS nº 5.051/2019), de autoria do Senador Styvenson Valentim²⁸⁸, ainda em tramitação, que tem a intenção de normatizar os princípios para o uso da inteligência artificial no Brasil. O mesmo senador propôs, em conjunto com aquele, um segundo Projeto de Lei do Senado, o de nº 5.691/2019, sobre a instituição da Política Nacional de Inteligência Artificial (PLS nº 5.691/2019)²⁸⁹.

Quanto a este último, entende-se que ele caminha, predominantemente, em conjunto com as iniciativas mundiais sobre inteligência artificial, se limitando a prever alguns princípios e diretrizes para “estimular a formação de um ambiente favorável ao desenvolvimento de tecnologias em Inteligência Artificial”²⁹⁰. Nesse documento, a “transparência, segurança e confiabilidade”²⁹¹ dos sistemas de inteligência artificial são exaltadas, assim como a necessidade de estas serem “inteligíveis, justificáveis e

²⁸⁷ Para uma pesquisa ampla sobre os princípios de IA de forma geral, adotados por diversos regulamentos no mundo, vide: ACHTEN, Nele *et al.* *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*. Berkman Klein Center Research. Publicação n. 2020-1, 2020. Disponível em: <https://ssrn.com/abstract=3518482>. Acesso em: 10 fev. 2022.

²⁸⁸ BRASIL. Senado Federal. *Projeto de Lei do Senado nº 5.051*, de 16 de setembro de 2019. Estabelece os princípios para o uso da Inteligência Artificial no Brasil. Brasília: Senado Federal, 2019. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/138790>. Acesso em: 10 fev. 2022.

²⁸⁹ BRASIL. Senado Federal. *Projeto de Lei do Senado nº 5.691*, de 25 de outubro de 2019. Institui a Política Nacional de Inteligência Artificial. Brasília: Senado Federal, 2019. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/139586>. Acesso em: 10 fev. 2022.

²⁹⁰ BRASIL. Senado Federal. *Projeto de Lei do Senado nº 5.691*, de 25 de outubro de 2019. Institui a Política Nacional de Inteligência Artificial. Artigo 1º. Brasília: Senado Federal, 2019. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/139586>. Acesso em: 10 fev. 2022.

²⁹¹ BRASIL. Senado Federal. *Projeto de Lei do Senado nº 5.691*, de 25 de outubro de 2019. Institui a Política Nacional de Inteligência Artificial. Artigo 2º. Brasília: Senado Federal, 2019. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/139586>. Acesso em: 10 fev. 2022.

acessíveis”²⁹². É previsto ainda que as decisões devem ser “rastreáveis e sem vieses discriminatórios ou preconceituosos”²⁹³.

O PLS nº 5.051/2019, por sua vez, também aborda os princípios da “transparência, a confiabilidade e a possibilidade de auditoria dos sistemas”²⁹⁴, em sintonia com o que foi proposto no PLS nº 5.691/2019 e também com os regulamentos dos países líderes em inteligência artificial, conforme analisado no tópico anterior.

Conforme muito bem sintetizado no *Policy Paper*, publicado pelo Centro de Pesquisa em Direito, Tecnologia e Inovação – DTIBR, o PLS nº 5.051/2019 tem os seus méritos:

Em síntese, essas virtudes são o pioneirismo no trato legislativo do tema no Brasil, bem como a abordagem principiológica da matéria, por meio de regras gerais, o que torna a lei mais maleável e capaz de acompanhar o progresso tecnológico, estratégia regulatória em sintonia com as principais iniciativas internacionais nessa matéria²⁹⁵.

Contudo, observam-se também grandes falhas no texto do referido Projeto de Lei relacionadas, principalmente, à estipulação da necessidade de supervisão dos sistemas e à matéria de responsabilidade civil. Sobre esses temas, na Justificação do projeto, é destacado que:

Nos termos da proposição, todo sistema de Inteligência Artificial terá a supervisão de uma pessoa humana, de forma compatível com cada aplicação. Com isso, é possível aliar as vantagens trazidas por essa inovação tecnológica com a necessária segurança, evitando que eventuais equívocos do sistema automatizado provoquem consequências indesejadas.

²⁹² BRASIL. Senado Federal. *Projeto de Lei do Senado nº 5.691*, de 25 de outubro de 2019. Institui a Política Nacional de Inteligência Artificial. Artigo 4º, IV. Brasília: Senado Federal, 2019. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/139586>. Acesso em: 10 fev. 2022.

²⁹³ BRASIL. Senado Federal. *Projeto de Lei do Senado nº 5.691*, de 25 de outubro de 2019. Institui a Política Nacional de Inteligência Artificial. Artigo 4º, VIII. Brasília: Senado Federal, 2019. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/139586>. Acesso em: 10 fev. 2022.

²⁹⁴ BRASIL. Senado Federal. *Projeto de Lei do Senado nº 5.051*, de 16 de setembro de 2019. Estabelece os princípios para o uso da Inteligência Artificial no Brasil. Artigo 2º, IV. Brasília: Senado Federal, 2019. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/138790>. Acesso em: 10 fev. 2022.

²⁹⁵ ALVES, Tárk César Oliveira e *et al.* *Policy Paper: Projeto de Lei do Senado n. 5.051/2019, que “estabelece os princípios para o uso da inteligência artificial no Brasil”*. Centro de Pesquisa em Direito Tecnologia e Inovação – DTIBR, 2019, p. 5. Disponível em: <https://www.dtibr.com/livros-e-artigos>. Acesso em: 10 fev. 2022.

Além disso, a fim de dirimir eventuais dúvidas acerca da responsabilidade por danos decorrentes de sistemas de inteligência artificial, o projeto define que a responsabilidade será, sempre, do supervisor humano do sistema²⁹⁶.

Com isso, os artigos do projeto preveem que os sistemas decisórios baseados em inteligência artificial serão, sempre, auxiliares à tomada de decisão humana e que, portanto, toda decisão automatizada deverá ser supervisionada por um humano de forma compatível com a “gravidade e as implicações da decisão submetida aos sistemas de Inteligência Artificial”²⁹⁷. Em primeiro lugar, entende-se que determinar de forma ampla que todas as decisões automatizadas sejam objeto de supervisão humana é um equívoco, por ter o potencial de refrear a inovação e os benefícios que poderiam ser colhidos do funcionamento autônomo da inteligência artificial. Além disso, existem algumas decisões automatizadas cujo impacto não é significativo, ou que não geram implicações no âmbito jurídico. Lembra-se, neste ponto, de alguns usos corriqueiros da inteligência artificial, como em pequenos eletrodomésticos (torradeiras e aspiradores de pó), cujas decisões, em sua maioria, não apresentarão problemas jurídicos, ou de nenhum outro tipo, que faça com que o usuário possa reclamar uma explicação. Por outro lado, existem aplicações de inteligência artificial que retornam decisões imediatas e não supervisionadas (já sendo utilizadas globalmente), como o caso dos carros autônomos e mecanismos de análise de crédito²⁹⁸. Fica claro, portanto, o potencial nocivo da obrigatoriedade de supervisão de todas as decisões automatizadas, indistintamente, e o seu desalinhamento tanto com a atual realidade do emprego da IA quanto com as iniciativas regulatórias internacionais sobre o tema. Em suma, conforme aduz Fabio Hartmann Peixoto e Marina de Alencar Araripe Coutinho:

Vincular o uso de IA a uma ideia de supervisor responsável também implica em reduzir as suas possibilidades, à medida que a própria IA inaugura uma outra dimensão nos mais variados campos do saber. Portanto, muito mais interessante seria indicar (regular) boas práticas e princípios nos campos de responsabilidade ética e responsabilidade normativa, que devem refletir nas

²⁹⁶ BRASIL. Senado Federal. *Projeto de Lei do Senado n° 5.051*, de 16 de setembro de 2019. Estabelece os princípios para o uso da Inteligência Artificial no Brasil. Justificação. Brasília: Senado Federal, 2019. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/138790>. Acesso em: 10 fev. 2022.

²⁹⁷ BRASIL. Senado Federal. *Projeto de Lei do Senado n° 5.051*, de 16 de setembro de 2019. Estabelece os princípios para o uso da Inteligência Artificial no Brasil. Artigo. 4º caput e parágrafo 1º. Brasília: Senado Federal, 2019. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/138790>. Acesso em: 10 fev. 2022.

²⁹⁸ ALVES, Târik César Oliveira *et al.* *Policy Paper: Projeto de Lei do Senado n. 5.051/2019, que “estabelece os princípios para o uso da inteligência artificial no Brasil”*. Centro de Pesquisa em Direito Tecnologia e Inovação – DTIBR, 2019, p. 19-20. Disponível em: <https://www.dtibr.com/livros-e-artigos>. Acesso em: 10 fev. 2022.

etapas de validação, verificação, controle e segurança tanto no desenvolvimento, quanto no uso de uma IA²⁹⁹.

Ainda mais preocupante do que a previsão a respeito da supervisão é a fixação de um regime de responsabilidade em que o supervisor seria sempre o responsável. Conforme levantado pelos pesquisadores do DTIBR, no *Policy Paper* sobre o PLS nº 5.051/2019, não há, sequer, no referido projeto uma definição de quem seria o supervisor do sistema³⁰⁰, o que nem sempre é de fácil aferição (seria sempre o seu usuário?; o seu desenvolvedor?; a pessoa física ou jurídica, no caso de responsabilização do fornecedor?). Por fim, ao estabelecer que a responsabilidade é sempre do supervisor, independentemente de culpa, há uma opção implícita pela aplicação do regime de responsabilidade objetiva, o que é feito de forma imponderada no projeto. Além disso, ao se predeterminar o responsável pelo dano, exprime-se uma ideia de descarte do nexo de causalidade como elemento necessário para demonstração da responsabilidade civil.

Conforme já se expôs neste trabalho, são inúmeros os desafios de se prever um regime de responsabilidade civil genérico, os quais não foram levados em conta no citado Projeto de Lei, o que é ainda mais grave neste caso em específico, haja vista que o regime objetivo de responsabilidade é imposto sobre um sujeito indefinido (o supervisor) e sem a necessária aferição do nexo causal entre conduta e dano³⁰¹.

²⁹⁹ PEIXOTO, Fabio Hartmann; COUTINHO, Marina de Alencar Araripe. Inteligência Artificial e Regulação: uma análise do Projeto de Lei 5.051/2019. *Revista Em Tempo*, [S.l.], v. 19, n. 1, ago. 2020. Disponível em: <https://revista.univem.edu.br/emtempo/article/view/3129>. Acesso em: 10 fev. 2022.

³⁰⁰ “Na prática, porém, a tarefa de identificar quem ou quais são os supervisores se revela muito mais complicada. Quem seria o supervisor de um software desenvolvido por equipe de programadores, onde cada um deles tem função específica, mas nenhum conhece os detalhes do projeto como um todo? Se esse mesmo software for licenciado para uso de terceiros, supervisor será o desenvolvedor originário ou o quem efetivamente utiliza o programa no dia a dia? Pense, ainda, em quem seria o supervisor de um assistente vocal instalado em telefone celular: o fabricante do aparelho (que desenvolveu o sistema) ou o próprio usuário, proprietário do celular (que ‘calibrou’ como o sistema irá funcionar no caso concreto, ao fornecer os inputs cotidianamente, conforme suas preferências pessoais)? Se o celular do exemplo anterior for revendido, isto muda o supervisor do sistema? No caso de um carro autônomo, que sabidamente reúne milhares de componentes, provenientes de fabricantes distintos, quem é o supervisor? Seria o fabricante de cada componente individualmente considerado; o fabricante final do carro; o revendedor/distribuidor; ou o usuário que utiliza o veículo (lembre-se que esses veículos podem ser utilizados no mesmo dia por dezenas de pessoas diferentes, como no caso de um carro autônomo em aplicativo de transporte, como o Uber).” (ALVES, Tárík César Oliveira e *et al.* *Policy Paper*: Projeto de Lei do Senado n. 5.051/2019, que “estabelece os princípios para o uso da inteligência artificial no Brasil”. Centro de Pesquisa em Direito Tecnologia e Inovação – DTIBR, 2019, p. 21. Disponível em: <https://www.dtibr.com/livros-e-artigos>. Acesso em: 10 fev. 2022.)

³⁰¹ ALVES, Tárík César Oliveira *et al.* *Policy Paper*: Projeto de Lei do Senado n. 5.051/2019, que “estabelece os princípios para o uso da inteligência artificial no Brasil”. Centro de Pesquisa em Direito Tecnologia e Inovação – DTIBR, 2019, p. 21. Disponível em: <https://www.dtibr.com/livros-e-artigos>. Acesso em: 10 fev. 2022.

Sobre o PLS nº 5.051/2019, em resumo, o que se observa é que é digna de aplausos a opção principiológica do texto, assim como foi feito no PLS nº 5.691/2019, o que os torna maleáveis aos diferentes tipos de inteligência artificial e ao seu avanço tecnológico ao longo do tempo. Em contrapartida, o fato de o próprio PLS nº 5.051/2019 determinar o regime de responsabilidade e quem deverá responder em caso de danos gerados por sistemas de inteligência artificial vai na contramão do que é proposto pelo próprio texto legal em relação à abordagem principiológica do tratamento da inteligência artificial, o que, segundo esta autora, deveria ser revisitado.

Apenas dois anos após o início da tramitação dos Projetos de Lei aludidos acima, em setembro de 2021, foi aprovado pela Câmara dos Deputados o Projeto de Lei nº 21-A/2020 (PL nº 21-A/2020), de autoria do deputado Eduardo Bismarck. Na justificativa do projeto, que levou à ambiciosa designação de “Marco Legal da Inteligência Artificial no Brasil”, é mencionada a necessidade de tornar “obrigatórios os princípios consagrados no âmbito internacional e disciplinando direitos e deveres”³⁰². Nessa toada, observa-se, novamente, o destaque ao princípio da transparência, considerado como “o direito das pessoas de serem informadas de maneira clara, acessível e precisa sobre a utilização das soluções de inteligência artificial, salvo disposição legal em sentido contrário e observados os segredos comercial e industrial”³⁰³ em hipóteses predeterminadas no próprio Projeto de Lei. Não se encontra no projeto, contudo, menção explícita ao princípio da prestação de contas, ou mesmo à necessidade de elaboração de relatórios de impacto pelos desenvolvedores do sistema, obrigação esta que muitas vezes é trazida de forma acessória e complementar ao princípio da *accountability*. Apesar disso, pode-se perceber a influência desse princípio no art. 6º, inciso VI do projeto, relativo à responsabilidade civil dos agentes que atuam na cadeia de desenvolvimento e de operação dos sistemas de inteligência artificial:

VI – responsabilidade: as normas sobre responsabilidade dos agentes que atuam na cadeia de desenvolvimento e operação de sistemas de inteligência artificial deverão, salvo disposição legal em contrário, pautar-se na responsabilidade subjetiva e levar em consideração a efetiva participação desses agentes, os danos específicos que se deseja evitar ou remediar e a forma como esses agentes podem demonstrar adequação às normas aplicáveis, por

³⁰² BRASIL. Câmara dos Deputados. *Projeto de Lei nº 21*, de 04 de fevereiro de 2020. Justificativa. Brasília, Câmara dos Deputados, 2020. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 10 fev. 2022.

³⁰³ BRASIL. Câmara dos Deputados. *Projeto de Lei nº 21-A*, de 29 de setembro de 2021. Artigo 5º, V. Brasília, Câmara dos Deputados, 2021. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 10 fev. 2022.

meio de esforços razoáveis compatíveis com os padrões internacionais e as melhores práticas de mercado³⁰⁴.

Ao mencionar “a forma como esses agentes podem demonstrar adequação às normas aplicáveis, por meio de esforços razoáveis compatíveis com os padrões internacionais e as melhores práticas de mercado” compreende-se que há aí um empenho no sentido de predefinir padrões mínimos aceitáveis aos quais os desenvolvedores ou fornecedores da inteligência artificial devem se fiar para evitar ou mitigar sua responsabilidade, o que vai ao encontro das normas internacionais sobre *accountability*.

Por outro lado, nota-se, uma vez mais, o equívoco do projeto em tentar, sem sucesso, definir o tipo de responsabilidade aplicável aos casos de danos gerados por sistemas de inteligência artificial. Nesse caso, a definição da responsabilidade civil subjetiva é contrária às recomendações feitas neste trabalho, visto que o tipo de responsabilidade aplicável dependerá do mercado no qual a aplicação de inteligência artificial é utilizada. E mesmo que fosse possível definir, genericamente, um tipo de responsabilidade para os casos de danos gerados pelas inúmeras aplicações de IA, tendo em vista o desenvolvimento do instituto da responsabilidade civil e a consequente preferência pela responsabilidade objetiva em casos envolvendo acidentes de consumo e proteção de dados (que muitas vezes estarão ligados aos danos causados por IA), de acordo com Anderson Schreiber, o projeto “cria uma fratura no ordenamento brasileiro” ao optar pela responsabilidade subjetiva³⁰⁵. O mesmo autor ainda afirma que:

Soa, ademais, inconsistente com o próprio texto do Projeto de Lei, que alude todo o tempo aos “riscos concretos” trazidos pela Inteligência Artificial (art. 6º, III) e à necessidade de sua comparação com “os riscos apresentados por sistemas similares que não envolvam inteligência artificial” (art.6º, III, alínea b)³⁰⁶.

É importante salientar que a redação final aprovada pela Câmara dos Deputados é substancialmente diferente da redação original do Projeto. Porém, caso esta tivesse se

³⁰⁴ BRASIL. *Projeto de Lei nº 21-A*, de 29 de setembro de 2021. Artigo 6º, VI. Brasília: Câmara dos Deputados, 2021. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 10 fev. 2022.

³⁰⁵ SCHREIBER, Anderson. *PL da Inteligência Artificial cria fratura no ordenamento jurídico brasileiro*. Jota, 2021, p. 5. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/coluna-do-anderson-schreiber/pl-inteligencia-artificial-cria-fratura-no-ordenamento-juridico-02112021>. Acesso em: 10 fev. 2022.

³⁰⁶ SCHREIBER, Anderson. *PL da Inteligência Artificial cria fratura no ordenamento jurídico brasileiro*. Jota, 2021, p. 5. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/coluna-do-anderson-schreiber/pl-inteligencia-artificial-cria-fratura-no-ordenamento-juridico-02112021>. Acesso em: 10 fev. 2022.

mantido, ela seria igualmente merecedora de críticas. Em primeiro lugar, havia uma tentativa de se prever qual a instituição responsável pelo sistema de IA, ao se mencionar que é um direito das partes interessadas “a ciência da instituição responsável pelo sistema de inteligência artificial”³⁰⁷. Em contrapartida, era previsto o dever do agente de IA “divulgar publicamente a instituição responsável pelo estabelecimento de sistema de inteligência artificial”³⁰⁸. Resta saber quem definiria, então, quais são as instituições responsáveis pelo sistema. Seria o próprio setor privado? Havendo a definição apriorística do responsável, ter-se-ia, necessariamente, responsabilidade objetiva? Como fica o papel do nexo de causalidade na aferição de responsabilidade? Isso parece contraditório com o próprio texto do Projeto original que, mais adiante, determina que os agentes de inteligência artificial devem “responder, na forma da lei, pelas decisões tomadas por um sistema de inteligência artificial”³⁰⁹. Ora, se se responde na forma da lei, não pode haver uma atribuição pelo próprio interessado de quem será o responsável, uma vez que essa determinação decorre da lei.

Além disso, também era previsto o seguinte dever dos agentes de inteligência artificial:

IV – implantar um sistema de inteligência artificial somente após avaliação adequada de seus objetivos, benefícios e riscos relacionados a cada fase do sistema e, caso seja o responsável pelo estabelecimento do sistema, encerrar o sistema se o seu controle humano não for mais possível [...] Nota-se que a supervisão humana já prevista no PLS nº 5051/2019 aparece no texto original do Projeto de Lei nº 21/2020 sob uma nova figura: a da necessidade de controle humano, sob pena de encerramento do sistema. Dessa forma, a redação original do Projeto de Lei nº 21/2020 inviabilizaria o uso de inúmeros sistemas de inteligência artificial já altamente intrincados na sociedade. Por exemplo, não seria possível ter mecanismos de recomendação de filmes em plataformas de streaming, não se poderia utilizar a inteligência artificial em ferramentas de pesquisas, como o Google, e redes sociais, como o Instagram, não poderiam mais ser utilizadas no Brasil, em razão dos algoritmos de escolha de conteúdo

³⁰⁷ As “partes interessadas” eram definidas no Projeto de Lei como “todos aqueles envolvidos ou afetados, direta ou indiretamente, por sistemas de inteligência artificial”. (BRASIL. *Projeto de Lei nº 21*, de 04 de fevereiro de 2020. Artigo 2º, V. Brasília: Câmara dos Deputados, 2020. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 10 fev. 2022.)

³⁰⁸ Os “agentes de inteligência artificial” eram definidas no Projeto de Lei como “pessoas físicas ou jurídicas, de direito público ou privado, e entes sem personalidade jurídica”. (BRASIL. *Projeto de Lei nº 21*, de 04 de fevereiro de 2020. Artigo 2º, IV. Brasília: Câmara dos Deputados, 2020. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 10 fev. 2022.)

³⁰⁹ BRASIL. Câmara dos Deputados. *Projeto de Lei nº 21*, de 04 de fevereiro de 2020. Artigo 9º, V. Brasília: Câmara dos Deputados, 2020. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 10 fev. 2022.

não serem supervisionados. Tem-se aí o disparate da redação originalmente proposta no Projeto de Lei nº 21/2020³¹⁰.

Percebe-se com isso que os Projetos de Lei do Brasil sobre inteligência artificial e responsabilidade civil apresentados até o momento não são muito animadores. Todos os três Projetos mencionados acima demonstram atecnia e inconsistência entre os artigos de um mesmo Projeto, bem como com o sistema legal brasileiro. Não poderia ser diferente, uma vez que eles não tiveram o tempo necessário para maturação. O PL nº 21-A/20, por exemplo, foi, desnecessariamente, tramitado em caráter de urgência, o que levou à aprovação de um Projeto pela Câmara sem a essencial participação dos especialistas na área. Conforme aduz André Lucas Fernandes, em crítica ao PL nº 21-A/20:

Regular de forma cautelar é regular com ampla participação de setores, com debates maduros, em oposição ao que ocorreu na aprovação apressada do PL pela Câmara dos Deputados, justo em tema tão crítico. Registre-se que a Europa está em meio aos debates regulatórios da inteligência artificial, num acúmulo que já conta alguns anos³¹¹.

Parte-se, por fim, para análise da Estratégia Brasileira de Inteligência Artificial, ou EBIA, como ficou conhecida a Portaria nº 4.617 de abril de 2021, lançada pelo Ministério da Ciência, Tecnologia e Inovações. Trata-se de documento concebido “durante os anos de 2019 e 2020 a partir de visões diversas e da convergência de estudos, reflexões, pesquisas e de consulta aos especialistas, empresas, pesquisadores e órgãos públicos, tendo como propósito o planejamento tecnológico de longo prazo para o país”³¹²; o que foi feito seguindo três principais eixos, que são: a legislação, a regulação e o uso ético; a governança da IA; e os aspectos internacionais de governança da inteligência artificial³¹³. O relatório teve grande influência das recomendações da OCDE

³¹⁰ BRASIL. Câmara dos Deputados. *Projeto de Lei nº 21*, de 04 de fevereiro de 2020. Artigo 9º, IV. Brasília: Câmara dos Deputados, 2020. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 10 fev. 2022.

³¹¹ FERNANDES, André Lucas. *PL da Inteligência Artificial erra ao criar regime de responsabilidade subjetiva*. Jota, 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/pl-21-2020-inteligencia-artificial-01112021>. Acesso em: 10 fev. 2022.

³¹² BRASIL. *Portaria nº 4.617*, de 06 de abril de 2021. Institui a Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos. Brasília: Ministério da Ciência, Tecnologia e Inovações, 2021, p. 6. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial>. Acesso em: 10 fev. 2022.

³¹³ BRASIL. *Portaria nº 4.617*, de 06 de abril de 2021. Institui a Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos. Brasília: Ministério da Ciência, Tecnologia e Inovações, 2021, p. 7. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial>. Acesso em: 10 fev. 2022.

sobre inteligência artificial³¹⁴, às quais o Brasil foi um dos países aderentes. Muito em função disso, a EBIA explora os princípios para desenvolvimento da inteligência artificial já consagrados pela regulamentação internacional³¹⁵.

Dessa forma, pensa-se ser salutar que a EBIA “importa” as discussões sobre esses princípios, traçando desde já a necessidade de que o desenvolvimento da inteligência artificial esteja pautado na explicabilidade, na transparência e na prestação de contas (*accountability*). Um dos méritos da EBIA relaciona-se ao fato de que nela estão descritas as formas pelas quais a prestação de contas pode ser colocada em prática:

Contudo, a ideia de *accountability* (aqui traduzida como responsabilidade e a prestação de contas) impõe que, a depender da aplicação de IA e dos riscos a ela associados, sejam estabelecidas estruturas de governança de IA, que possam assegurar a adoção de princípios para IA confiável e implementar mecanismos para sua observância. Tais mecanismos podem incluir: (i) a designação de indivíduos ou de grupos específicos dentro da organização para promover a conformidade com os princípios; (ii) a adoção de medidas para aumentar a conscientização interna sobre a necessidade dessa conformidade, inclusive por meio de orientações e treinamentos em toda a empresa; e (iii) a implementação de um processo de escalação por meio do qual os funcionários possam levantar preocupações de conformidade e resolver essas preocupações. Podem, ainda, envolver a criação de selos, certificações e códigos de conduta corporativos ou governamentais. Em relação aos processos de revisão ética, muitas organizações já têm ou estão considerando a criação de conselhos de revisão de dados ou comitês de ética em relação à IA, que podem ser internos ou externos a tais organizações. Essa é vista como uma maneira de impulsionar *accountability* dentro das corporações³¹⁶.

Além disso, é mencionada a importância da elaboração de relatórios de impacto para uso de inteligência artificial de alto risco, bem como há recomendação pela criação de conselhos destinados à revisão e à análise dos possíveis efeitos colaterais de sistemas de inteligência artificial em companhias destinadas ao desenvolvimento destes.

Quanto à transparência e à explicabilidade, o uso desses termos pelo EBIA gera uma certa confusão. Ao mesmo tempo que se tenta definir a transparência como a “adoção

³¹⁴ OCDE. *Recommendation of the Council on Artificial Intelligence*. Secretary-General of the OECD, 2019. Disponível em: <https://oecd.ai/en/ai-principles>. Acesso em: 10 fev. 2022.

³¹⁵ “Alinhada às diretrizes da OCDE endossadas pelo Brasil, a EBIA fundamenta-se nos cinco princípios definidos pela Organização para uma gestão responsável dos sistemas de IA, quais sejam: (i) crescimento inclusivo, o desenvolvimento sustentável e o bem-estar; (ii) valores centrados no ser humano e na equidade; (iii) transparência e explicabilidade; (iv) robustez, segurança e proteção; e (v) a responsabilização ou a prestação de contas (*accountability*)”. (BRASIL. *Portaria nº 4.617*, de 06 de abril de 2021. Institui a Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos. Brasília: Ministério da Ciência, Tecnologia e Inovações, 2021, p. 17. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial>. Acesso em: 10 fev. 2022.)

³¹⁶ BRASIL. *Portaria nº 4.617*, de 06 de abril de 2021. Institui a Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos. Brasília: Ministério da Ciência, Tecnologia e Inovações, 2021, p. 24. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial>. Acesso em: 10 fev. 2022.

de metodologias transparentes e auditáveis quanto ao desenvolvimento dos sistemas de IA”³¹⁷ – o que, por si só, já é uma definição imprecisa e pouco técnica do termo –, ao longo do documento a transparência e a explicabilidade são tratadas muitas vezes como sinônimos. Concluindo os tópicos destinados à necessidade de interpretabilidade, de modo geral, dos sistemas de inteligência artificial, é admitido no EBIA que:

Ainda que tecnologias de machine learning sejam frequentemente caracterizadas como “sistemas fechados”, é possível incorporar tais ideias aos sistemas de IA de várias maneiras, inclusive por meio da implementação de mecanismos para facilitar a rastreabilidade do processo decisório e do desenvolvimento e emprego de ferramentas e de técnicas de explicabilidade.

O trecho acima é visto com bons olhos, uma vez que remete à explicabilidade no sentido de se preservar a caixa-preta do sistema, permitindo que seus usuários colham os frutos de um sistema mais preciso, mas, ao mesmo tempo, tenham a possibilidade de compreender os elementos (dados) que fizeram com que uma determinada decisão fosse tomada.

Por fim, sobre a regulamentação ou não do tema no Brasil, compreende-se também que a EBIA tem uma abordagem correta ao prever que um eventual regulamento brasileiro sobre inteligência artificial deve “(i) desenvolver estruturas legais existentes; (ii) adotar abordagem regulatória baseada em princípios e resultados; (iii) fazer um ‘teste de equilíbrio de riscos/benefícios’ centrado no indivíduo humano e (iv) fazer avaliação de impacto contextual”³¹⁸.

Apesar disso, a estratégia brasileira foi criticada em alguns aspectos, principalmente por lhe “faltar materialidade”³¹⁹ e por “não apresentar nenhuma meta, orçamento, organização ou planejamento de implementação”³²⁰. Ao se analisar as estratégias dos demais países estudados nesta dissertação, percebe-se que há uma previsão

³¹⁷ BRASIL. *Portaria nº 4.617*, de 06 de abril de 2021. Institui a Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos. Brasília: Ministério da Ciência, Tecnologia e Inovações, 2021, p. 25. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial>. Acesso em: 10 fev. 2022.

³¹⁸ BRASIL. *Portaria nº 4.617*, de 06 de abril de 2021. Institui a Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos. Brasília: Ministério da Ciência, Tecnologia e Inovações, 2021, p. 22. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial>. Acesso em: 10 fev. 2022.

³¹⁹ XAVIER, Fábio Correa. A Estratégia Brasileira de Inteligência Artificial. *MIT Technology Review*, 2021. Disponível em: <https://mittechreview.com.br/a-estrategia-brasileira-de-inteligencia-artificial/#:~:text=Segundo%20o%20MCTIC%2C%20a%20Estrat%C3%A9gia,e%20em%20prol%20de%20>. Acesso em: 10 fev. 2022.

³²⁰ LEMOS, Ronaldo. Estratégia Brasileira de IA é patética. *Folha de S.Paulo*, 2021. Disponível em: <https://www1.folha.uol.com.br/colunas/ronaldolemos/2021/04/estrategia-de-ia-brasileira-e-patetica>. Acesso em: 10 fev. 2022.

de investimento, objetivos claros para curto, médio e longo prazo, além da criação de órgãos e comitês especializados para o desenvolvimento de meios para se alcançar esses objetivos, o que não é visto na estratégia brasileira. O próprio documento menciona em diversas oportunidades³²¹ a defasagem tecnológica do Brasil, o que demandaria proposições mais enquadradas aos problemas vivenciados pelo país, em vez de as estratégias se basearem tão intensamente na perspectiva de países que estão em outro estágio de desenvolvimento e de aplicação da inteligência artificial. Por outro lado, admite-se que a reverberação dos princípios já consagrados internacionalmente é um ponto positivo da EBIA, pois, independentemente do avanço das discussões sobre a responsabilidade civil e a inteligência artificial no Brasil e o seu futuro regulamento, é importante que esta abordagem esteja em sintonia com as estratégias internacionais, dado o caráter global do uso da tecnologia e dos dados e do emaranhado das relações sociais que transcendem o âmbito nacional.

3.5 PRINCÍPIOS: PRESTAÇÃO DE CONTAS, EXPLICABILIDADE E TRANSPARÊNCIA

O que se observa da análise do arcabouço regulatório sobre a inteligência artificial nos países estudados acima é que todos levam em consideração, com grande destaque, os princípios de prestação de contas (*accountability*), explicabilidade e transparência. Apesar de esses termos aparecerem em todos os regulamentos analisados dos quatro países, nota-se que não houve grande preocupação em defini-los bem, ou mesmo identificar como eles seriam colocados em prática.

Para fins deste trabalho, no entanto, é imprescindível sistematizar melhor esses princípios e identificar a sua correlação com a responsabilidade civil. Para tanto,

³²¹ “Segundo o Índice de Inovação Global de 2019, o Brasil está situado na 66ª posição, tendo como dois de seus maiores desafios o investimento no ambiente de negócios e a infraestrutura tecnológica. Dados do Banco Mundial corroboram tal diagnóstico, ao classificar o Brasil em 138º lugar quanto à facilidade de se iniciar um negócio e em 124º lugar no que se refere à facilidade de se realizar um negócio. Além disso, o Brasil investe 1,27% do seu PIB em P&D, enquanto a média de investimento dos países-membros da OCDE é de 2,39%. [...] No que diz respeito à Inteligência Artificial, os desafios ainda são consideráveis. Em 2019, enquanto os EUA investiram 224 milhões USD em *startups* de IA, e a China 45 milhões USD, o Brasil investiu apenas 1 milhão USD10. Desse modo, é essencial para a Estratégia Brasileira de Inteligência Artificial estabelecer diretrizes que habilitem o fomento de um ambiente de empreendedorismo em I A no país”. (BRASIL. *Portaria n° 4.617*, de 06 de abril de 2021. Institui a Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos. Brasília: Ministério da Ciência, Tecnologia e Inovações, 2021, p. 25. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial>. Acesso em: 10 fev. 2022.)

preliminarmente, faz-se necessário definir com maior precisão o que se entende por *accountability*, o que, conforme já exposto anteriormente, tem sido traduzido para o português como “prestação de contas” e/ou “responsabilidade”. Nesse sentido, são valiosos os estudos de Stephen McGrath e Stephen Whitty, que buscam definir e diferenciar os termos *accountability* e *responsability*³²². Nesse estudo, eles se valem de um rigoroso método de refino da definição de cada palavra, removendo todos os seus elementos não essenciais, que podem produzir sobreposições entre as definições, para se alcançar a versão mais apurada de cada conceito³²³.

Um dos passos da pesquisa consistiu na análise de diversas definições do termo *accountability* encontradas em diferentes dicionários, conforme sintetizado pelos autores na tabela abaixo:

Figura 1 – Definições de “accountability” em diversos dicionários

Dictionary	Definition of accountability (all sourced on 5 June 2017)
Business	The obligation of an individual or organisation to account for its activities, accept responsibility for them, and to disclose the results in a transparent manner. It also includes the responsibility for money or other entrusted property
Cambridge	A situation in which someone is responsible for things that can happen and can give a satisfactory reason for them
Collins	None given
Concise Oxford	None given
Dictionary.com	The state of being accountable, liable, or answerable
Longman	None given
Macmillan	A situation in which people know who are responsible for something and can ask them to explain its state or quality
Macquarie	The state of being liable to be called to account
Merriam-Webster	The quality or state of being accountable
Oxford	The fact or condition of being accountable; responsibility
The free dictionary	The state of being accountable, liable or answerable
Wiktionary	The state of being accountable; liability to be called on to render an account; accountableness; responsible for; answerable for

Fonte: WHITTY, Jonathan; MCGRATH, Stephen Keith (p. 687-707).

³²² WHITTY, Jonathan; MCGRATH, Stephen Keith. Accountability and responsibility defined. *International Journal of Managing Projects in Business*, v. 11, p. 687-707.

³²³ “We seek to resolve definitional confusion through scientific analysis using deduction rather than induction, as Popper (1979, p. 86) noted ‘Hume had shown induction invalid’. We therefore do not survey multitudes of current usages of the term that may not have grasped the essence of the difference between them [...] We therefore seek a method that is based on determining essential characteristics”. / “Procuramos resolver a confusão na definição através da análise científica utilizando o método dedutivo ao invés do indutivo, como Popper (1979, p. 86) observou ‘Hume demonstrou a indução inválida’ [...] Buscamos, portanto, um método que se baseie na identificação de características essenciais.” (WHITTY, Jonathan; MCGRATH, Stephen Keith. Accountability and responsibility defined. *International Journal of Managing Projects in Business*, v. 11, 687-707, p. 690, tradução nossa.)

A partir dessas definições, observa-se a grande interconexão entre os termos *accountability* e *liability*³²⁴ que, segundo os autores, é bem-vinda:

Liability is stronger than responsibility, carrying some legal force or implication of possible penalty for not doing so, which being called to account comprises and so it clearly distinguishes accountability from responsibility. [...] Liability implies being held to account anyway and so even though accountability is derived from the word account, its definition, provided it is compatible with its root term, does not actually need to explicitly include these words and in this case, it would be tautological if it did³²⁵.

Ao final de toda a análise, os autores chegam à conclusão de que a melhor definição para o termo *accountability* seria “liability for ensuring a task is satisfactorily done”³²⁶⁻³²⁷, o que poderia ser traduzido por “ser responsável (no sentido de responder legal ou administrativamente, inclusive com a aplicação de penas) por garantir que uma tarefa seja cumprida de forma satisfatória” (tradução nossa). Tem-se claro aí a correlação entre a chamada *accountability* e a responsabilidade civil no Brasil, uma vez que se trata do dever de se responder legalmente por uma tarefa realizada. O prestador de contas tem a obrigação de assegurar que a tarefa foi cumprida a contento, sob pena de responder no caso contrário.

³²⁴ Conforme elucida Nelson Rosenvald: “No *common law* há um termo que se ajusta perfeitamente ao clássico sentido civilístico da responsabilidade. Trata-se da “liability”. Várias teorias desenvolvem a liability no contexto da responsabilidade civil. Em comum, remetem à uma indenização cujo núcleo consiste em um nexo causal entre uma conduta e um dano, acrescida por outros elementos conforme o nexo de imputação concreto, tendo em consideração as peculiaridades de cada jurisdição”. (ROSELVALD, Nelson. A polissemia da responsabilidade civil na LGPD. Migalhas de Proteção de Dados, 2020. Disponível em: <https://www.migalhas.com.br/coluna/migalhas-de-protecao-de-dados/336002/a-polissemia-da-responsabilidade-civil-na-lgpd>. Acesso em 10 fev. 2022.)

³²⁵ “‘Liability’ é um termo ‘mais forte’ do que a ‘responsibility’, por carregar aspecto de força legal, ou de implicação de uma possível penalidade em caso de descumprimento. Considerando que ser chamado a prestar contas evidentemente compreende a possibilidade de penalidades em caso de não o fazer de forma satisfatória, a ‘accountability’ é afastada do termo ‘responsibility’ [...] ‘Liability’ implica ser chamado a prestar contas e, portanto, embora o termo ‘accountability’ seja derivado da palavra ‘conta’, sua definição, desde que seja compatível com seu termo raiz, não precisa incluir explicitamente essas palavras e, neste caso, seria tautológico se o fizesse.” (WHITTY, Jonathan; MCGRATH, Stephen Keith. Accountability and responsibility defined. *International Journal of Managing Projects in Business*, v. 11, 687-707, p. 698, tradução nossa.)

³²⁶ WHITTY, Jonathan; MCGRATH, Stephen Keith. *Accountability and responsibility defined. International Journal of Managing Projects in Business*, v. 11, 687-707, p. 701.

³²⁷ *Responsibility* por sua vez é definido como “an obligation to satisfactorily perform a task”. (WHITTY, Jonathan; MCGRATH, Stephen Keith. Accountability and responsibility defined. *International Journal of Managing Projects in Business*, p. 701.) A diferença encontra-se no ponto de que a responsabilidade para os autores se traduz como uma obrigação lato sensu de performar uma tarefa, enquanto *accountability* traduz-se no dever de responder legalmente perante as tarefas realizadas.

Acudindo na análise comparativa entre o termo inglês e a legislação nacional, Nelson Rosenvald distingue faz a distinção entre *accountability ex ante* e *accountability ex post*:

No plano ex ante a *accountability* é compreendida como um guia para controladores e operadores, protagonistas do tratamento de dados pessoais, mediante a inserção de regras de boas práticas que estabeleçam procedimentos, normas de segurança e como planificação para os riscos de maior impacto negativo [...]. Já na vertente ex post, a *accountability* atua como um guia para o magistrado e outras autoridades, tanto para identificar e quantificar responsabilidades, como para estabelecer os remédios mais adequados. Assim, ao invés do juiz se socorrer da discricionariedade para aferir o risco intrínseco de uma certa atividade por sua elevada danosidade - o desincentivo ao empreendedorismo é a reação dos agentes econômicos à insegurança jurídica -, estabelecem-se padrões e garantias instrumentais que atuam como parâmetros objetivos para a mensuração do risco em comparação com outras atividades³²⁸.

Nesse sentido de estabelecimento de padrões e garantias objetivos para análise de risco, o termo *accountability* é também muitas vezes referenciado para fazer menção à necessidade de comprovação de que uma determinada tarefa foi realizada de forma satisfatória, seguindo os parâmetros exigidos para tanto. É o caso, por exemplo, do Accountability Algorithm Act, dos Estados Unidos, já mencionado neste trabalho, no qual é prevista a necessidade de desenvolvedores de sistemas de inteligência artificial, que expõem os indivíduos a decisões consideradas de alto risco, elaborarem um relatório de impacto dos respectivos sistemas. Em última análise, o relatório serve como medida de comprovação de que o sistema foi modelado de acordo com as regras estabelecidas para tanto e, se assim comprovado, o seu desenvolvedor mereceria redução, ou mesmo a exclusão, da sanção ou responsabilização por eventuais danos causados. No Brasil, a Lei Geral de Proteção de Dados vai no mesmo sentido ao correlacionar a “responsabilização e prestação de contas” (em uma clara tradução do termo *accountability*, utilizado no GDPR) com a “demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas”³²⁹. Tem-se aí, mais uma vez, o paralelismo entre a *accountability* e a responsabilidade, no sentido de que, a demonstração do atendimento

³²⁸ ROSELVALD, Nelson. A polissemia da responsabilidade civil na LGPD. Migalhas de Proteção de Dados, 2020. Disponível em: <https://www.migalhas.com.br/coluna/migalhas-de-protecao-de-dados/336002/a-polissemia-da-responsabilidade-civil-na-lgpd>. Acesso em 10 fev. 2022.

³²⁹ BRASIL. Art. 6º, X. Assembleia Legislativa. Lei 13.709/2018. Lei Geral de Proteção de Dados. Disponível em http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm. Acesso em: 10 fev. 2022.

de requisitos mínimos impostos servirá para aferição dos riscos tomados e da ilicitude da conduta.

Delimitado, então, o conceito de *accountability*, que se traduz na necessidade de se responder por uma tarefa (seja no sentido de reparar os danos causados, seja no sentido de demonstrar que a tarefa foi realizada seguindo os parâmetros predeterminados), pode-se concluir que os princípios de transparência e de explicabilidade são decorrentes dessa obrigação de prestação de contas: para que seja possível a prestação de contas, o algoritmo deve ser explicável, transparente, interpretável. Conforme Madalina Busuioc:

Accountability is closely linked to “answerability”, and a key element thereof is that of explanation or justification [...] Information and explanation requirements serve as the foundation that allows the accountability forum to assess and judge whether decisions or actions undertaken were appropriate. For meaningful accountability to have taken place — and to the extent that failings have been identified — there should be a possibility to extract consequences³³⁰.

A explicabilidade e a transparência, então, funcionam, na verdade, como métodos para se garantir o princípio da *accountability*. Entende-se, assim, que estes primeiros não devem ser utilizados como princípios ao lado de *accountability* e, muito menos, de forma conjunta nos regulamentos, considerando que são meios distintos de se alcançar a prestação de contas.

Quanto à diferença entre a transparência e a explicabilidade nos sistemas de *machine learning*, esse tema já é velho conhecido entre os estudiosos da inteligência artificial, podendo-se dizer que a diferença, já bem assimilada na comunidade acadêmica, é no sentido de que a *transparência* converge com a ideia de se desvendar o código-fonte do algoritmo, para que se conheça toda a linguagem da codificação, para, então, ser possível identificar o que levou a uma determinada decisão pelo sistema. Por outro lado, o método de *explicação* prescinde do conhecimento do código-fonte, partindo da aplicação de outros modelos para que se alcance uma explicação de determinada decisão,

³³⁰ “A prestação de contas está intimamente ligada à necessidade de se prover uma resposta, e um elemento-chave dela é a explicação ou justificativa [...] Os requisitos de informação e explicação servem como a base que permite que o público avalie e julgue se as decisões ou ações tomadas foram apropriadas. Para que uma prestação de contas significativa tenha ocorrido – e na medida em que as falhas tenham sido identificadas – deve haver a possibilidade de se aplicar penalidades.” (BUSUIOC, Madalina. *Accountable Artificial Intelligence: Holding Algorithms to Account. Public Administration Review*, Florida, v. 81, p. 3, 2020, tradução nossa.)

que seja apreensível ao usuário, ou operador do sistema, sem necessidade de conhecimento técnico sobre a complexidade algorítmica que o molda.

A diferença entre transparência e explicabilidade, então, pode ser resumida da seguinte forma: abrir a caixa-preta do algoritmo (no caso da transparência) ou preservá-la, porém sendo capaz de explicá-la (no caso da explicabilidade). Mencionou-se acima que há certo consenso científico sobre essa diferenciação entre transparência e explicabilidade, uma vez que renomados artigos acadêmicos e mesmo publicações governamentais assumem o termo “transparência” como sendo a transparência técnica e tecnológica, no sentido de se poder acessar o código-fonte, enquanto a explicabilidade é identificada como outros métodos de se apreender o racional de uma decisão sem necessariamente se conhecer a linguagem do seu código³³¹. Sob esse mesmo ângulo, Bruno Bioni e Maria Luciano declaram:

Ao exigir informações sobre a racionalidade de uma decisão específica, o direito a explicação não se confunde com a transparência pura e simples. Variações nos dados de raça, por exemplo, já poderiam fornecer o impacto e a maneira como esse tipo de dado impacta uma decisão, sem, contudo, demandar a revelação de todo o sistema automatizado envolvido naquela decisão³³².

Sobre a explicabilidade, o artigo “Accountability of AI Under the Law: The Role of Explanation” corrobora que:

Explanations expose information about specific individual decisions without necessarily exposing the precise mechanics of the decision-making process. Explanations can be used to prevent or rectify errors and increase trust. Explanations can also be used to ascertain whether certain criteria were used appropriately or inappropriately in case of a dispute³³³.

E, ainda, segundo Marco Túlio Ribeiro, Samee Singh e Carlos Guestrin:

³³¹ Apesar disso, encontra-se também estudos que utilizam os termos como sinônimos, com o que não se concorda. Nota-se que pelos regulamentos estudados no tópico acima, há uma certa preferência pelo termo “transparência”, mas sem que isso necessariamente corresponda à maior inclinação ao método de abertura do código-fonte. Ao contrário, parece a esta autora que o termo foi muitas vezes utilizado como sinônimo de explicabilidade ou, de forma genérica, como qualquer método que faça com que o sistema se torne rastreável ou interpretável.

³³² BIONI, Bruno; LUCIANO, Maria. *O Princípio da Precaução na Regulação de Inteligência Artificial: seriam as leis de proteção de dados o seu portal de entrada?* p. 220. In: FRAZÃO, Ana. MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Revista dos Tribunais, 2019, p. 209-231.

³³³ “As explicações expõem informações sobre decisões individuais específicas sem necessariamente expor a mecânica precisa do processo de tomada de decisão automatizada. Elas podem ser usadas para prevenir, ou corrigir erros e aumentar a confiança. Além disso, as explicações também podem ser usadas para verificar se determinados critérios foram usados de forma adequada ou inadequada em caso de disputa.” (DOSHI-VELEZ, Finale; KORTZ, Mason. *Accountability of AI Under the Law: The Role of Explanation*. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society Working Paper, 2017, tradução nossa. Disponível em: <https://dash.harvard.edu/handle/1/34372584>. Acesso em: 10 fev. 2022.)

By “explaining a prediction”, we mean presenting textual or visual artifacts that provide qualitative understanding of the relationship between the instance’s components (e.g. words in text, patches in an image) and the model’s prediction. We argue that explaining predictions is an important aspect in getting humans to trust and use machine learning effectively, if the explanations are faithful and intelligible. [...] An essential criterion for explanations is that they must be interpretable, i.e., provide qualitative understanding between the input variables and the response. We note that interpretability must take into account the user’s limitations³³⁴.

Em complemento, no famoso artigo de Sandra Wachter *et al.* – “Counterfactual Explanations Without Opening The Black Box: Automated Decisions And The GDPR”³³⁵ – é mencionado que a explicação de decisões automatizadas não depende necessariamente do entendimento do público em geral sobre como os sistemas algorítmicos funcionam. Argumentam ainda que, mesmo que tal possibilidade de interpretação do sistema seja de grande importância e deva ser perseguida, explicações podem, em princípio, ser oferecidas sem que seja necessária a abertura da caixa-preta³³⁶.

Por fim, a publicação “AI in UK” também carrega a discussão sobre a necessidade de se ter sistemas de inteligência artificial interpretáveis, analisando-se os métodos de transparência e explicação. Sobre o primeiro é previsto que, com a transparência técnica dos sistemas, *experts* poderiam compreender como o algoritmo foi desenvolvido, outorgando-lhes acesso ao código-fonte. Contudo, argumenta-se que esse método não necessariamente revelaria a razão pela qual o sistema tomou uma decisão em determinada situação, como também não daria a oportunidade de se avaliar os dados que foram

³³⁴ “Por ‘explicar uma previsão’, queremos dizer apresentar artefatos textuais ou visuais que fornecem compreensão qualitativa da relação entre os componentes inseridos no sistema (por exemplo, palavras de um texto, ou patches em uma imagem) e o resultado da previsão do modelo. Argumentamos que explicar as previsões é um aspecto importante da inteligência artificial, para fazer com que os humanos confiem e usem o aprendizado de máquina de forma eficaz, certos de que as explicações são fiéis e inteligíveis [...] Um critério essencial para as explicações é que elas devem ser interpretáveis, ou seja, fornecer entendimento qualitativo entre as variáveis de entrada e a resposta. Ressaltamos que a interpretabilidade deve levar em consideração as limitações do usuário.” (RIBEIRO, Marco Túlio; SINGH, Samee; GUESTIN, Carlos. “*Why Should I Trust You?*”: Explaining the Predictions of Any Classifier. KDD 2016 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: 1135–1144, 2016, p. 1135-1136, tradução nossa. Disponível em: https://www.researchgate.net/publication/305342147_Why_Should_I_Trust_You_Explaining_the_Predictions_of_Any_Classifier. Acesso em: 10 fev. 2022.)

³³⁵ MITTELSTADT, Brent; RUSSELL, Chris; WACHTER, Sandra. Counterfactual Explanations Without Opening The Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law & Technology*, v. 31, n. 2, 2018.

³³⁶ MITTELSTADT, Brent; RUSSELL, Chris; WACHTER, Sandra. Counterfactual Explanations Without Opening The Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law & Technology*, v. 31, n. 2, p. 843, 2018.

inseridos no sistema, ou como eles foram processados para alcançar tal decisão³³⁷. Com isso em vista, percebe-se que a conquista da completa transparência técnica é difícil, para não se dizer impossível, para certos tipos de sistemas de inteligência artificial já em uso atualmente, e, mesmo que o seu alcance fosse sempre possível, em muitos casos o método não seria apropriado ou útil³³⁸.

Sandra Wachter *et al.*, também nesse sentido, sustentam que:

In modern machine learning, the internal state of the algorithm can consist of millions of variables intricately connected in a large web of dependent behaviours. Conveying this state to a layperson in a way that allows them to reason about the behaviour of an algorithm is extremely challenging³³⁹.

Ainda em análise sobre o “AI in UK”, como uma alternativa, é avaliado o método de explicação, em que um sistema de inteligência artificial é desenvolvido de modo que ele seja passível de interpretação. Nas pesquisas realizadas para a elaboração do documento, foram identificadas várias iniciativas atualmente em desenvolvimento, que podem ajudar a traduzir a lógica do algoritmo para formas compreensíveis ao operador humano³⁴⁰. Entre outras grandes companhias, a Google, a IBM e a Microsoft foram entrevistadas e informaram estar trabalhando em mecanismos para tornar os algoritmos explicáveis, incluindo-se o desenvolvimento de sistemas como o Google Glassbox e o Microsoft Explainable Boosting Machine.

O ponto de vista desta autora é coincidente com o dos estudos aqui expostos, considerando que a transparência técnica pode não acudir o usuário leigo de inteligência artificial nas decisões controversas e que tenham reflexo no âmbito jurídico, sendo necessário sempre a análise de um *expert* – o que necessariamente encareceria e burocratizaria o processo de demonstração dos “porquês” de uma decisão automatizada. Ademais, a transparência pode ser de difícil aferição técnica e, também, não revelar

³³⁷ “One solution to the question of intelligibility is to try to increase the technical transparency of the system, so that experts can understand how an AI system has been put together. This might, for example, entail being able to access”. (REINO UNIDO. *AI in the UK: ready, willing and able?* House of Lords, Select Committee on Artificial Intelligence. p. 38, 2018. Disponível em: <https://www.parliament.uk/ai-committee/>. Acesso em: 09 fev. 2022.)

³³⁸ REINO UNIDO. *AI in the UK: ready, willing and able?* House of Lords, Select Committee on Artificial Intelligence. p. 38, 2018. Disponível em: <https://www.parliament.uk/ai-committee/>. Acesso em: 09 fev. 2022.

³³⁹ MITTELSTADT, Brent; RUSSELL, Chris; WACHTER, Sandra. Counterfactual Explanations Without Opening The Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law & Technology*, v. 31, n. 2, p. 845, 2018.

³⁴⁰ REINO UNIDO. *AI in the UK: ready, willing and able?* House of Lords, Select Committee on Artificial Intelligence. p. 39, 2018. Disponível em: <https://www.parliament.uk/ai-committee/>. Acesso em: 09 fev. 2022.

informações satisfatórias à necessidade de justificativa de certa decisão. E ainda que seja possível a abertura da caixa-preta e sua interpretação em relação à decisão em questão, isso poderia infringir os direitos de propriedade intelectual do titular do código-fonte do sistema. Este último fator é compreendido como o maior obstáculo para a viabilidade do método de transparência na prática, pois a sua aplicabilidade teria grande impacto no mercado de inovação, uma vez que os investimentos na criação de um sistema próprio poderiam não ser recompensados, na medida em que seriam facilmente copiáveis quando divulgados no intuito de se demonstrar as razões de uma determinada decisão automatizada.

Defende-se neste trabalho, portanto, o uso da explicabilidade em detrimento da transparência para que se alcance a responsabilidade e a prestação de contas almejada nos sistemas de inteligência artificial. Dado que a necessidade de explicabilidade e prestação de contas dos algoritmos foi colocada de forma recorrente como princípio para o uso e a disseminação da inteligência artificial, viu-se um aumento na preocupação científica de como tornar esses princípios aplicáveis na prática. Em pesquisa pela expressão “Explainable AI” no Google Trends de 2016, até atualmente, nota-se um crescimento considerável no interesse pela matéria:

Figura 2 – Retorno do *Google Trends* para pesquisa global pelo termo “Explainable AI”

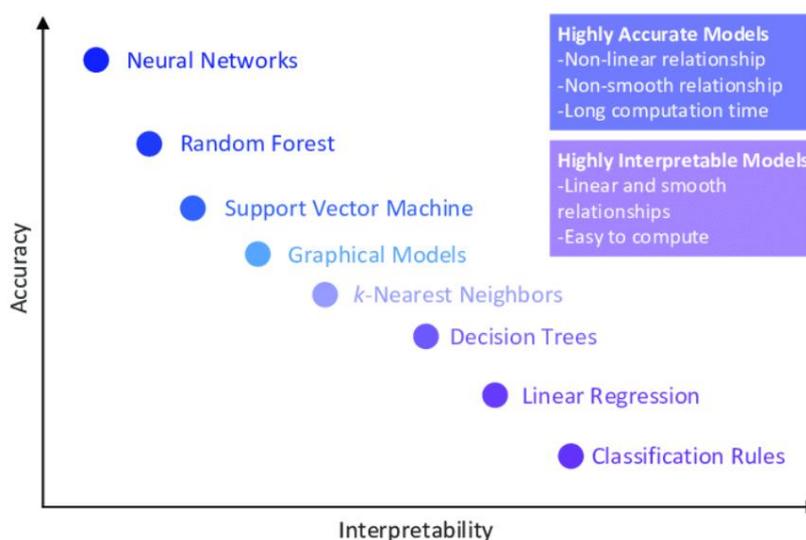


Fonte: *Google Trends*.

Resta compreender, então, como a explicabilidade tem sido colocada em prática, considerando os desafios dos algoritmos de *machine learning* mais complexos e a opacidade característica desses sistemas.

Em primeiro lugar, é preciso lembrar que alguns sistemas de aprendizado de máquina mais simples são explicáveis por natureza, enquanto outros representam grande desafio em razão da sua complexidade. Ocorre que quanto mais complexos, mais precisos são os modelos algorítmicos, e justamente em razão disso é que surge o interesse na capacidade de se explicar algoritmos complexos³⁴¹. A partir da explicabilidade, seria possível aproveitar da precisão provida por esses sistemas, sendo também possível explicá-los em caso de decisões que gerem danos. No gráfico abaixo está representada a situação de *trade-off* entre interpretabilidade e precisão dos sistemas de *machine learning*:

Figura 3 – Gráfico sobre a possibilidade de interpretação de um modelo algorítmico X precisão do modelo



Fonte: LEE, Haeyoung; MOROCHO-CAYAMCELA, Manuel Eugenio; LIM, Wansu (2019).

Pela análise do gráfico acima, conclui-se, então, que existem os programas de *machine learning* que são naturalmente inteligíveis e que, portanto, não exigem mecanismos adjacentes de explicação para sua compreensão. Os três primeiros tipos de *machine learning* identificados no canto inferior direito do gráfico acima (árvores de decisão, regressão linear e regras de classificação) são modelos que possuem o atributo chamado de *interpretability by design*, ou interpretabilidade intrínseca. Apesar dessa

³⁴¹ Para maiores informações sobre o *trade-off* entre previsão e interpretabilidade, vide: LEE, Haeyoung; MOROCHO-CAYAMCELA, Manuel Eugenio; LIM, Wansu. Machine Learning for 5G/B5G Mobile and Wireless Communications: Potential, Limitations, and Future Directions. *IEEE Access*, v. 7, Sept. 19, 2019. Disponível em: https://www.researchgate.net/publication/335937022_Machine_Learning_for_5GB5G_Mobile_and_Wireless_Communications_Potential_Limitations_and_Future_Directions. Acesso em: 10 fev. 2022.

facilidade na interpretação dos sistemas, para algumas funções mais complexas, esses algoritmos não são os mais recomendados, ou podem mesmo não ter qualquer utilidade. Para modelos algorítmicos mais intrincados há, por outro lado, os métodos de explicação *post-hoc*, que são aqueles utilizados para explicar o sistema de *machine learning* não inteligível, em razão da sua opacidade³⁴².

Outro tipo de diferenciação entre as possíveis formas de explicação de sistemas de inteligência artificial diz respeito ao método ser específico para um determinado tipo de *machine learning*, ou ser utilizado independentemente do modelo do algoritmo. É a diferença conhecida por *model specific* e *model agnostic*. No modelo específico, o sistema de explicação é apenas aplicável a um tipo de *machine learning*, por exemplo, redes neurais. No caso do método agnóstico, este pode ser utilizado para diferentes tipos de *machine learning*. Sendo assim, neste último caso, as ferramentas de explicação não têm acesso à estrutura e aos pesos conferidos a cada variável no sistema, mas apenas aos seus *inputs* e *outputs*³⁴³.

Além dessas diferenças, uma última qualidade dos métodos de explicação que deve ser exposta, para que se facilite a compreensão dos seus possíveis usos, é a distinção entre explicação local e explicação global, ou seja, entre as ferramentas que explicam uma decisão específica ou o modelo algorítmico de forma global³⁴⁴. Do ponto de vista de

³⁴² “Intrinsic or post hoc? This criteria distinguishes whether interpretability is achieved by restricting the complexity of the machine learning model (intrinsic) or by applying methods that analyze the model after training (post hoc). Intrinsic interpretability refers to machine learning models that are considered interpretable due to their simple structure, such as short decision trees or sparse linear models. Post hoc interpretability refers to the application of interpretation methods after model training. Permutation feature importance is, for example, a post hoc interpretation method. Post hoc methods can also be applied to intrinsically interpretable models. For example, permutation feature importance can be computed for decision trees”. / “Intrínseca ou post hoc? Esse critério distingue se a interpretabilidade é alcançada restringindo a complexidade do modelo de aprendizado de máquina (intrínseco) ou aplicando métodos que analisam o modelo após o treinamento (post hoc). A interpretabilidade intrínseca refere-se a modelos de aprendizado de máquina considerados interpretáveis devido à sua estrutura menos complexa, como árvores de decisão curtas, ou modelos lineares esparsos. A interpretabilidade post hoc refere-se à aplicação de métodos de interpretação após o treinamento do modelo.” (MOLNAR, Christoph. *Interpretable machine learning: A guide for making black boxes models explained*. Leanpub, 2019, p. 21, tradução nossa.)

³⁴³ MOLNAR, Christoph. *Interpretable machine learning: A guide for making black boxes models explained*. Leanpub, 2019, p. 22.

³⁴⁴ MOLNAR, Christoph. *Interpretable machine learning: A guide for making black boxes models explained*. Leanpub, 2019, p. 22.

viabilidade técnica, alguns estudos demonstram³⁴⁵ que a explicação local é mais viável do que a explicação global, o que se deve pelo fato de que os dados inseridos no sistema serem capazes de alterar de forma significativa os resultados de uma decisão automatizada, impedindo que haja uma explicação abstrata, que poderia ser usada em diversas circunstâncias³⁴⁶.

Com essa breve introdução a respeito dos tipos de explicação possíveis, compreende-se que o maior desafio em relação à interpretação de sistemas de *machine learning* encontra-se nos modelos complexos, que não são intrinsecamente inteligíveis. Em relação a estes, como as redes neurais, sabe-se que a explicação *post-hoc* já é possível por meio de métodos como o *Local interpretable model-agnostic explanations* – LIME³⁴⁷ e o *Shapley Values* – SHAP³⁴⁸.

Neste trabalho, no entanto, optou-se por focar na explicação pelo método contrafactual, que ganhou notoriedade para os estudantes de Direito e inteligência artificial, principalmente por ter sido o método eleito como o mais eficiente em termos de decisões que têm impacto no âmbito jurídico em renomados artigos publicados pela *Harvard Journal of Law & Technology*³⁴⁹ e pelo Berkman Klein Center³⁵⁰.

Abreviadamente, esse método pode ser explicado da seguinte forma: diante de um sistema caixa-preta em que apenas se conhecem os *inputs* e os *outputs*, a partir do

³⁴⁵ Alguns desses estudos são:

DOSHI-VELEZ, Finale; KORTZ, Mason. *Accountability of AI Under the Law: The Role of Explanation*. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper, 2017. Disponível em: <https://dash.harvard.edu/handle/1/34372584>. Acesso em: 10 fev. 2022.

CITRON, Danielle Keats; PASQUALE, Frank A. The Scored Society: Due Process for Automated Predictions. *Washington Law Review*, v. 89, p. 01, 2014. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2376209. Acesso em: 10 fev. 2022.

³⁴⁶ DOSHI-VELEZ, Finale; KORTZ, Mason. *Accountability of AI Under the Law: The Role of Explanation*. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper, p. 13, 2017. Disponível em: <https://dash.harvard.edu/handle/1/34372584>. Acesso em: 10 fev. 2022.

³⁴⁷ Para saber mais sobre esse tipo de explicação algorítmica, vide: RIBEIRO, Marco Túlio; SINGH, Samee; GUESTRIN, Carlos. “*Why Should I Trust You?*”: Explaining the Predictions of Any Classifier. *KDD 2016 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*: 1135–1144, 2016.

³⁴⁸ Para saber mais sobre esse tipo de explicação algorítmica, vide: MOLNAR, Christoph. *Interpretable machine learning: A guide for making black boxes models explained*. Leanpub, 2019.

³⁴⁹ MITTELSTADT, Brent; RUSSELL, Chris; WACHTER, Sandra. Counterfactual Explanations Without Opening The Black Box: Automated Decisions And The GDPR. *Harvard Journal of Law & Technology*, v. 31, n. 2, 2018.

³⁵⁰ DOSHI-VELEZ, Finale; KORTZ, Mason. *Accountability of AI Under the Law: The Role of Explanation*. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper, 2017. Disponível em: <https://dash.harvard.edu/handle/1/34372584>. Acesso em: 10 fev. 2022.

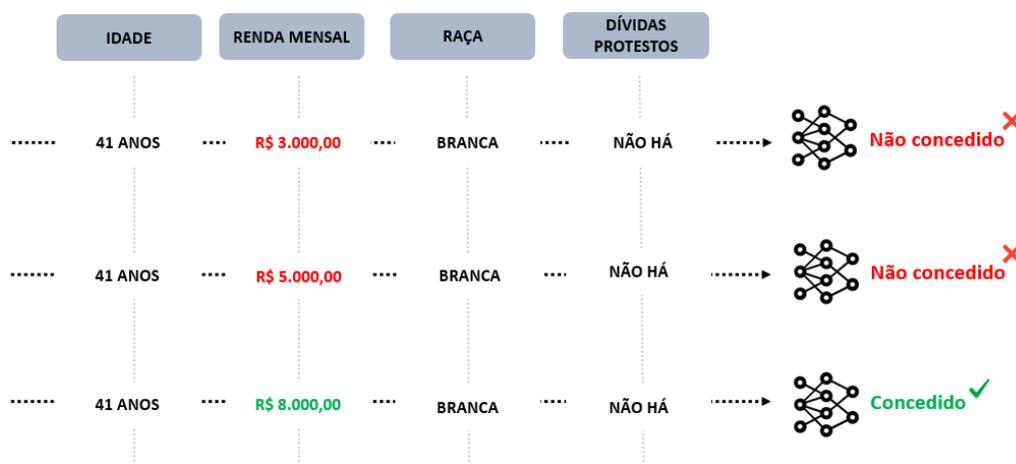
contrafactual, é possível fazer com que o sistema demonstre quais *inputs* devem ser alterados para que se alcance o resultado desejado. Dessa forma, é possível aferir quais foram os dados inseridos no sistema que, caso fossem alterados, levariam a resultados significativamente diferentes; conseqüentemente, se conheceria qual foi a variável que teve peso para uma determinada decisão, e, a partir daí, identificar-se-ia se a decisão foi juridicamente correta ou não. Segundo Wachter *et al.*, o modelo contrafactual pode ser demonstrado pela seguinte sentença: “You were denied a loan because your annual income was £30,000. If your income had been £45,000, you would have been offered a loan”³⁵¹. Enquanto outros métodos de explicação indicarão, de forma direta, o *input* que teve mais peso para uma decisão, o contrafactual informa o que fazer de forma diferente para alcançar um resultado desejado.

Para que se identifique o dado que teve maior relevância no resultado, é preciso que os testes sejam realizados com a menor perturbação possível dos dados inseridos no sistema. Em outras palavras, deve-se chegar ao *input* específico que gerou o *output* em questão, todo o mais constante, para que não haja dúvida que aquele específico dado gerou o resultado indesejado e que, à vista disso, a sua alteração faria com que fosse possível se alcançar o *output* desejado³⁵². Em outras palavras, contrafactual é o método pelo qual se busca alcançar o resultado desejado a partir da menor alteração possível nos *inputs* do sistema. Continuando no exemplo do empréstimo, o contrafactual poderia ser representado, de maneira bastante simplificada, da seguinte forma:

³⁵¹ “O empréstimo lhe foi negado, porque a sua receita anual é de £30,000. Caso a sua receita anual fosse de £45,000, você teria conseguido um empréstimo.” (MITTELSTADT, Brent; RUSSELL, Chris; WACHTER, Sandra. Counterfactual Explanations Without Opening The Black Box: Automated Decisions And The GDPR. *Harvard Journal of Law & Technology*, v. 31, n. 2, p. 844, 2018, tradução nossa.)

³⁵² “The concept of the ‘closest possible world’, or the smallest change to the world that can be made to obtain a desirable outcome, is key throughout the discussion of counterfactuals”. / “O conceito de ‘universo mais próximo possível’, ou a menor mudança possível no universo de *inputs* que pode ser feita para obter um resultado desejável, é fundamental em toda a discussão de contrafactuais.” (Counterfactual Explanations Without Opening The Black Box: Automated Decisions And The GDPR. *Harvard Journal of Law & Technology*, v. 31, n. 2, p. 845, 2018, tradução nossa.)

Figura 4 – Representação do contrafactual para explicação algorítmica



Fonte: Elaborado pela autora.

É claro que essa representação, assim como a explicação do contrafactual disposta acima, são amplamente simplificadas, uma vez que este não é um trabalho que pretende demonstrar a complexa matemática por trás dos sistemas de explicação, mas apenas identificar que a explicabilidade, tão cara à sociedade, já é viável na prática. Assim, o que se pretende demonstrar com a imagem acima é que, executando uma série de testes com a menor perturbação possível dos dados de entrada do sistema, é possível se ter uma visão bastante clara do limite entre uma decisão desejada ou indesejada (no caso exemplificado, entre se ter um empréstimo concedido ou não).

Ainda sobre o contrafactual, segundo Wachter *et al.*, esse tipo de explicação é intencionalmente restrito, de modo a prover o mínimo de informação capaz de alterar uma decisão de forma impactante. Assim, esse sistema não requer que o sujeito interessado em entender uma determinada decisão tenha que compreender a lógica algorítmica do sistema em termos técnicos de linguagem computacional³⁵³. Sintetizando as vantagens de tal método, os autores concluem:

Principally, counterfactuals bypass the substantial challenge of explaining the internal workings of complex machine learning systems. Even if technically feasible, such explanations may be of little practical value to data subjects. In contrast, counterfactuals provide information to the data subject that is both

³⁵³ Apesar de o contrafactual poder ser representado de forma simples e proporcional à fácil compreensão ao sujeito que requer a explicação, vale dizer que o método, na prática, é baseado em fórmulas matemáticas complexas. Para uma compreensão mais ampla sobre o tema, sob o ponto de vista da ciência computacional, com exemplos de fórmulas utilizadas, vide: MOLNAR, Christoph. *Interpretable machine learning: A guide for making black boxes models explained*. Leanpub, 2019.

easily digestible and practically useful for understanding the reasons for a decision, challenging them, and altering future behaviour for a better result³⁵⁴.

Reforçam esse entendimento Doshi-Velez e Kortz., que, traçando um paralelo com as explicações que são exigidas do humano de acordo com a legislação, afirmam que identificar o contrafactual não exige conhecimento sobre o fluxo de *bits* em um sistema de inteligência artificial, assim como explicações exigidas dos humanos também não requerem o conhecimento do fluxo das sinapses dos neurônios³⁵⁵.

Há, contudo, desafios a serem superados para que se possa aplicar o contrafactual de forma eficiente. Entre eles, cita-se o fato de que a inteligência artificial não conhece os conceitos dos seus *inputs*, tais como o humano os entende³⁵⁶. Ou seja, os dados inseridos em um sistema são lidos por este como uma sequência de *pixels*, e não como conceitos (por exemplo, gênero, sexo, idade, entre outros). Segundo Doshi-Velez e Kortz, apesar de este problema apresentar um obstáculo a mais, ele é contornável por meio de treinamento em massa dos algoritmos³⁵⁷. Tem-se aí um segundo problema em relação à aplicação da explicabilidade, considerando-se o atual estado da técnica: necessariamente, a disponibilização de explicações contrafactuais (ou por qualquer outro método disponível) encarecerá os sistemas de inteligência artificial, por demandarem engenheiros qualificados, grande disponibilidade de dados e treinamento em massa dos sistemas. Isso implica maior vantagem para companhias já consolidadas, aumentando ainda mais o seu poder e a disparidade de recursos quando comparadas com empresas menores de tecnologia. Outro possível lado negativo da necessidade de explicação é que, tendo em

³⁵⁴ “Principalmente, os contrafactuais contornam o desafio substancial de explicar o funcionamento interno de sistemas complexos de aprendizado de máquina. Mesmo que tecnicamente viáveis, tais explicações podem ter pouco valor prático para os titulares dos dados. Em contrapartida, os contrafactuais fornecem informações ao titular dos dados que são facilmente digeríveis e úteis, na prática, para entender os motivos de uma decisão, desafiá-los e alterar o comportamento futuro para um melhor resultado.” (MITTELSTADT, Brent; RUSSELL, Chris; WACHTER, Sandra. Counterfactual Explanations Without Opening The Black Box: Automated Decisions And The GDPR. *Harvard Journal of Law & Technology*, v. 31, n. 2, p. 860, 2018, tradução nossa.)

³⁵⁵ DOSHI-VELEZ, Finale; KORTZ, Mason. *Accountability of AI Under the Law: The Role of Explanation*. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper, p. 12, 2017. Disponível em: <https://dash.harvard.edu/handle/1/34372584>. Acesso em: 10 fev. 2022.

³⁵⁶ DOSHI-VELEZ, Finale; KORTZ, Mason. *Accountability of AI Under the Law: The Role of Explanation*. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper, p. 14, 2017. Disponível em: <https://dash.harvard.edu/handle/1/34372584>. Acesso em: 10 fev. 2022.

³⁵⁷ DOSHI-VELEZ, Finale; KORTZ, Mason. *Accountability of AI Under the Law: The Role of Explanation*. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper, p. 8, 2017. Disponível em: <https://dash.harvard.edu/handle/1/34372584>. Acesso em: 10 fev. 2022.

vista que isso implica custos, pode ser que os desenvolvedores optem por fazer sistemas menos sofisticados e precisos, mas com alta explicabilidade, conforme identificado na Figura 3³⁵⁸. Nesse ponto, é importante lembrar que existem autores que defendem a aplicabilidade apenas de algoritmos interpretáveis, como Cynthia Rudin, em seu artigo cujo título já indica o seu posicionamento: “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead”³⁵⁹. Segundo ela, em casos de aplicação de *machine learning* para decisões de alto impacto, a explicação *post hoc* deveria ser preterida em relação ao uso de modelos que são inerentemente interpretáveis³⁶⁰; no entanto, este trabalho discorda desse ponto de vista, por entender que, em razão das vantagens dos sistemas complexos relacionadas à sua maior precisão, deve haver incentivo à aplicação e aos avanços nas técnicas de explicação desses modelos.

Por fim, é importante destacar que não são todos os resultados de um sistema que precisam de explicação. De acordo com o “AI in UK”, as decisões automatizadas que precisam ser interpretáveis são as seguintes: (i) decisão judiciais; (ii) decisões relacionadas a diagnósticos e a tratamentos na área de saúde; (iii) decisões relacionadas a certos tipos de produtos e a serviços financeiros, tais como seguros e empréstimos; (iv) decisões de veículos autônomos; e (v) decisões de sistemas bélicos³⁶¹.

Doshi-Velez e Kortz, por sua vez, sugerem uma abordagem mais ampla – e, sob o ponto de vista desta autora, mais acertada – ao propor que se deve exigir a explicação dos sistemas nas mesmas situações legais em que ela é exigida para as decisões humanas. O estudo partiu da análise da legislação dos Estados Unidos da América, do Reino Unido, da França e da Alemanha, em que foram identificadas quatro situações nas quais as decisões são exigidas do humano. Essas situações são: decisões tomadas por juízes; decisões tomadas por agências administrativas ou reguladoras; determinadas decisões

³⁵⁸ DOSHI-VELEZ, Finale; KORTZ, Mason. *Accountability of AI Under the Law: The Role of Explanation*. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper, p. 10, 2017. Disponível em: <https://dash.harvard.edu/handle/1/34372584>. Acesso em: 10 fev. 2022.

³⁵⁹ “Parem de explicar modelos de aprendizado de máquina do tipo caixa-preta para decisões de alto risco e usem modelos interpretáveis em vez disso.” (RUDIN, Cynthia. Stop Explaining Black Box Machine Learning Models For High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, v. 1, p. 206-215, May 2019, tradução nossa)

³⁶⁰ RUDIN, Cynthia. Stop Explaining Black Box Machine Learning Models For High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, v. 1, p. 206, May 2019.

³⁶¹ REINO UNIDO. *AI in the UK: ready, willing and able?* House of Lords, Select Committee on Artificial Intelligence. p. 37, 2018. Disponível em: <https://www.parliament.uk/ai-committee/>. Acesso em 09 fev. 2022.

tomadas no âmbito privado em alguns mercados (por exemplo, explicações relacionadas à governança corporativa em uma empresa); e, por fim, as explicações exigidas de indivíduos, pontualmente, em caso de litígio³⁶². Por se tratar de hipóteses amplas em que as decisões são legalmente exigidas, este mesmo arcabouço pode ser aplicado em relação às situações nas quais a legislação brasileira exige explicações sobre decisões humanas, somando-se a estas a previsão legal de revisão de decisões automatizadas, prevista na Lei Geral de Proteção de Dados. Todas essas hipóteses englobam o tipo de decisão que importa a este trabalho, isto é, decisões de maior impacto, capazes de gerar dano a um indivíduo ou à coletividade.

3.6 REGULAÇÃO DA INTELIGÊNCIA ARTIFICIAL SOB A ÓTICA DA PREVENÇÃO

Por tudo que foi dito neste capítulo, não seria possível finalizá-lo sem mencionar que a regulação da inteligência artificial, tal como estudada neste trabalho, aponta para a aplicação da responsabilidade civil sob suas outras funções, que não a puramente reparatória. Longe de ser a pretensão deste trabalho esgotar o tema, mas, antes, com o intuito de abrir os caminhos para próximos trabalhos, conclui-se que o princípio da *accountability* e seus deveres anexos, consubstanciados na transparência, na explicabilidade e em outras obrigações informacionais e colaborativas (tal como a elaboração de relatórios de impacto) trazem luz às medidas precaucionárias e preventivas para lidar com o risco e com a incerteza gerados pelo emprego da inteligência artificial.

Segundo Nelson Rosenvald, “a prevenção é o cerne da responsabilidade civil contemporânea”³⁶³. Analisando o tema no contexto da Sociedade do Risco, conforme delineada por Ulrich Beck, Rosenvald afirma que a responsabilidade de outrora, de roupagem individualista, reativa e patrimonialista, propende, na contemporaneidade, a

³⁶² DOSHI-VELEZ, Finale; KORTZ, Mason. *Accountability of AI Under the Law: The Role of Explanation*. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper, p. 8-12, 2017. Disponível em: <https://dash.harvard.edu/handle/1/34372584>. Acesso em: 10 fev. 2022.

³⁶³ ROSENVALD, Nelson. *As funções da responsabilidade civil: a reparação e a pena civil*. 3. ed. São Paulo: Saraiva, 2017, p. 96.

um viés preventivo por razões éticas, comportamentais e econômicas³⁶⁴. Passa-se a ter, então, a responsabilidade civil como mecanismo de “prevenção de comportamentos e não apenas de danos”³⁶⁵.

Para Marcelo Milagres, “já se foi a época em que o estudo da responsabilidade civil se limitava ao prejuízo de ordem material ou à função reparatória/compensatória”³⁶⁶, o que, ainda segundo ele, “não significa, por si só, a despatrimonialização da responsabilidade civil ou mesmo responsabilidade sem dano”³⁶⁷.

No contexto digital de emprego da inteligência artificial, Ana Frazão *et. al.* asseveram que “a velocidade das transformações tecnológicas, a capacidade de propagação de informações e a dificuldade na contenção do fluxo de dados” levam a ainda maiores desafios à aplicação da lógica repressiva da responsabilidade civil³⁶⁸.

Assim, acredita-se que, além da sua vocação retrospectiva³⁶⁹, a responsabilidade civil sob o viés preventivo passaria a imprimir novas obrigações aos desenvolvedores, fornecedores e operadores dos sistemas de IA, no sentido de impor a “necessidade de tomada de medidas concretas, adequadas e proporcionais aos riscos gerados, com vistas à evitabilidade do dano, bem como à adoção de técnicas para comprovar efetivamente que assim agiu”³⁷⁰. Dessa forma, a partir da adoção de medidas relacionadas a deveres de informação e cooperação, os agentes envolvidos com o desenvolvimento, fornecimento e operação dos sistemas de inteligência artificial, de forma colaborativa com órgãos reguladores e com a sociedade de modo geral, passariam a fazer melhor julgamento quanto ao custo-benefício dos empregos de determinada tecnologia.

³⁶⁴ ROSENVALD, Nelson. *As funções da responsabilidade civil: a reparação e a pena civil*. 3. ed. São Paulo: Saraiva, 2017, p. 27-28.

³⁶⁵ ROSELVALD, Nelson. A polissemia da responsabilidade civil na LGPD. Migalhas de Proteção de Dados, 2020. Disponível em: <https://www.migalhas.com.br/coluna/migalhas-de-protecao-de-dados/336002/a-polissemia-da-responsabilidade-civil-na-lgpd>. Acesso em 10 fev. 2022.

³⁶⁶ MILAGRES, Marcelo de Oliveira. Breves notas sobre a (des)patrimonialização da responsabilidade civil: ainda a fundamentalidade do dano. In: MILAGRES, Marcelo. ROSENVALD, Nelson (Coord.). *Responsabilidade Civil: Novas Tendências*. 2. ed. Indaiatuba: Editora Foco, 2018, p. 189.

³⁶⁷ MILAGRES, Marcelo de Oliveira. Breves notas sobre a (des)patrimonialização da responsabilidade civil: ainda a fundamentalidade do dano. In: MILAGRES, Marcelo. ROSENVALD, Nelson (Coord.). *Responsabilidade Civil: Novas Tendências*. 2. ed. Indaiatuba: Editora Foco, 2018, p. 195.

³⁶⁸ FRAZÃO, Ana *et al.* Compliance de dados pessoais. In: FRAZÃO, Ana. TEPENDINO, Gustavo, OLIVA, Milena Donato (Coord.). *Lei Geral de Proteção de Dados e suas repercussões no direito brasileiro*. São Paulo: Thomson Reuters Brasil, 2019, p. 681.

³⁶⁹ ROSENVALD, Nelson. *As funções da responsabilidade civil: a reparação e a pena civil*. 3. ed. São Paulo: Saraiva, 2017. p. 32.

³⁷⁰ BUSATTA, Eduardo Luiz. Do dever de prevenção em matéria de proteção de dados pessoais. In: JÚNIOR, Marcos Ehrhardt Júnior; CATALAN, Marcos; MALHEIROS, Pablo (Coord.). *Direito Civil e Tecnologia*. Belo Horizonte: Forum, 2020, p. 42.

Especificamente sobre o princípio da *accountability* na regulação da inteligência artificial, Rosenvald anota que:

A "accountability", amplia o espectro da responsabilidade civil, mediante a inclusão de parâmetros regulatórios preventivos, que promovem uma interação entre a *liability* do Código Civil com uma regulamentação voltada à inserção de regras de boas práticas que estabeleçam procedimentos, normas de segurança e padrões técnicos³⁷¹.

Como um elemento a mais para se enriquecer este debate, não se pode deixar de mencionar o princípio da precaução, que ganhou forte notoriedade após a Conferência das Nações Unidas sobre o Meio Ambiente e o Desenvolvimento, ocorrida no Rio de Janeiro, em 1992, que culminou na publicação da Declaração do Rio sobre Meio Ambiente e Desenvolvimento, O princípio 15 dessa Declaração prevê que:

Com o fim de proteger o meio ambiente, o princípio da precaução deverá ser amplamente observado pelos Estados, de acordo com suas capacidades. Quando houver ameaça de danos graves ou irreversíveis, a ausência de certeza científica absoluta não será utilizada como razão para o adiamento de medidas economicamente viáveis para prevenir a degradação ambiental³⁷²

A despeito de algumas críticas no sentido de que a aplicação dessa diretriz no âmbito da tecnologia refrearia a inovação³⁷³, tal princípio, assim como propugna Stirling, não se reduz a determinar que certa atividade que envolva grave risco ou incerteza seja cessada. Ao contrário, o que se espera é que os desenvolvedores de tais atividades assumam deveres relacionados à *accountability*, identificando possíveis ações mitigatórias à atividade de risco (“it is about reinforcing qualities of understanding deliberation, and accountability, rather than just the stringency of the resulting actions”)³⁷⁴.

Inserindo tais lições no contexto da inteligência artificial, Bruno Bioni aposta que a Lei Geral de Proteção de Dados funcionará como feixe de entrada para o princípio da

³⁷¹ ROSENVALD, Nelson. A falácia da responsabilidade subjetiva na regulação da IA. Migalhas, 2022. Disponível em: [³⁷² NAÇÕES UNIDAS. Declaração do Rio sobre Meio Ambiente e Desenvolvimento, 1992. Disponível em: <http://www.onu.org.br/rio20/img/2012/01/rio92.pdf>. Acesso em: 25 jun. 2022.](https://www.migalhas.com.br/coluna/migalhas-de-protecao-de-dados/365913/a-falacia-da-responsabilidade-subjetiva-na-regulacao-da-ia#:~:text=O%20PL%2021%2F20201,artigo%206%C2%BA%2C%20que%20%22Art. Acesso em: 30 jul. 2022.</p>
</div>
<div data-bbox=)

³⁷³ SUSTEIN, Cass R. Beyond the Precautionary Principle. Law and Economics Working Paper, Vol 149, Chicago Unbound, 2013. Disponível em: <https://dash.harvard.edu/bitstream/handle/1/12876710/Beyond%20the%20Precautionary%20Principle.pdf?sequence=1&isAllowed=y>. Acesso em: 30 jun. 2022.

³⁷⁴ STIRLING, Andrew. Precaution in the Governance of Technology. Working Paper, Science Policy Research Unit, Brighton, 2016. p. 649

precaução na regulação de IA no Brasil, tendo em vista que muitas aplicações da tecnologia se alimentam de dados pessoais³⁷⁵. Os deveres de *accountability* e os relatórios de impacto previstos na LGPD, segundo ele, refletem o princípio da precaução na sua conotação de deliberação pública, na medida em que têm o condão de reduzir a assimetria informacional entre operadores e titulares de dados, possibilitando uma “empreitada regulatória cada vez mais colaborativa com quem está no ‘chão de fábrica’ e, sobretudo, que contas sejam prestadas [...] acerca das ações tomadas para avaliar e gerir os riscos em jogo”³⁷⁶.

Tal matéria – seja pelo viés da função preventiva da responsabilidade, seja pelo viés do princípio da precaução – contudo, dada sua complexidade e acirrados debates acerca dos seus contornos, merece trabalho próprio, ficando aqui apenas o aceno para futura pesquisa.

³⁷⁵ BIONI, Bruno; 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 (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Revista dos Tribunais, 2019, p. 215.

³⁷⁶ BIONI, Bruno; 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 (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Revista dos Tribunais, 2019, p. 216.

CONCLUSÃO

Em meio às promessas e aos perigos da inteligência artificial, às suas vantagens e aos seus riscos, entre seus fascínios e seus desencantos, este trabalho buscou situar a responsabilidade civil em relação aos casos de danos gerados pela tecnologia, observado o crescente uso de sistemas de *machine learning*.

Foi verificado nítido embate entre a conduta emergente da inteligência artificial e os elementos conformadores da responsabilidade civil, o que se observa em razão do esmorecimento do nexo de causalidade entre a conduta do humano envolvido no desenvolvimento, fornecimento ou operação do sistema e o dano gerado pela conduta emergente deste. Inobstante a conduta lícita do agente humano, o sistema, por si só, é capaz de gerar danos em razão do aprendizado de máquina, faltando-lhe, contudo, os requisitos de imputabilidade.

Apesar desse patente conflito entre a tecnologia e a responsabilidade civil, constatou-se não ser possível estabelecer uma resposta única e genérica para as dúvidas acerca de quem responde, ou sob quais fundamentos responde em caso de danos gerados pela inteligência artificial autônoma, haja vista a grande heterogeneidade de aplicações e riscos proporcionados pela tecnologia. Dessa premissa, decorreram as principais conclusões deste trabalho.

Em que pese não ser possível o estabelecimento de um regime único de responsabilidade civil para todas as situações de danos envolvendo inteligência artificial, vislumbrou-se a possibilidade de criação de diretrizes que servirão como guia para a regulação e aferição de responsabilidade civil nesses casos. Assim, recomendou-se a abordagem principiológica da regulação (aqui considerada como atividade intencional do Estado, visando moldar determinada atividade) sobre responsabilidade civil e inteligência artificial, o que se daria por meio do estabelecimento de um colchão de princípios aplicáveis de forma genérica, amparados, quando necessário, por microssistemas normativos setoriais, que regeriam a responsabilidade civil em determinados ramos de aplicação da IA que demandem regras específicas.

Observou-se que essa abordagem tem grande aceitação na doutrina e que já vem sendo aplicada por alguns entes reguladores, inclusive nos países líderes em desenvolvimento da inteligência artificial, a saber, Estados Unidos, China e Reino Unido. Analisando-se os regulamentos já existentes nesses países, e também no Brasil, concluiu-

se pela reincidência na previsão dos princípios de *accountability*, *explainability* e *transparency*, neste estudo traduzidos como “responsabilidade e prestação de contas”, “explicabilidade” e “transparência”, respectivamente.

Passou-se, então, à análise crítica desses princípios correlacionando-os à responsabilidade civil. De forma sintética, identificou-se a *accountability* como a necessidade de se demonstrar que certa tarefa é realizada a contento, seguindo parâmetros predeterminados. A partir disso, concluiu-se que a explicabilidade e a transparência são qualidades dos algoritmos capazes de proporcionar a referida prestação de contas.

Depreendeu-se da análise mais detida sobre a transparência e a explicabilidade, que a primeira pode ser compreendida como o desvelamento do código-fonte, com o intuito de se descobrir a lógica adjacente a uma determinada decisão automatizada, enquanto a explicabilidade independe da abertura da “caixa-preta” do algoritmo para que sejam alcançadas as respostas necessárias quanto a tais decisões, o que é feito por meio da análise de *inputs* e *outputs* do programa. Em razão disso, concluiu-se que a explicabilidade deve ser privilegiada em detrimento da transparência, vez que esta última implicaria possível infração a direitos de propriedade intelectual e conseqüente desincentivo à inovação. Além disso, considerando-se o público leigo, outorgar acesso ao código-fonte não significa, necessariamente, conceder esclarecimento inteligível quanto à decisão questionada.

Como importante conclusão deste trabalho, mencionou-se que não são todas as decisões algorítmicas que merecem ser explicadas, mas apenas aqueles de alto impacto. Sem a pretensão de esgotar o tema, cogitou-se como um possível parâmetro para identificação de quais decisões algorítmicas merecem explicação aquelas decisões em que a legislação demandaria explicação caso fossem tomadas pelo humano. De forma ampla, essas situações são as seguintes: decisões tomadas por juízes; decisões tomadas por agências administrativas ou reguladoras; determinadas decisões tomadas no âmbito corporativo privado (como aquelas relacionadas à governança corporativa); e as decisões tomadas por indivíduos que culminam em litígio por, supostamente, infringir direito de outro.

Muito distante de oferecer respostas definitivas, mas antes com o intuito de fomentar o debate a respeito do tema, concluiu-se neste trabalho que o princípio da “*accountability*” e as suas formas de concretização, consubstanciadas, principalmente, na explicabilidade e na transparência dos sistemas de IA, orientam para a aplicação do princípio da precaução e da responsabilidade civil sob a sua função preventiva. Por meio

da instituição de deveres de informação, de colaboração, de transparência, entre outras obrigações atribuídas aos desenvolvedores/fornecedores da IA – além da exigência de observância a parâmetros técnicos predefinidos – o princípio da *accountability* aponta no sentido de prevenção às condutas que poderiam levar a danos, a partir da coparticipação da sociedade e dos agentes envolvidos no desenvolvimento e operação da inteligência artificial nas escolhas regulatórias e na análise custo-benefício quanto ao emprego de determinada tecnologia de inteligência artificial.

REFERÊNCIAS

2020 IBM ANNUAL REPORT. Disponível em: <https://www.ibm.com/annualreport/>. Acesso em: 05 out. 2021.

ACHTEN, Nele *et al.* *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI*. Berkman Klein Center Research. Publicação n. 2020-1, 2020. Disponível em: <https://ssrn.com/abstract=3518482>. Acesso em: 10 fev. 2022.

ALEMANHA. *Artificial Intelligence Strategy of the German Federal Government*. Governo Federal Alemão, 2020, tradução nossa. Disponível em: https://knowledge4policy.ec.europa.eu/ai-watch/germany-ai-strategy-report_en. Acesso em: 10 fev. 2022.

ALVES, Tárík César Oliveira e *et al.* *Policy Paper: Projeto de Lei do Senado n. 5.051/2019, que “estabelece os princípios para o uso da inteligência artificial no Brasil”*. Centro de Pesquisa em Direito Tecnologia e Inovação – DTIBR, 2019. Disponível em: <https://www.dtibr.com/livros-e-artigos>. Acesso em: 10 fev. 2022.

ALVIM, Agostinho. *Da inexecução das obrigações e suas consequências*. 5. ed. São Paulo: Saraiva, 1980.

ASARO, Peter. The Liability Problem for Autonomous Artificial Agents. *AAAI Spring Symposium Series*, mar. 2016. Disponível em: <https://www.aaai.org/ocs/index.php/SSS/SSS16/paper/view/12699>. Acesso em: 23 maio 2022.

ANALYTICS INSIGHT. *Artificial Intelligence Investment by Top 10 Countries*. 2021. Disponível em: <https://www.analyticsinsight.net/artificial-intelligence-investment-by-top-10-countries/>. Acesso em: 10 fev. 2022.

ARTIFICIAL INTELLIGENCE INDEX REPORT 2021. Stanford University Human-Centered Artificial Intelligence, p. 16. Disponível em: https://aiindex.stanford.edu/wp-content/uploads/2021/03/2021-AI-Index-Report_Master.pdf. Acesso em: 05 out. 2021.

BARBOSA, Mafalda Miranda. O futuro da responsabilidade civil desafiada pela inteligência artificial: as dificuldades dos modelos tradicionais e caminhos de solução. *Revista de Direito da Responsabilidade*, p. 284, 2020. Disponível em: <https://revistadireitoresponsabilidade.pt/2020/o-futuro-da-responsabilidade-civil-desafiada-pela-inteligencia-artificial-as-dificuldades-dos-modelos-tradicionais-e-caminhos-de-solucao-mafalda-miranda-barbosa/> Acesso em: 23 maio 2022.

BECK, Ulrich. *Sociedade de Risco: Rumo a uma outra modernidade*. Tradução de Sebastião Nascimento. 2. ed. São Paulo: Editora 34, 2011.

BENJAMIN, Antonio Heman de Vasconcellos. Fato do produto ou serviço. In: BENJAMIN, Antonio Heman de Vasconcellos; BESSA, Leonardo Roscoe; MARQUES,

Cláudia Lima. *Manual de Direito do Consumidor*. 9. ed. São Paulo: Thomson Reuters, 2020.

BELLMAN, Richard. *An Introduction to Artificial Intelligence: Can Computers Think?* Boyd & Fraser Publishing Company: 1978.

BIONI, Bruno; 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 (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo, Revista dos Tribunais, 2019, p. 209-231.

BODEN, Margaret. *Inteligência Artificial: uma brevíssima introdução*. Tradução Fernando Santos. São Paulo: Editora Unesp, 2020.

BOSTROM, Nick. *Superinteligência: caminhos, perigos e estratégias para um novo mundo*. Rio de Janeiro: DarkSide, 2014.

BRASIL, Assembleia Legislativa. Lei 13.709/2018. Lei Geral de Proteção de Dados. Disponível em <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm> Acesso em: 10 fev. 2022.

BRASIL. Câmara dos Deputados. *Projeto de Lei nº 21*, de 04 de fevereiro de 2020. Justificativa. Brasília, Câmara dos Deputados, 2020. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 10 fev. 2022.

BRASIL. Câmara dos Deputados. *Projeto de Lei nº 21-A*, de 29 de setembro de 2021. Brasília: Câmara dos Deputados, 2021. Disponível em: <https://www.camara.leg.br/propostas-legislativas/2236340>. Acesso em: 10 fev. 2022.

BRASIL. *Lei n. 8.078*, de 11 de setembro de 1990. Institui o Código de Defesa do Consumidor. Brasília, 1990. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/18078compilado.htm. Acesso em: 23 maio 2022.

BRASIL. *Lei n. 10.406*, de 10 de janeiro de 2002. Institui o Código Civil. Brasília, 2002. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/2002/110406.htm. Acesso em: 23 maio 2022.

BRASIL. Senado Federal. *Projeto de Lei do Senado nº 5.051*, de 16 de setembro de 2019. Estabelece os princípios para o uso da Inteligência Artificial no Brasil. Brasília: Senado Federal, 2019. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/138790>. Acesso em: 10 fev. 2022.

BRASIL. Senado Federal. *Projeto de Lei do Senado nº 5.691*, de 25 de outubro de 2019. Institui a Política Nacional de Inteligência Artificial. Brasília: Senado Federal, 2019. Disponível em: <https://www25.senado.leg.br/web/atividade/materias/-/materia/139586>. Acesso em: 10 fev. 2022.

BRASIL. *Portaria nº 4.617*, de 06 de abril de 2021. Institui a Estratégia Brasileira de Inteligência Artificial e seus eixos temáticos. Brasília: Ministério da Ciência, Tecnologia

e Inovações, 2021, p. 6. Disponível em: <https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/inteligencia-artificial>. Acesso em: 10 fev. 2022.

BRIZ, Jaime Santos. *La Responsabilidad Civil*, p. 42 apud PEREIRA, Caio Mário da Silva. *Responsabilidade Civil*. 9. ed. Rio de Janeiro: Forense, 1999.

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

BRYNJOLFSSON, Erik; MCAFEE, Andrew. *Second Machine Age*. Nova York: W.W. Norton & Company, 2014.

BUSATTA, Eduardo Luiz. Do dever de prevenção em matéria de proteção de dados pessoais. In: JÚNIOR, Marcos Ehrhardt Júnior; CATALAN, Marcos; MALHEIROS, Pablo (Coord.). *Direito Civil e Tecnologia*. Belo Horizonte: Forum, 2020.

BUSUIOC, Madalina. Accountable Artificial Intelligence: Holding Algorithms to Account. *Public Administration Review*, Florida, v. 81, 2020.

CASALONE, Carlo *et al.* *Human-centric AI: from principles to actionable and shared policies*. G20 AI Principles. Sept. 30, 2021. Disponível em: https://www.g20-insights.org/policy_briefs/human-centric-ai-from-principles-to-actionable-and-shared-policies/. Acesso em: 10 fev. 2022.

CASTRO, Daniel; CHIVOT, Eline; MCLAUGHLIN, Michael. *Who is Winning the AI Race: China, EU or the United States?* Center for Data Innovation. 2019. Disponível em: <https://datainnovation.org/2019/08/who-is-winning-the-ai-race-china-the-eu-or-the-united-states/>. Acesso em: 10 fev. 2022.

CATH, Corinne *et al.* *Artificial Intelligence and the “Good Society”*: the US, EU, and UK approach. *Sci Eng Ethics Vol.* **24**, p. 505–528, 2018.

CELLAN-JONES, Nicholas Rory. *Uber's self-driving operator charged over fatal crash*. BBC News, 2020. Disponível em: <https://www.bbc.com/news/technology-54175359>. Acesso em 05 out. 2021.

ČERKA, Paulius; GRIGIENĖ, Jurgita; SIRBIKYTĖ, Gintarė. Liability for damages caused by Artificial Intelligence. *Computer Law & Security Review*, v. 31, n. 3, p. 376-389, 2015. Disponível em: https://is.muni.cz/el/1422/podzim2017/MV735K/um/ai/Cerka_Grigiene_Sirbikyte_Liability_for_Damages_caused_by_AI.pdf. Acesso em: 23 maio 2022.

CHARNIAK, Eugene; MCDERMOTT, Drew. *Introduction to Artificial Intelligence*, 1985. Addison-Wesley Publishing Company.

CHINA. *Artificial Intelligence Standardization White Paper*. Instituto de Padronização Eletrônica da China. Trad. Etcetera Language Group, Inc. p. 2, 2018. Disponível em: <https://cset.georgetown.edu/publication/artificial-intelligence-standardization-white-paper/>. Acesso em: 10 fev. 2022.

CHINA. *New Generation of Artificial Intelligence Development Plan*. Conselho do Estado Chinês. p. 6, 2017. Disponível em: <https://www.newamerica.org/cybersecurity-initiative/digichina/blog/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/>. Acesso em: 10 fev. 2022.

CHOPRA, Samir; WHITE, Laurence F. *A Legal Theory for Autonomous Artificial Agents*. Ann Arbor: University of Michigan Press, 2011.

CITRON, Danielle Keats; PASQUALE, Frank A. The Scored Society: Due Process for Automated Predictions. *Washington Law Review*, v. 89, 2014. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2376209. Acesso em: 10 fev. 2022.

COSTA, Isac Silveira da. *High frequency trading (HFT) é uma forma de insider trading?* In: PARENTONI, Leonardo (Coord.). *Direito, Tecnologia e Inovação*: volume 1. Belo Horizonte: D'Plácido, 2018.

DARRACH, Brad. *Meet Shaky, the first eletronic person*. The fascinating and fearsome reality of a machine with a mind of its own. *Life Magazine*, 1970, Nov. 20. Disponível em:

<https://books.google.com.br/books?id=2FMEAAAAMBAJ&printsec=frontcover&hl=pt-BR&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false>. Acesso em 27 jun. 2022.

DELTEK. *Federal Artificial Intelligence Landscape*. GovWin IQ. 2022. Disponível em: <https://info.deltek.com/Federal-Artificial-Intelligence-Landscape-2022-Summary-GovWin-Deltek>. Acesso em: 10 fev. 2022.

DIAS, Daniel. Implementação de seguro obrigatório de responsabilidade civil no contexto da inteligência artificial. In: SILVA, Rodrigo da Guia; TEPEDINO, Gustavo (Coord.). *O Direito Civil na era da inteligência artificial*. São Paulo: Thomson Reuters Brasil, 2020.

DIAS, José de Aguiar. *Da responsabilidade civil*. 11. ed. Atualizada de acordo com o Código Civil de 2002 e aumentada por Rui Belford Dias. Rio de Janeiro: Renovar, 2006.

DING, Jeffrey. *Deciphering China's AI Dream: The context, components, capabilities, and consequences of China's strategy to lead the world in AI*. Future of Humanity Institute: Universidade de Oxford, 2018.

DOMINGOS, Pedro. *The master algorithm: How the quest for the ultimate learning machine will remake our world*. Nova York: Basic Books, 2015.

DOSHI-VELEZ, Finale; KORTZ, Mason. *Accountability of AI Under the Law: The Role of Explanation*. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society Working Paper, 2017, tradução nossa. Disponível em: <https://dash.harvard.edu/handle/1/34372584>. Acesso em: 10 fev. 2022.

EASTERBROOK, Frank Hoover. *Cyberspace and the Law of the Horse*. Chicago: University of Chicago Legal Forum, 1996. Disponível em <

https://chicagounbound.uchicago.edu/cgi/viewcontent.cgi?referer=&httpsredir=1&article=2147&context=journal_articles> Acesso em 10 fev. 2022.

ESTADOS UNIDOS DA AMÉRICA. *Four Principles of Explainable Artificial Intelligence*. Instituto Nacional de Normas e Tecnologia, Departamento de Comércio do Governo Federal, 2020. Disponível em: <https://www.nist.gov/publications/four-principles-explainable-artificial-intelligence-draft>. Acesso em: 10 fev. 2022.

ESTADOS UNIDOS DA AMÉRICA. *Guidance for Regulation of Artificial Intelligence Applications*. Memorandum for the Heads of Executive Departments and Agencies, p. 4, tradução nossa. Disponível em: <https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf>. Acesso em: 10 fev. 2022.

ESTADOS UNIDOS DA AMÉRICA. *House Resolution 2231*, de 10 de abril de 2019. House of Representatives. Algorithmic Accountability Act. Disponível em: <https://www.congress.gov/bill/116th-congress/housebill/2231>. Acesso em: 10 fev. 2022.

ESTADOS UNIDOS DA AMÉRICA. *National Artificial Intelligence Research and Development Strategic Plan*. National Science and Technology Council. p. 21, 2019, tradução nossa. Disponível em: <https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf>. Acesso em: 10 fev. 2022.

ESTADOS UNIDOS DA AMÉRICA. *Preparing for the Future of Artificial Intelligence*. National Science and Technology Council. p. 17, 2016, tradução nossa. Disponível em: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf. Acesso em: 10 fev. 2022.

FARIAS, Cristiano Chaves de; ROSENVALD, Nelson; NETTO, Felipe Peixoto Braga. *Curso de direito civil: responsabilidade civil*. 5. ed. Salvador: Juspodivm, 2018.

FERNANDES, André Lucas. *PL da Inteligência Artificial erra ao criar regime de responsabilidade subjetiva*. Jota, 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/artigos/pl-21-2020-inteligencia-artificial-01112021>. Acesso em: 10 fev. 2022.

FERREIRA, Matheus Costa. *Progresso digital exponencial: o direito no ponto de inflexão*. In: PARENTONI, Leonardo (Coord.). *Direito, Tecnologia e Inovação: volume 1*. Belo Horizonte: D'Plácido, 2018.

FENWICK, Mark; KAAL, Wulf A.; VERMEULEN, Erik. Regulation Tomorrow: What Happens When Technology is Faster than the Law? *American University Business Law Review*, vol. 6, n. 3, p. 24, 2017, tradução nossa. Disponível em: <https://ssrn.com/abstract=2834531>. Acesso em: 09 fev. 2022.

FILHO, Carlos Edison do Rêgo Monteiro; ROSENVALD, Nelson. Responsabilidade Civil indireta e inteligência artificial. In: BARBOSA, Mafalda Miranda *et al.* (Coord.). *Direito Digital e Inteligência Artificial: diálogos entre Brasil e Europa*. São Paulo: Editora Foco, 2021.

FILHO, Carlos Edison do Rêgo Monteiro; ROSENVALD, Nelson. Riscos e responsabilidades na inteligência artificial e noutras tecnologias digitais emergentes. In:

O Direito Civil na era da inteligência artificial. São Paulo: Thomson Reuters Brasil, 2020.

FILHO, Calixto Salomão. Regulamentação da atividade empresarial para o desenvolvimento. In: FILHO, Calixto Salomão (Org.). *Regulação e desenvolvimento: novos temas*. São Paulo: Editora Malheiros, 2012.

FILHO, Sérgio Cavalieri. *Programa de Responsabilidade Civil*. 9. ed. São Paulo: Atlas, 2010.

FRAZÃO, Ana et al. Compliance de dados pessoais. In: FRAZÃO, Ana. TEPENDINO, Gustavo, OLIVA, Milena Donato (Coord.). *Lei Geral de Proteção de Dados e suas repercussões no direito brasileiro*. São Paulo: Thomson Reuters Brasil, 2019.

FREITAS, Wallace Almeida de. Artificial Intelligence and Solvable Civil Responsibility. In: CARDOSO, Renato César; PARENTONI, Leonardo Netto (Org.). *Law, Technology and Innovation – v. II: Insights on Artificial Intelligence and the Law*. Belo Horizonte: Editora Expert, 2021, p. 291-308. Disponível em: <https://www.dtibr.com/livros-e-artigos>. Acesso em: 23 maio 2022.

GASSER, Urs; ALMEIDA, Virgílio Augusto Fernandes. A Layered Model for AI Governance. *IEEE Internet Computing*, v. 21, n. 6, p. 3-4, 2017. Disponível em: <https://dash.harvard.edu/handle/1/34390353>. Acesso em: 08 maio 2021.

GIBBS, Samuel. Google buys UK artificial intelligence startup Deepmind for £400m. *The Guardian*, 2014. Disponível em: <https://www.theguardian.com/technology/2014/jan/27/google-acquires-uk-artificial-intelligence-startup-deepmind>. Acesso em: 10 fev. 2022.

GOPANI, Avi. Meet The New Player In The US-China AI Arms Race: United Kingdom. *Analytics India Magazine*, 2021. Disponível em: <https://analyticsindiamag.com/meet-the-new-player-in-the-us-china-ai-arms-race-united-kingdom/>. Acesso em: 10 fev. 2022.

GUSTIN, Miracy Barbosa de Sousa; DIAS, Maria Tereza Fonseca. *(Re)pensando a pesquisa jurídica: teoria e prática*. 4. ed. Belo Horizonte: Del Rey, 2013.

HAUGELAND, John. *Artificial Intelligence: The Very Idea*. 1985. MIT Press: Cambridge.

HUBBARD, F. Patrick. Allocating the risk of physical injury from “sophisticated robots”: Efficiency, fairness, and innovation. In: CALO, Ryan; FROMKIN, A. Michael; KERR, Ian. *Robot Law*. Northampton: Edward Elgar Pub, 2016.

INTELIGÊNCIA ARTIFICIAL In: DICIONÁRIO online de Oxford. Disponível em: <https://www.oxfordreference.com/view/10.1093/oi/authority.20110803095426960>. Acesso em: 05 out. 2021.

JAPÃO. *Social Principles of Human-Centric Artificial Intelligence*. Conselho para Ciência, Tecnologia e Inovação, p. 7, 2019, tradução nossa. Disponível em: <https://www8.cao.go.jp/cstp/english/humancentricai.pdf>. Acesso em: 10 fev. 2022.

JARVIS, David; LOUCKS, Jeff; HUPFER, Susanne; MURPHY, Tomothy. *Future in the balance?* How countries are pursuing an AI advantage. Deloitte Insights, 2019. Disponível em: <https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/ai-investment-by-country.html>. Acesso em: 10 fev. 2022.

KAPLAN, Jerry. *Artificial Intelligence: What everyone needs to know*. Oxford: Oxford University Press, 2016.

KARNOW, Curtis E. A. The application of traditional tort theory to embodied machine intelligence. In: CALO, Ryan; FROOMKIN, A. Michael; KERR, Ian. *Robot Law*. Northampton: Edward Elgar Pub, 2016.

KURZWEIL, Raymond. *The Age of Intelligent Machines*, 1990. MIT Press: Cambridge.

KURZWEIL, Raymond. *The Singularity is Near: When humans transcend biology*. Nova York: Penguin Books, 2005.

LEE, Haeyoung; MOROCHO-CAYAMCELA, Manuel Eugenio; LIM, Wansu. Machine Learning for 5G/B5G Mobile and Wireless Communications: Potential, Limitations, and Future Directions. *IEEE Access*, v. 7, Sept. 19, 2019. Disponível em: https://www.researchgate.net/publication/335937022_Machine_Learning_for_5GB5G_Mobile_and_Wireless_Communications_Potential_Limitations_and_Future_Directions. Acesso em: 10 fev. 2022.

LEE, Kai-Fu. *AI Superpowers: China, Silicon Valley, and the new world order*. Nova York: Houghton Mifflin Harcourt, 2018.

LEMOES, Ronaldo. Estratégia Brasileira de IA é patética. *Folha de S.Paulo*, 2021. Disponível em: <https://www1.folha.uol.com.br/colunas/ronaldolemos/2021/04/estrategia-de-ia-brasileira-e-patetica>. Acesso em: 10 fev. 2022.

LIKENS, Scott; ROA, Anand; SHEBAB, Michael. *AI Predictions 2021*. PWC, 2020. Disponível em: <https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions.html>. Acesso em: 05 out. 2021.

LODGE, Martin; KOOP, Christel. What is regulation? An interdisciplinary concept analysis. *Regulation and Governance*, v. 11, p. 5, 14 Jul. 2015.

LOUKIDES, Mike; LORICA, Ben. *What's Artificial Intelligence?* Sebastopol: O'Reilly Media, 2016.

MAGRANI, Eduardo; SILVA, Priscilla; VIOLA, Rafael. Novas perspectivas sobre ética e responsabilidade de inteligência artificial. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Thomsom Reuters Brasil, 2019.

MARQUES, Cláudia Lima. Campo de aplicação do CDC. In: BENJAMIN, Antonio Heman de Vasconcellos; BESSA, Leonardo Roscoe; MARQUES, Cláudia Lima. *Manual de Direito do Consumidor*. 9. ed. São Paulo: Thomson Reuters, 2020.

MCCARTHY, John; MINSKY, Leon; ROCHESTER, Nathanael; SHANNON, Claude. *A proposal for the Dartmouth Summer Research Project on Artificial Intelligence*. 1955. Disponível em: <http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf>. Acesso em: 05 out. 2021.

MCCULLOCH, Warren. S.; PITTS, Walter. *A logical calculus of the ideas immanent in nervous activity*. Bulletin of Mathematical Biophysics. vol. 5, p. 115-133 (1943).

MEDON, Filipe. *Inteligência Artificial e Responsabilidade Civil: autonomia, riscos e solidariedade*. Editora JusPodivm: São Paulo, 2022.

MENDES, Pedro Manuel Pimenta. Inteligência Artificial e Responsabilidade Civil: as possíveis “soluções” .do ordenamento jurídico português. In: BARBOSA, Mafalda Miranda *et al.* (Coord.) *Direito Digital e Inteligência Artificial: diálogos entre Brasil e Europa*. São Paulo: Editora Foco, 2021.

MILAGRES, Marcelo de Oliveira. A robótica e as discussões sobre a personalidade eletrônica. In: JÚNIOR, Marcos Ehrhardt Júnior; CATALAN, Marcos; MALHEIROS, Pablo (Coord.). *Direito Civil e Tecnologia*. Belo Horizonte: Forum, 2020.

MILAGRES, Marcelo de Oliveira. Breves notas sobre a (des)patrimonialização da responsabilidade civil: ainda a fundamentalidade do dano. In: MILAGRES, Marcelo. ROSENVALD, Nelson (Coord.). *Responsabilidade Civil: Novas Tendências*. 2ª ed. Indaiatuba: Editora Foco, 2018.

MITTELSTADT, Brent; RUSSELL, Chris; WACHTER, Sandra. Counterfactual Explanations Without Opening The Black Box: Automated Decisions and the GDPR. *Harvard Journal of Law & Technology*, v. 31, n. 2, 2018.

MOLNAR, Christoph. *Interpretable machine learning: A guide for making black boxes models explained*. Leanpub, 2019.

MOORE. Gordon E. *Cramming more components onto integrated circuits*. Electronics Magazine, vol. 38, p. 2, 1965. Tradução nossa. Disponível em: <https://newsroom.intel.com/wp-content/uploads/sites/11/2018/05/moores-law-electronics.pdf>. Acesso em: 05 out. 2021.

MORAES, Maria Celina Bodin de. *Danos à pessoa humana: uma leitura civil-constitucional dos danos morais*. Rio de Janeiro: Renovar, 2003.

MOORE, Samuel; SCHNEIDER, David; STRICKLAND, Eliza. *How Deep Learning Works: Inside the neural networks that power today's AI*. IEEE Spectrum, 2021. Disponível em: <https://spectrum.ieee.org/what-is-deep-learning>. Acesso em 27 jun. 2022.

MULHOLLAND, Caitlin Sampaio. *A responsabilidade civil por presunção de causalidade*. Rio de Janeiro: GZ Editora, 2010.

MULHOLLAND, Caitlin. *Responsabilidade civil e processos decisórios autônomos em sistemas de Inteligência Artificial (IA): autonomia, imputabilidade e responsabilidade*. In: FRAZÃO, Ana. MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo, Revista dos Tribunais, 2019.

NAÇÕES UNIDAS. Declaração do Rio sobre Meio Ambiente e Desenvolvimento, 1992. Disponível em: <http://www.onu.org.br/rio20/img/2012/01/rio92.pdf>. Acesso em: 25 jun. 2022.

NATURE INDEX. *Top 25 countries/territories in artificial intelligence*. Nature Index 2020 Artificial Intelligence. 2020. Disponível em: <https://www.natureindex.com/supplements/nature-index-2020-ai/tables/countries>. Acesso em: 10 fev. 2022.

NILSSON, Nils. John. *Artificial Intelligence: A New Synthesis*. 1998. Morgan Kaufmann Publishers: San Francisco.

OCDE. *Recommendation of the Council on Artificial Intelligence*. Secretary-General of the OECD, 2019. Disponível em: <https://oecd.ai/en/ai-principles>. Acesso em: 10 fev. 2022.

PAGALLO, Ugo. *The Laws of Robots: Crimes, Contracts, and Torts*. Heidelberg: Springer Netherlands, 2013.

PADRÃO, Vinicius; SOUZA, Carlos Affonso Souza. *Responsabilidade civil e os “novos turcos”*: o papel da conduta humana nos danos relacionados à inteligência artificial, p. 664. In: SILVA, Rodrigo da Gaia; TEPEDINO, Gustavo (Coord.). *O Direito Civil na era da inteligência artificial*. 1 ed. São Paulo: Thomson Reuters Brasil, 2020.

PEIXOTO, Fabio Hartmann; COUTINHO, Marina de Alencar Araripe. *Inteligência Artificial e Regulação: uma análise do Projeto de Lei 5.051/2019*. *Revista Em Tempo*, [S.l.], v. 19, n. 1, ago. 2020. Disponível em: <https://revista.univem.edu.br/emtempo/article/view/3129>. Acesso em: 10 fev. 2022.

PEREIRA, Caio Mário da Silva. *Responsabilidade Civil*. 9. ed. Rio de Janeiro: Forense, 1999.

PRICEWATERHOUSECOOPERS GLOBAL. *Sizing the price: What’s the real value of AI for your business and how can you capitalise?* p. 8, 2017. Disponível em: <https://www.pwc.com/gx/en/issues/analytics/assets/pwc-ai-analysis-sizing-the-prize-report.pdf>. Acesso em: 10 fev. 2021.

PRICE II, William Nicholson. *Artificial Intelligence in Health Care: Applications and Legal Issues*. *U of Michigan Public Law Research Paper*, Michigan, n. 599, nov. 2017. Disponível em: <https://ssrn.com/abstract=3078704>. Acesso em: 05 out. 2021.

POLIDO, Fabrício Bertini Pasquot. *Novas perspectivas para a regulação da Inteligência Artificial: diálogos entre as políticas domésticas e os processos legais transnacionais*. In:

FRAZÃO, Ana; MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Revista dos Tribunais, 2019.

POOLE, David; MACKWORTH, Alan; GOEBEL, Randy. *Computational intelligence: A logical approach*. 1998. Oxford University Press: New York.

REINO UNIDO. *AI in the UK: ready, willing and able?* House of Lords, Select Committee on Artificial Intelligence, p. 113, 16 Apr. 2018. Disponível em: <https://www.parliament.uk/ai-committee/>. Acesso em: 09 fev. 2022.

REINO UNIDO. *National AI Strategy*. Office for Artificial Intelligence, p. 52, Sept. 2021. Disponível em: <https://www.gov.uk/government/publications/national-ai-strategy>. Acesso em: 10 fev. 2022.

REINO UNIDO. *Robotics and artificial intelligence*. House of Commons. 2016. Disponível em: <https://publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/14501.htm>. Acesso em: 10 fev. 2022.

RIBEIRO, Júlia Melo Carvalho. Inteligência artificial e responsabilidade civil sob a perspectiva do direito do consumidor. In: CHAVES, Natália Cristina; COLOMBINI, Henry (Org.). *Direito e Tecnologia: Novos modelos e tendências*. Porto Alegre: Editora Fi, 2021.

RIBEIRO, Júlia Melo Carvalho; PENNA, Thomaz Murta. A (des)regulação da tecnologia blockchain: uma análise da experiência regulatória brasileira. In: PARENTONI, Leonardo; MILAGRES, Marcelo de Oliveira; VAN DE GRAAF, Jeroen (Coord.); MOREIRA, Arthur Salles de Paula; CHAGAS, Ciro Costa; SANTANA, Mariana Damiani (Org.). *Direito, Tecnologia e Inovação – Vol. III: aplicações jurídicas de blockchain*. Belo Horizonte: Editora Expert, 2021.

RIBEIRO, Marco Túlio; SINGH, Samee; GUESTRIN, Carlos. “*Why Should I Trust You?*”: Explaining the Predictions of Any Classifier. KDD 2016 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: 1135–1144, 2016, p. 1135-1136, Disponível em: https://www.researchgate.net/publication/305342147_Why_Should_I_Trust_You_Explaining_the_Predictions_of_Any_Classifier. Acesso em: 10 fev. 2022.

RICH, Elaine; KNIGHT, Kevin. *Artificial Intelligence*. 2. ed. 1991. McGraw-Hill Education: New York.

RUSSEL, Stuart; NORVIG, Peter. *Artificial intelligence: a modern approach*. 3. ed. Nova Jersey: Prentice Hall, 2010.

RICHARDS, Neil M.; SMART, Willian D. How should the law think about robots? In: CALO, Ryan; FROOMKIN, A. Michael; KERR, Ian. *Robot Law*. Northhampton: Edward Elgar Pub, 2016.

ROBERTS, Huw *et al.* The Chinese approach to artificial intelligence: an analysis of policy, ethics, and regulation. *AI & Society Review*, 2 Mar. 2021. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3469784. Acesso em: 10 fev. 2022.

RODRIGUES, Silvio. *Direito Civil: Responsabilidade Civil*. São Paulo: Saraiva, 2002.

ROSEVALD, Nelson. A falácia da responsabilidade subjetiva na regulação da IA. Migalhas, 2022. Disponível em: <https://www.migalhas.com.br/coluna/migalhas-de-protecao-de-dados/365913/a-falacia-da-responsabilidade-subjetiva-na-regulacao-da-ia#:~:text=O%20PL%2021%2F2021,artigo%206%C2%BA%2C%20que%20%22Art.> Acesso em 30 jul. 2022.

ROSELVALD, Nelson. A polissemia da responsabilidade civil na LGPD. Migalhas de Proteção de Dados, 2020. Disponível em: <https://www.migalhas.com.br/coluna/migalhas-de-protecao-de-dados/336002/a-polissemia-da-responsabilidade-civil-na-lgpd>>. Acesso em 10 fev. 2022.

ROSEVALD, Nelson. As funções da responsabilidade civil: a reparação e a pena civil; 3. Ed. São Paulo: Saraiva, 2017.

RUDIN, Cynthia. Stop Explaining Black Box Machine Learning Models For High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, v. 1, p. 206-215, May 2019.

SAMUELSON, Pamela. Five Challenges for Regulating the Global Information Society. *Social Science Research Network*, California, 9 Aug. 2000. Disponível em: <https://ssrn.com/abstract=234743>. Acesso em: 09 fev. 2022.

SCHERER, Matthew U. *Regulating Artificial Intelligence Systems*. Harvard Journal of Law & Technology, v. 29, n. 2, p. 358, 2016. Disponível em: <https://ssrn.com/abstract=2609777>. Acesso em: 08 de maio de 2021.

SCHREIBER, Anderson. *Novos paradigmas da Responsabilidade Civil: da erosão dos filtros da reparação à diluição dos danos*. 6. ed. VitalSource Bookshelf, Grupo GEN, 2015.

SCHREIBER, Anderson. *PL da Inteligência Artificial cria fratura no ordenamento jurídico brasileiro*. Jota, 2021. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/coluna-do-anderson-schreiber/pl-inteligencia-artificial-cria-fratura-no-ordenamento-juridico-02112021>. Acesso em: 10 fev. 2022.

SCHUMPETER, Joseph Alois. *Capitalismo, Socialismo e Democracia*. São Paulo: Editora Unesp, 2017.

SILVA, Rodrigo da Guia; TEPEDINO, Gustavo. Inteligência Artificial e elementos da responsabilidade civil. In: FRAZÃO, Ana; MULHOLLAND, Caitlin (Coord.). *Inteligência artificial e direito: ética, regulação e responsabilidade*. São Paulo: Thomsom Reuters Brasil, 2019.

SOLUM, Lawrence. *Legal Personhood for Artificial Intelligences*. North Carolina Law Review, v. 70, 1992. Disponível em: <http://scholarship.law.unc.edu/nclr/vol70/iss4/4>. Acesso em: 23 mai. 2022.

SOUZA, Eduardo Nunes de. Personalidade jurídica e inteligência artificial. In: SILVA, Rodrigo da Guia; TEPEDINO, Gustavo (Coord.). *O Direito Civil na era da inteligência artificial*. São Paulo: Thomson Reuters Brasil, 2020.

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

STIRLING, Andrew. Precaution in the Governance of Technology. Working Paper, Science Policy Research Unit, Brighton, 2016.

STOCO, Rui. *Tratado de responsabilidade civil: doutrina e jurisprudência*. 7. ed. São Paulo: Revista dos Tribunais, 2007.

STONE, Peter; BROOKS, Rodney; BRYNJOLFSSON, Erik; CALO, Ryan; ETZIONI, Oren; HAGER, Greg; HIRSCHBERG, Julia; KALYANAKRISHNAN, Shivaram; KAMAR, Ece. KRAUS, Sarit; LEYTON-BROWN, Kevin; PARKES, David; PRESS, William; SAXENIAN, AnnaLee; SHAH, Julie; TAMBE, Milind; e TELLER, Astro. *Artificial Intelligence and Life in 2030*. One Hundred Year Study on Artificial Intelligence: Report of the 2015-2016 Study Panel, Stanford University, Stanford, CA, 2016. Disponível em: <https://ai100.stanford.edu/2016-report> . Acesso em: 05 out. 2021.

SU, Jean Baptiste. *Venture Capital Funding For Artificial Intelligence Startups Hit Record High In 2018*. Forbes, 2019. Disponível em: <https://www.forbes.com/sites/jeanbaptiste/2019/02/12/venture-capital-funding-for-artificial-intelligence-startups-hit-record-high-in-2018/> . Acesso em: 05 out. 2021.

SUSTEIN, Cass R. Beyond the Precautionary Principle. Law and Economics Working Paper, Vol 149, Chicago Unbound, 2013. Disponível em: <https://dash.harvard.edu/bitstream/handle/1/12876710/Beyond%20the%20Precautionary%20Principle.pdf?sequence=1&isAllowed=y>>. Acesso em 30 jun. 2022.

TAULLI, Tom. *Introdução à Inteligência Artificial*. Tradução A Press Media, LLC. São Paulo: Novatec, 2020.

TORTOISE MEDIA. *The Global AI Index*. 2021. Disponível em: <https://www.tortoisemedia.com/intelligence/global-ai/>. Acesso em: 10 fev. 2022.

TURING, Alan M. *Computing Machinery and Intelligence*. In Mind, vol. 49, p. 433-460. 1950

TURING, Alan M. *Intelligent Machinery*. National Physical Laboratory. 1948.

UNIÃO EUROPEIA. Parlamento Europeu. *Disposições de Direito Civil sobre robótica*. Resolução do Parlamento Europeu, de 16 de fevereiro de 2017, que contém

recomendações à Comissão sobre disposições de Direito Civil sobre robótica (2015/2103(INL)). Estrasburgo, 16 fev. 2017. Disponível em: http://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_PT.html. Acesso em: 23 maio 2022.

UNIÃO EUROPEIA. Parlamento Europeu. Regime de responsabilidade civil aplicável à inteligência artificial. Resolução do Parlamento Europeu, de 20 de outubro de 2020, que contém recomendações à Comissão sobre o regime de responsabilidade civil aplicável à inteligência artificial (2020/2014(INL)). Bruxelas, 20 out. 2017. Disponível em: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_PT.html. Acesso em: 08 de jun. de 2021.

UNIÃO EUROPEIA. Regulamento (UE) n° 2016/679 do Parlamento Europeu e do Conselho, de 23 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). *Jornal Oficial da União Europeia*, Estrasburgo, 04/05/2016, . Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016R0679&from=PT>. Disponível em: 10 fev. 2022.

VLADÉCK, David C. Machines Without Principals: Liability Rules and Artificial Intelligence. *Washington Law Review*, v. 89, n. 1, p. 146-147, 2014.

WESENDONCK, Tula. A responsabilidade civil pelos danos decorrentes dos riscos do desenvolvimento dos produtos postos em circulação. In: MILAGRES, Marcelo; ROSENVALD, Nelson (Coord.). *Responsabilidade Civil: novas tendências*. 2. ed. Indaiatuba: Editora Foco, 2018.

WESENDONCK, Tula. Inteligência Artificial e Responsabilidade Civil pelos Riscos do Desenvolvimento: um estudo comparado entre as propostas de regulamentação da matéria na união Europeia e o ordenamento vigente brasileiro. In: BARBOSA, Mafalda Miranda *et al.* (Coord.). *Direito Digital e Inteligência Artificial: diálogos entre Brasil e Europa*. São Paulo: Editora Foco, 2021.

WESTRHEIDE, Fabian. China: The First Artificial Intelligence Superpower. *Forbes*, 2020, Disponível em: <https://www.forbes.com/sites/cognitiveworld/2020/01/14/china-artificial-intelligence-superpower/?sh=7f4050152f05>. Acesso em: 10 fev. 2022.

WHITTY, Jonathan; MCGRATH, Stephen Keith. Accountability and responsibility defined. *International Journal of Managing Projects in Business*, v. 11, p. 687-707.

WINSTON, Patrick Henry. *Artificial Intelligence*. 3. ed., 1992. Addison-Wesley Publishing Company.

XAVIER, Fábio Correa. A Estratégia Brasileira de Inteligência Artificial. *MIT Technology Review*, 2021. Disponível em: <https://mittechreview.com.br/a-estrategia-brasileira-de-inteligencia-artificial/#:~:text=Segundo%20o%20MCTIC%2C%20a%20Estrat%C3%A9gia,e%20em%20prol%20de%20>. Acesso em: 10 fev. 2022.

ZIMMERMAN, Evan J. *Machine Minds: Frontiers in Legal Personhood*. 2015. Disponível em: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2563965. Acesso em: 08 de maio de 2021.