



RISCO DO DESENVOLVIMENTO: O PROBLEMA DO ALINHAMENTO DE VALORES E A IMPREVISIBILIDADE DOS COMPORTAMENTOS DA INTELIGÊNCIA ARTIFICIAL

*Alan Araújo de Sousa*¹

RESUMO

O presente artigo objetiva explorar as possíveis classificações legais, à luz do Código de Defesa do Consumidor, dos atos imprevisíveis da inteligência artificial que provocam danos aos consumidores. A metodologia utilizada baseia-se em uma pesquisa qualitativa, bibliográfica e documental, aplicando-se os métodos indutivo, dedutivo e hipotético-dedutivo. O estudo apresentará, inicialmente, noções gerais sobre a inteligência artificial, salientando os novos riscos gerados por essa tecnologia, com enfoque na imprevisibilidade dos seus atos e na dificuldade de harmonizar os seus objetivos com os interesses dos seres humanos. Para melhor visualização da problemática, serão utilizados como exemplo os carros autônomos. Na segunda parte do artigo, será apresentado ao leitor o conceito de risco do desenvolvimento, bem como serão expostos aspectos jurídicos que permitem classificá-lo como um defeito de concepção à luz do Código de Defesa do Consumidor. Ao final, o estudo evidencia a possibilidade de classificar como risco de desenvolvimento os atos imprevisíveis dessa tecnologia e que causam danos. O artigo conclui que os atos da inteligência artificial, ainda que imprevisíveis, são condicionados pelos objetivos programados na máquina, devendo serem alcançados pelo risco da atividade empresarial e, se causarem danos, classificados como defeito de concepção.

Palavras-chave: Inteligência artificial. Carros autônomos. Atos imprevisíveis. Risco do desenvolvimento. Responsabilidade civil.

¹ Mestrando em Direito Civil pela Pontifícia Universidade Católica de São Paulo. E-mail: alan.asousa.adv@gmail.com.



1 INTRODUÇÃO

As preocupações relacionadas à implementação em larga escala de máquinas dotadas de inteligência artificial vêm se consolidando na sociedade contemporânea, à medida em que essa tecnologia se torna cada vez mais presente em nossa realidade.

Os debates sobre as repercussões da adoção de produtos e serviços que operam por meio de um sistema inteligência artificial não são, contudo, recentes. Há muito tempo obras literárias e cinematográficas fomentam discussões acerca dos impactos positivos e negativos dessa tecnologia nos mais variados aspectos da vida humana. A título exemplificativo, pode-se mencionar a inquietação em relação ao progresso tecnológico retratada no romance *Frankenstein*, também conhecido como *O Prometeu Moderno*, de Mary Shelley; no livro *Eu, Robô*, de Isaac Asimov; na série televisiva *Black Mirror*, criada por Charlie Brooker; e no filme *AI – Inteligência Artificial*, dirigido pelo renomado Steven Spielberg.

As obras artísticas que tratam dessa temática refletem, em linhas gerais, o temor de criar máquinas com inteligências artificiais prejudiciais ao ser humano. A preocupação é bem sintetizada nas célebres Leis da Robótica, também conhecidas como Leis de Asimov: (1) um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano seja ferido; (2) um robô deve obedecer às ordens dadas por seres humanos, exceto nos casos em que tais ordens entrem em conflito com a primeira lei; (3) um robô deve proteger sua própria existência, desde que esta proteção não entre em conflito com a primeira ou segunda lei; e (0) um robô não pode causar mal à humanidade ou, por omissão, permitir que a humanidade sofra algum mal.

Todas essas histórias refletem a relevância de criação de sistemas de inteligência artificial cujos objetivos estejam em harmonia com os interesses ou valores humanos. Do contrário, consequências nefastas poderão advir dessa tecnologia.

No campo jurídico, o avanço de produtos e serviços que operam por meio de inteligência artificial também tem originado inúmeros debates e propostas legislativas a nível mundial². Interessa, no presente estudo, examinar apenas as repercussões dessa tecnologia no

² Na União Europeia, a discussão se consolidou em 2015, por ocasião da apresentação do relatório preliminar com recomendações acerca da legislação civil e robótica (*Draft Report with recommendations on civil law rules and robotics*), o qual inspirou a Resolução de 16 de fevereiro de 2017, com recomendações à Comissão de Direito Civil sobre Robótica (“Disposições de Direito Civil sobre Robótica”). Posteriormente, foram produzidos outros relatórios, como o Relatório de Orientações Éticas para uma Inteligência Artificial de Confiança, de 08 de abril de 2019, do Grupo de Peritos de Alto Nível em Inteligência Artificial, que, posteriormente, serviu de base para a Resolução de 20 de outubro de 2020, do Parlamento Europeu, que trata do regime de responsabilidade civil aplicável à inteligência artificial. No Brasil, podem ser citados os Projetos de Lei de nº 5.051, de 2019; nº 21, de 2020; e nº 872, de 2021, todos do Senado Federal, os quais estão instruídos e serão apreciados com o



âmbito da responsabilidade civil nas relações de consumo, em especial as decorrentes dos comportamentos imprevisíveis da inteligência artificial de produtos e serviços que geram situações de risco e danos aos consumidores.

Com esse escopo, o artigo apresentará, inicialmente, noções gerais sobre a inteligência artificial, expondo um breve histórico de seu conceito, forma de funcionamento, riscos gerados, entre outros aspectos. Para melhor compreensão desta tecnologia, serão trabalhadas questões relacionadas aos carros autônomos, pois se estima que serão os primeiros produtos verdadeiramente autônomos.

Em seguida, será fornecida uma base teórica acerca do risco do desenvolvimento, a fim de viabilizar, na última etapa do estudo, a qualificação legal dos atos imprevisíveis da inteligência artificial que provocam danos.

2 INTELIGÊNCIA ARTIFICIAL E CARROS AUTÔNOMOS

Alan Turing foi a primeira pessoa a expor uma concepção de inteligência artificial, o que fez em seu conhecido ensaio “*Computing machinery and intelligence*”, publicado no ano de 1950. No referido artigo, o usualmente considerado “pai da computação” propunha que, em vez de nos perguntarmos se as máquinas podem pensar, deveríamos questionar se as máquinas são capazes de passar em um teste de inteligência comportamental, que posteriormente ficou conhecido como *teste de Turing*. Em síntese, um programa de computador é aprovado no teste se um interrogador humano, após propor algumas perguntas por escrito, não conseguir descobrir se as respostas foram formuladas por uma pessoa ou uma máquina (NORVIG, RUSSEL, 2022)³.

A expressão *inteligência artificial*, todavia, somente foi oficialmente adotada na Conferência de Dartmouth College de 1956, nos Estados Unidos da América (EUA), organizada por John McCarthy, que sugeriu a realização de um estudo com duração de dois meses, conduzido com a conjectura básica de que as características da inteligência poderiam

relatório final da comissão de juristas responsável pela elaboração do denominado Marco Legal da Inteligência Artificial, entregue àquela casa legislativa em dezembro de 2022.

³ Explicando mais detalhadamente, o teste foi proposto sob a forma de um *jogo da imitação* (“*The imitation game*”, nas exatas palavras de Alan Turing), que envolve três participantes: uma pessoa e um programa de computador, cujas identidades permanecem ocultas a um interrogador humano, que deve manter uma breve conversa com os interrogados apenas por meio de mensagens digitadas. O objetivo do programa de computador é enganar o interrogador sobre sua identidade, ao passo que o objetivo da pessoa interrogada é ajudar o interrogador na tarefa de identificação dos participantes, apenas fazendo afirmações verdadeiras. Após alguns minutos, o interrogador humano deve responder qual participante é um programa e qual participante é uma pessoa. Caso o programa de computador consiga enganar seu interrogador, será aprovado no teste (TURING, 1950).



ser descritas com tamanha precisão a ponto de viabilizar a construção de uma máquina para simulá-la. Assim, a máquina poderia usar linguagem, formar abstrações, conceitos, aperfeiçoar-se e resolver todos os tipos de problemas que eram reservados aos seres humanos (NORVIG, RUSSEL, 2022). A ideia central era a de que a inteligência artificial corresponderia à “capacidade de uma máquina realizar tarefas, que, acaso fossem executadas por um ser humano, seriam consideradas ‘inteligentes’” (MEIRELLES, SANTANA, 2022, p. 7).

Conforme expõem Russel e Norvig (2022), historicamente as pesquisas têm adotado diferentes definições de inteligência artificial, por vezes relacionando a inteligência à fidelidade ao pensamento e/ou comportamento humano, ora definindo em termos de pensar e/ou agir racionalmente (fazer a “coisa certa”). Existem, pois, diferentes abordagens de inteligência artificial, que, desse modo, não tem um conceito amplamente aceito na comunidade científica⁴.

Não há neste estudo predileção por alguma das abordagens aludidas. Interessa mais, para o fim deste artigo, esclarecer brevemente a diferença crucial entre os programas de computador tradicionais e aqueles dotados de tecnologia de inteligência artificial.

De fato, na programação tradicional, um computador somente consegue realizar tarefas e alcançar objetivos cujas etapas e ações estiverem prévia e pormenorizadamente descritas por seu programador, por meio de algoritmos, que consistem, em linhas gerais, em uma sequência de instruções que informa o computador o que fazer (PIRES, SILVA, 2017).

O advento da inteligência artificial deu origem a uma revolução tecnológica, na medida em que os programas de computador ganharam a habilidade de acumular experiências próprias e delas extrair aprendizado (*machine learning*⁵), para tornar suas ações futuras mais eficientes (PIRES, SILVA, 2017). Assim, máquinas que possuem tecnologia de inteligência artificial são capazes de realizar atos autônomos, alcançando resultados criativos e, por vezes, não previstos por seus programadores (PIRES, SILVA, 2017).

É nesse contexto que se inserem os veículos autônomos, que podem ser compreendidos como um tipo de transporte que opera por meio de um sistema computacional, integrado a um conjunto de sensores e atuadores, e que é apto a realizar o transporte de pessoas ou bens, de forma parcial ou totalmente autônoma, isto é, com ou sem nenhum

⁴ Recomenda-se, neste ponto, a leitura da síntese elaborada por Thatiane Cristina Fontão Pires e Rafael Peteffi da Silva (2017) acerca da concepção de inteligência artificial adotada por Stuart J. Russel e Peter Norvig.

⁵ Para aprofundamento do estudo do aprendizado de máquina, conferir “NORVIG, Peter; RUSSEL, Stuart J. Aprender a partir de exemplos. In: *Inteligência artificial: uma abordagem moderna*. 4. ed. Rio de Janeiro: GEN, 2022, p. 590-744. *E-book*.



auxílio de um ser humano (RODRIGES, 2020). Nos mais elevados níveis de autonomia⁶, estes veículos operam por meio de um sistema computacional com tecnologia de inteligência artificial.

Um veículo dessa espécie pode ser considerado um *agente racional baseado em objetivos*. *Agente*, porque é capaz de perceber seu ambiente por meio de *sensores* e sobre ele agir por intermédio de seus *atuadores* (NORVIG, RUSSEL, 2022)⁷. *Racional*, na medida em que para cada sequência de percepções e com base no seu conhecimento interno, adota um comportamento que visa maximizar seu desempenho (NORVIG, RUSSEL, 2022). *Baseado em objetivos*, porque não age aleatoriamente, mas adota comportamentos visando alcançar situações desejáveis para o ser humano (NORVIG, RUSSEL, 2022), como chegar ao destino informado por seu usuário e transportá-lo de forma segura.

De fato, os carros autônomos que operam nos níveis mais elevados de autonomia⁸ têm um sistema de direção composto por câmeras e sensores diversos (sensores de ultrassom, radares, sensores infravermelho e sistema de varredura a laser, conhecido pela sigla “LIDAR”)⁹, que coletam os dados do entorno do veículo para posterior transmissão à sua rede neural, conhecida pela sigla CNN (*Convolution Neural Network*). A rede neural, que corresponde à inteligência artificial, viabiliza que as percepções se transformem em ações e, assim, realiza a tomada de decisões referentes à condução, como o ajuste da velocidade e do ângulo de viragem do volante (CHEN *et al*, 2020).

Há um consenso científico de que os carros autônomos representam uma tecnologia disruptiva e que serão adotados em larga escala, tendo em vista os inúmeros benefícios que

⁶ Segundo a Sociedade Internacional dos Engenheiros Automotivos – SAE INTERNATIONAL (2021), os veículos são classificados em seis níveis de autonomia (0 a 5). O nível 0 corresponde à ausência de automação, de modo que todas as ações devem ser realizadas pelo motorista humano. O nível mais elevado, considerado de automação completa (nível 05), é aquele que dispensa qualquer intervenção humana (RODRIGES, 2020), podendo, inclusive, não possuir volante. Atualmente, os carros ditos inteligentes que estão em circulação possuem autonomia de nível 03 (automação condicional), em que o sistema operacional do veículo gerencia e realiza todas as tarefas relacionadas à condução, ao passo que o motorista humano só intervém quando requerido pelo sistema.

⁷ Para melhor compreensão acerca dos termos *sensores* e *atuadores*, menciona-se os exemplos de Stuart J. Russel e Peter Norvig (2022): agente humano e agente robótico. O primeiro tem como sensores os olhos, ouvidos e outros órgãos, e como atuadores as mãos, pernas, boca e outras partes do corpo. Um agente robótico, por sua vez, tem como sensores câmeras e detectores de distância por infravermelho e, na condição de atuadores, os seus motores.

⁸ Neste artigo, são considerados níveis mais elevados de autonomia os níveis 03, 04 e 05, em que o carro autônomo consegue realizar a maior parte ou todas as tarefas relacionadas à condução do veículo sem intervenção humana, que pode ser solicitada eventualmente nos níveis 03 e 04, mas que é completamente dispensada no nível 05.

⁹ A sigla “LIDAR” corresponde à expressão em inglês *Light Detection and Ranging*, que é um sistema de varredura que emite diversos pulsos de laser em 360 graus, identificando todos os objetos no entorno do veículo.



proporcionarão à sociedade¹⁰. Entende-se que o ponto de inflexão ocorrerá quando esses veículos chegarem a 10% de todos os automóveis em uso em determinado país, o que alguns acreditam que acontecerá ainda nessa década (SCHWAB, 2019¹¹).

Apesar das vantagens prometidas, os carros autônomos também poderão se envolver em acidentes de consumo, que terão origem em defeitos comuns aos veículos tradicionais ou em defeitos exclusivos de veículos autônomos, como aqueles oriundos de sensores imprecisos e/ou insuficientes; dificuldade de obter e processar rapidamente dados para a tomada de decisões da inteligência artificial; vulnerabilidade a ciberataques¹² etc. (MEIRELLES, SANTANA, 2022). Ainda quando os carros autônomos estejam livres destes defeitos e sem falhas em sua programação, a imprevisibilidade inerente à inteligência artificial, em decorrência do aprendizado de máquina, pode dar origem a ações perigosas e provocar danos aos consumidores, sobretudo se os objetivos do produto não estiverem alinhados adequadamente com os interesses dos seres humanos.

Com efeito, a inteligência artificial se concentra no estudo e no desenvolvimento de agentes racionais, isto é, que buscam fazer a “coisa certa”. A “coisa certa”, por sua vez, é definida por meio dos objetivos programados no agente (NORVIG, RUSSEL, 2022), que não são necessariamente especificados em sua totalidade. Acerca do tema, Norvig e Russel ponderam:

O modelo padrão tem constituído um guia útil para a pesquisa de IA desde o princípio, mas provavelmente não é o modelo certo a longo prazo. O motivo é que o modelo padrão pressupõe que daremos à máquina um objetivo totalmente especificado. Para uma tarefa definida artificialmente, como no xadrez ou no cálculo do caminho mais curto, a tarefa vem com um objetivo embutido – portanto, o modelo padrão é aplicável. Porém, ao avançarmos para o mundo real, torna-se cada vez mais difícil especificar o objetivo de forma completa e correta. Por exemplo, ao projetar um carro autônomo, pode-se pensar que o objetivo é chegar ao destino com segurança. Mas dirigir em qualquer estrada incorre em riscos de danos causados por outros motoristas errantes, falha de equipamento, e assim por diante; desse modo, uma meta estrita de segurança exigiria a permanência na garagem. Há um certo

¹⁰ A título exemplificativo, pode-se mencionar a redução considerável do congestionamento nas grandes cidades, a diminuição dos custos para locomoção, o maior controle do tráfego, o aumento do compartilhamento de veículos e atenuação dos indicadores de poluição e uma queda drástica do número de acidentes fatais no mundo, decorrentes predominantemente de falha humana (LITMAN, 2019).

¹¹ A previsão de Klaus Schwab tem como base a circulação de automóveis nos Estados Unidos da América.

¹² A título exemplificativo, menciona-se que, em 2015, dois hackers demonstraram que podiam invadir o sistema computacional de um carro autônomo em pleno movimento e controlar suas funções, utilizando-se do painel de entretenimento (SCHWAB, 2019). Acerca deste problema, Gustavo Tepedino e Rodrigo da Guia Silva (2019) ponderam sobre a existência de um aparente paradoxo, na medida em que, com a difusão dos carros autônomos, espera-se uma diminuição do número total de acidentes, ao mesmo tempo em que se teme por gerar lesões a pessoas que jamais seriam atingidas sem a presença de carros autônomos. Assim, por exemplo, mencionam os riscos relacionados à atuação dos *hackers*, que podem interferir no tráfego de carros autônomos. Ainda, apontam que essas circunstâncias evidenciam a periculosidade da atividade automotiva com o recurso à inteligência artificial.



compromisso entre prosseguir em direção ao destino e incorrer em risco de lesão. Como esse compromisso deve ser feito? Além disso, até que ponto podemos permitir que o carro execute ações que afetariam o comportamento de outros motoristas? Até que ponto o carro deve ser moderado em sua aceleração, direção e frenagem, a fim de evitar sacudir o passageiro? Essas perguntas são difíceis de serem respondidas *a priori*. São particularmente problemáticas na área geral da interação humano-robô, de que o carro autônomo é um exemplo (2022, p. 4).

Dessume-se, pois, que não constitui tarefa simples obter uma harmonia entre as verdadeiras preferências dos seres humanos e os objetivos das máquinas que operam por meio de inteligência artificial. Trata-se do “problema de alinhamento de valores”¹³, expressão utilizada por Norvig e Russel, que afirmam, ainda, que: “Consequências negativas surgirão de um sistema implantado com um objetivo incorreto. Além disso, quanto mais inteligente for o sistema, mais negativas serão as consequências” (2022, p. 4).

Os comportamentos adotados pela inteligência artificial dos carros autônomos, ainda que imprevisíveis, serão definidos em função dos objetivos especificados em sua programação, os quais, caso tenham sido programados em desarmonia com os interesses humanos, poderão provocar danos indesejados e que poderiam ser evitados. Neste caso, poderá o fornecedor do veículo (em especial o fabricante e o desenvolvedor do *software* inteligente¹⁴) ser responsabilizado civilmente?

Além da análise de diversos aspectos da responsabilidade pelo fato do produto, a resposta a esta questão demanda um exame sobre o risco do desenvolvimento, o qual este estudo se propõe a analisar.

3 RISCO DO DESENVOLVIMENTO: UMA ESPÉCIE DE DEFEITO DE CONCEPÇÃO

O sistema de responsabilidade que vigora no Código de Defesa do Consumidor é o de natureza objetiva¹⁵. E assim tende a continuar relativamente aos fornecedores de produtos que operam por meio de um sistema de inteligência artificial, conforme dispõe o art. 29, da

¹³ Para melhor compreensão do *problema de alinhamento de valores*, Peter Norvig e Stuart J. Russel (2022) citam a famosa mitologia grega do Rei Midas, rei lendário que desejou que tudo que tocasse se tornasse ouro. Teve seu desejo atendido, mas veio a se arrepender após tocar sua comida, bebida e membros da família.

¹⁴ Acerca do impacto das inteligências artificiais na definição da cadeia de fornecimento para fins de responsabilidade solidária entre os fornecedores, conferir MEIRELLES, Arthur; SANTANA, Agatha Gonçalves. A responsabilidade civil envolvendo inteligências artificiais em carros autônomos: repercussões no Código de Defesa do Consumidor. *Civilistica.com*. Rio de Janeiro, a. 11, n. 2, 2022.

¹⁵ Ressalva-se a responsabilidade dos profissionais liberais, que, nos termos do art. 14, § 4º, do CDC, será apurada mediante a verificação da culpa. Trata-se de uma exceção ao princípio da objetivação da responsabilidade civil por danos nas relações de consumo (DENARI, 2022).



minuta de substitutivo elaborada para instruir a apreciação dos Projetos de Lei de nº 5.051, de 2019; nº 21, de 2020; e nº 872, de 2021, constante do relatório final da comissão de juristas responsável pela elaboração do denominado Marco Legal da Inteligência Artificial¹⁶, apresentado ao Senado Federal em dezembro de 2022. Em síntese, o referido artigo estabelece que as hipóteses de responsabilização decorrentes de danos causados por sistemas de inteligência artificial permanecem sujeitas às regras do Código de Defesa do Consumidor.

Apesar de fundada no risco da atividade empresarial, o fato gerador da responsabilidade pelo fato do produto ou serviço é o defeito (CAVALIERI FILHO, 2022). Assim, o CDC adotou a teoria mitigada do risco ou teoria do risco da atividade defeituosa (AGUIAR JR., 2012), de modo que aquele que exerce uma atividade em que o produto pode ser defeituoso, havendo defeito, responde independentemente de culpa.

Na seara consumerista, pois, o defeito assume especial relevância dentre os pressupostos da responsabilidade pelo fato¹⁷, podendo ser considerado, inclusive, o seu elemento central (MEIRELLES, SANTANA, 2022). De fato, não há violação de dever jurídico, apto a gerar a responsabilização do agente, em razão do mero exercício de uma atividade que dá origem a produtos ou serviços perigosos. A responsabilidade somente resta configurada quando há violação do dever jurídico de segurança, que vem estabelecido no art. 12, § 1º, do CDC. O dispositivo legal criou uma cláusula geral de segurança¹⁸, que consiste no dever de o fornecedor não introduzir no mercado de consumo produto com defeito. Assim, criar um produto defeituoso significa violar o dever jurídico de zelar pela segurança dos consumidores.

O art. 12, § 1º, do CDC, apresenta o conceito legal de defeito, definindo-o como a frustração da legítima expectativa de segurança do consumidor, bem como elenca algumas circunstâncias relevantes para a sua análise, dentre as quais, a apresentação do produto (inciso I); o uso e os riscos que razoavelmente dele se esperam (inciso II); bem como a época em que foi colocado em circulação (inciso III). Este último critério é considerado o fundamento legal do risco do desenvolvimento nas relações de consumo.

O processo de modernização e o surgimento da sociedade de consumo se devem ao

¹⁶ Os projetos em discussão têm como objetivo estabelecer princípios, regras, diretrizes e fundamentos para regular o desenvolvimento e a aplicação da inteligência artificial no Brasil.

¹⁷ Segundo Ruy Rosado de Aguiar Jr. (2012), são pressupostos da responsabilidade pelo fato do produto: a) a colocação do produto ou serviço no mercado de consumo; b) a existência de defeito; c) o dano injusto; d) a relação de causalidade (nexo causal); e e) o fator de atribuição fundado na teoria mitigada do risco da atividade.

¹⁸ A Terceira Turma do Superior Tribunal de Justiça já reconheceu a existência da cláusula geral de segurança, em 04 de dezembro de 2018, por ocasião do julgamento do REsp n. 1.762.224/SP, de relatoria do Ministro Marco Aurélio Bellizze.



progresso científico e tecnológico. Todavia, os inúmeros benefícios gerados pelo avanço do conhecimento científico e da tecnologia são acompanhados, inevitavelmente, por malefícios, que são oriundos dos riscos da atividade empresarial. Assim, há uma forte ligação entre a sociedade de risco e a sociedade de consumo, marcadas pela “relação paradoxal entre progresso e risco” (KROETZ, L. SILVA, 2016, p. 83). É nesse contexto que se inserem os riscos do desenvolvimento.

Os riscos do desenvolvimento são aqueles desconhecidos pela comunidade científica no momento de introdução do produto ou serviço no mercado de consumo, em razão do estado da ciência e da técnica, e que somente são descobertos após um certo período de uso do bem de consumo e com o avanço do conhecimento científico. São responsáveis, em geral, por provocarem danos significativos aos consumidores, de que são exemplos os efeitos colaterais de medicamentos como o Cotergan-Talidomida, o anticolesterol MER-29 e o talco Morhange, que causaram, respectivamente, o nascimento de bebês com má formação congênita¹⁹, cegueira e intoxicação.

Embora haja na doutrina respeitável posicionamento no sentido de que o risco do desenvolvimento deve ser considerado uma excludente da responsabilidade do fornecedor²⁰, entende-se mais acertada a concepção de que deve ser qualificado juridicamente como um *defeito de concepção*. Assim se posiciona Benjamin, segundo o qual o risco do desenvolvimento corresponde a um defeito que “decorre da carência de informações científicas, à época da concepção, sobre os riscos inerentes à adoção de uma determinada tecnologia nova” (2016, p. 188-189). Em sentido semelhante se posicionam Cavalieri Filho²¹ (2022) e Denari (2022), sendo que este jurista, além de afirmar que o inciso III, do art. 12, § 1º, do CDC, não representa a adoção do risco do desenvolvimento como uma excludente da responsabilidade do fornecedor, ressalta que: “Quando estão em causa vidas humanas, as eximentes de responsabilidade devem ser recebidas pelo aplicador da norma com muita

¹⁹ O caso mais emblemático e consternador de risco de desenvolvimento envolvendo medicamentos é o da Talidomida, que era um anestésico frequentemente receitado a mulheres grávidas no final da década de 1950. Com o avanço do conhecimento científico, descobriu-se que o princípio ativo do medicamento era responsável pela má formação congênita dos bebês, que ficaram conhecidos ao redor do mundo como portadores da “Síndrome da Talidomida”. Para aprofundamento sobre a temática e casos paradigmáticos, recomenda-se a leitura de WESENDONCK, Tula. A responsabilidade civil pelos riscos do desenvolvimento: evolução histórica e disciplina no Direito Comparado. *Direito e Justiça*, v. 38, n. 2, p. 212-227, jul./dez. 2012.

²⁰ Para aprofundamento da temática, recomenda-se a leitura de WESENDONCK, Tula. A responsabilidade civil pelos riscos do desenvolvimento: evolução histórica e disciplina no Direito Comparado. In: *Direito e Justiça*, v. 38, n. 2, p. 212/227, jul./dez. 2012. Disponível em: <<https://revistaseletronicas.pucrs.br/ojs/index.php/fadir/article/view/12549>>. Acesso em: 12.06.2022.

²¹ O referido autor, além de qualificar o risco do desenvolvimento como um defeito de concepção, afirma que também representa um fortuito interno. Ainda, sustenta ser irrelevante a previsibilidade do defeito, uma vez que esta circunstância é irrelevante no sistema de responsabilidade do Código de Defesa do Consumidor, em que o fornecedor responde independentemente de culpa.



reserva e parcimônia” (DENARI, 2022, p. 212).

Este critério de interpretação das excludentes apresentado por Zelmo Denari aparenta, a princípio, apoiar-se excessivamente em um valor moral, relegando o aspecto jurídico. Contudo, o parâmetro oferecido é revestido de grande força jurídica. Afinal, a responsabilidade civil é instituto que está intimamente atrelado à tutela da dignidade humana e dos direitos fundamentais (KROETZ, L. SILVA, 2016)²², além de estar diretamente relacionado a um dos princípios basilares do direito, a saber, o princípio do *alterum non laedere* (a ninguém lesar). Ademais, a reparação integral dos danos constitui princípio fundamental, de modo que o direito da pessoa lesada somente deve ser afastado de forma excepcional. Assim, tendo em vista que o risco do desenvolvimento não está expressamente previsto como uma excludente da responsabilidade pelo fato, eventual interpretação que o qualifique como tal parece contrariar o sistema de responsabilidade do Código de Defesa do Consumidor.

Conforme ponderam Kroetz e Luiz Silva:

Vale frisar, mais uma vez, o fato de que as lesões decorrentes de acidentes de consumo, em especial as que envolvem os riscos do desenvolvimento, com frequência recaem sobre valores existenciais de grande importância, como a vida, a saúde e a integridade psicofísica dos consumidores. E tais valores, por força constitucional, devem ter sua tutela necessariamente prioritária. Assim, em eventual conflito entre a proteção dos direitos fundamentais dos consumidores e a proteção conferida à atividade econômica, esta deverá ceder espaço àquela” (2016, p. 12-13).

É sob essa perspectiva que os atos imprevisíveis da inteligência artificial deverão ser interpretados.

4 A IMPREVISIBILIDADE DOS COMPORTAMENTOS DA INTELIGÊNCIA ARTIFICIAL DOS CARROS AUTÔNOMOS E OS RISCOS DO DESENVOLVIMENTO

Além de apontarem a recorrente invocação da doutrina do risco do desenvolvimento nessa temática, Gustavo Tepedino e Rodrigo da Guia Silva (2020) alertam para a grande dificuldade que recairá na tarefa de definição de falha no funcionamento do código de programação, nos casos envolvendo atos imprevisíveis da inteligência artificial. Assim

²² Acerca da interpretação do risco do desenvolvimento em função da tutela dos direitos fundamentais, recomenda-se a leitura de KROETZ, Maria Cândida Pires Vieira do; SILVA, Luiz Augusto da. Um Prometeu pós-moderno? Sobre desenvolvimento, riscos e a responsabilidade civil nas relações de consumo. *Revista Brasileira de Direito Civil*, Belo Horizonte, v. 9, p. 81-101, jul./set. 2016.



ponderam os referidos juristas: “Afigura-se tênue, com efeito, a linha divisória entre o dano (que se espera não previsto, em homenagem à presunção de boa-fé subjetiva) produzido por sistema autônomo defeituoso e o dano produzido por sistema autônomo não defeituoso” (2020, p. RB-12.3).

De fato, tendo em vista que a inteligência artificial, para alcançar os melhores resultados, pode adotar comportamentos que não são prévia e pormenorizadamente descritos por seu programador, por meio de algoritmos, eventuais ações e omissões irão ocorrer sem que seja possível prevê-las. Acredita-se que, em geral, essas condutas criativas adotadas pela inteligência artificial de produtos ou serviços lançados no mercado de consumo serão seguras. Haverá problemas, todavia, nas situações em que os comportamentos inesperados derem origem a riscos e causarem danos.

Sabe-se que decisões autônomas de um sistema de inteligência artificial são esperadas e desejadas, bem assim que o nível de segurança esperado jamais será absoluto, de modo que é difícil afirmar, ao menos sem uma profunda incursão teórica, que uma decisão autônoma constitui um defeito, sobretudo porque a existência de riscos em tais decisões é reconhecida, aceita e tecnicamente impossível de afastar (CAMARA, ROBERTO, 2018)²³.

Assim, repisa-se, neste momento do estudo, alguns aspectos da teoria da responsabilidade pelo fato e da tecnologia de inteligência artificial que, analisados em conjunto, permitem qualificar como risco do desenvolvimento (defeito de concepção) os comportamentos imprevisíveis da inteligência artificial que geram danos.

Conforme exposto alhures, o campo de estudo da inteligência artificial visa desenvolver agentes racionais, que, na definição de Russel e Norvig (2022), são agentes que buscam fazer a “coisa certa”, selecionando, para cada sequência de percepções possível, uma ação que se espera maximizar sua medida de desempenho. A “coisa certa”, afirmam os referidos autores, é definida pelo objetivo implementado no sistema de inteligência artificial. Esse paradigma geral é chamado de *modelo padrão* (NORVIG, RUSSEL, 2022), que, todavia, não é o modelo certo a longo prazo, uma vez que as máquinas mais avançadas não terão objetivos totalmente especificados.

Neste ponto, retoma-se o problema do alinhamento de valores. Acerca dessa questão, Russel e Norvig obtemperam:

²³ Oportuno mencionar que a teoria da computação confirma que é impossível assegurar que qualquer programa de computador seja isento de defeitos. De fato, no processo de desenvolvimento de *softwares*, é demonstrável matematicamente a imprevisibilidade de todos os defeitos; além disso, alguns erros não são oriundos da fase de desenvolvimento do *software*, pois surgem posteriormente em razão das variáveis do ecossistema em que está inserido, o qual se altera ao longo do tempo (BORGES, MELLO, PINHEIRO, 2019).



É impossível antecipar todas as maneiras pelas quais uma máquina que busca um objetivo fixo pode se comportar mal. Logo, há uma boa razão para pensar que o modelo padrão é inadequado. Não queremos máquinas que sejam inteligentes no sentido de perseguir *seus* objetivos; queremos que elas busquem os *nossos* objetivos. Se não pudermos transferir esses objetivos perfeitamente para a máquina, então precisamos de uma nova formulação – uma em que a máquina esteja perseguindo nossos objetivos, mas necessariamente *não tenha certeza* de quais são eles. Quando uma máquina sabe que não conhece o objetivo completo, ela tem um incentivo para agir com cautela, pedir permissão, aprender mais sobre nossas preferências por meio da observação e submeter-se ao controle humano. Em última análise, queremos agentes *comprovadamente benéficos* para os humanos” (2022, p. 5, destaques dos autores).

Independentemente do modelo implementado na máquina, entende-se que os objetivos programados, total ou parcialmente, condicionarão as condutas do agente racional na busca pela realização da “coisa certa”. Assim, ainda que imprevisíveis, as ações ou omissões da máquina serão realizadas em função desses objetivos programados pelo desenvolvedor do *software* de inteligência artificial.²⁴

Retoma-se, aqui, para melhor visualização do problema, o carro autônomo. Supondo que a inteligência artificial desse veículo tenha sido programada para assegurar, acima de tudo, o transporte seguro do usuário, pode ocorrer que, diante de certas adversidades, o carro autônomo delibere por realizar manobras inusitadas que, embora resguardem a integridade física de seu usuário, coloquem em perigo e causem danos à incolumidade pessoal de pedestres. Segundo o objetivo programado, seria a coisa certa a se fazer.

Embora seja um exemplo simplório e com finalidade meramente didática, cenários semelhantes ao descrito poderão ocorrer quando da implementação em larga escala de carros autônomos. Afinal, conforme alertam Russel e Norvig: “devemos aceitar a possibilidade de que possamos colocar o propósito errado na máquina – exatamente o problema do Rei Midas” (2022, p. 36). Tem-se, aqui, o risco da atividade empresarial, fundamento da responsabilidade pelo fato do Código de Defesa do Consumidor.

Apesar de poderem ser considerados imprevisíveis e certamente não serem desejados pelos programadores, esses comportamentos têm origem na programação implementada na fase de concepção do produto²⁵. Desse modo, se essas ações se tornarem um padrão de

²⁴ Trata-se, aqui, de apenas uma ponderação, e não de uma afirmação categórica. O tema demanda o exame de inúmeros aspectos técnicos da inteligência artificial, relacionados à forma de tomada de decisão, à ética implementada no sistema de programação, às preferências dos usuários aprendidas pelo sistema de inteligência artificial com o regular uso do produto ou serviço, entre outros pontos que merecem a devida consideração em estudos mais abrangentes.

²⁵ Oportuno observar que, embora a fase de concepção do produto seja usualmente vista como uma etapa que antecede o lançamento do bem no mercado de consumo, é certo que, tratando-se de produtos com tecnologia de inteligência artificial, eventuais atualizações de *software* serão implementadas após a colocação do bem no mercado e, desse modo, se derem origem a defeitos, estes deverão ser considerados defeitos de concepção, uma



comportamento e, com o avanço do conhecimento científico, constatar-se que seria possível, em situações de adversidade, resguardar a integridade física de todas as pessoas (usuários dos carros autônomos e pedestres), desde que o transporte seguro do usuário não fosse programado na condição de objetivo prevalecente, restará configurado o risco do desenvolvimento.

Neste ponto, oportuna a afirmação de Cavalieri Filho (2022), segundo o qual o dever de responder por eventuais vícios ou defeitos dos bens e serviços fornecidos pertence ao fornecedor, ou seja, àquele que se dispõe a exercer alguma atividade no mercado de consumo. Não se adotou a teoria do risco do consumo, pois, como obtempera o referido jurista: “O consumidor não pode assumir os riscos das relações de consumo, não pode arcar sozinho com os prejuízos decorrentes dos acidentes de consumo, ou ficar sem indenização” (2022, p. 350). Assim, considerando que o consumidor não é o responsável pela definição e programação dos objetivos das máquinas dotadas de inteligência artificial, as consequências negativas dos atos imprevisíveis, realizados em função destes objetivos, não devem ser arcadas exclusivamente pela parte mais vulnerável da relação de consumo.

Sob essa perspectiva, entende-se possível imputar a responsabilidade pelos atos imprevisíveis aos fornecedores de produtos que operam por meio de um sistema de inteligência artificial, uma vez que o risco do desenvolvimento constitui, na essência, um defeito de concepção (fato gerador da responsabilidade pelo fato). Portanto, se presentes os demais pressupostos da responsabilidade, restará configurada a obrigação de indenizar do fornecedor.

5 CONSIDERAÇÕES FINAIS

A preocupação com as repercussões negativas da inteligência artificial é, de fato, legítima, e o problema do alinhamento de valores das máquinas dotadas de inteligência artificial com os interesses humanos uma questão relevante em sede de responsabilidade civil. Afinal, um sistema computacional autônomo programado com um objetivo incorreto é fonte de danos em potencial.

Sem pretensão de atingir a completude, em razão da complexidade da matéria, o presente artigo visou oferecer uma breve base teórica para auxiliar os juristas na tarefa de qualificação legal dos atos imprevisíveis da inteligência artificial, optando por tratar do risco

vez que relacionados ao momento de idealização do produto.



do desenvolvimento, em razão de sua recorrente invocação como matéria de defesa, visando afastar o dever de reparar os danos imprevisíveis.

Evidenciou-se, com o desenvolvimento do trabalho, a relevância de um estudo profundo, técnico e interdisciplinar sobre a temática, a fim de evitar classificações equivocadas dos fatos jurídicos e provocar injustiças em matéria de reparação de danos, que está diretamente relacionada à tutela da dignidade da pessoa humana e de direitos fundamentais.

REFERÊNCIAS

AGUIAR JR., Ruy Rosado de. Os pressupostos da responsabilidade civil no CDC e as ações de indenização por danos associados ao consumo de cigarros. **Civilística**, a. 1, n. 1, 2012.

BENJAMIN, Antônio Herman V.; BESSA, Leonardo Roscoe Bessa; MARQUES, Cláudia Lima. **Manual de Direito do Consumidor**. 7. ed. São Paulo: Revista dos Tribunais, 2016.

BORGES, Maria Ruth; MELLO, Flávio Luís de; PINHEIRO, Guilherme Pereira. Danos envolvendo veículos autônomos e a responsabilidade civil do fornecedor. **Revista Brasileiro de Direito Civil**. Belo Horizonte, v. 21, p. 247-267, jul./set. 2019. Disponível em: <<https://rbdcivil.ibdcivil.org.br/rbdc/article/view/472>>. Acesso em: 19 jun. 2022.

CAMARA, Dennys; ROBERTO, Enrico. **Danos causados por carros autônomos**. E se fosse no Brasil?. *Jota*, 6 abr. 2018. Disponível em: <<https://www.jota.info/opiniao-e-analise/artigos/danos-causados-por-carros-autonomos-06042018>>. Acesso em: 25 jan. 2023.

CAVALIERI FILHO, Sergio. **Programa de direito do consumidor**. 6. ed. Barueri: Atlas, 2022. *E-book*.

CHEN, Chen et al. An analysis of adversarial attacks and defenses on autonomous driving models. **International Conference on Pervasive Computing and Communications (PerCom)**. IEEE, 2020. p. 1-10. Disponível em: <<https://arxiv.org/pdf/2002.02175.pdf>>. Acesso em: 26 nov. 2022.

DENARI, Zelmo. Da qualidade de produtos e serviços, da prevenção e da reparação dos danos. In: BENJAMIN, Antônio Herman V.; DENARI, Zelmo; FILOMENO, José Geraldo Brito; FINK, Daniel Roberto; GRINOVER, Ada Pelegrini; JÚNIOR, Nelson Nery; MARQUES, Cláudia Lima; PFEIFFER, Roberto; WATANABE, Kazuo. **Código Brasileiro de Defesa do Consumidor**. 13. ed. Rio de Janeiro: Forense, 2022, p. 195-258. *E-book*.

KROETZ, Maria Cândida Pires Vieira do; SILVA, Luiz Augusto da. Um Prometeu pós-moderno? Sobre desenvolvimento, riscos e a responsabilidade civil nas relações de consumo. **Revista Brasileira de Direito Civil**, Belo Horizonte, v. 9, p. 81-101, jul./set. 2016. Disponível em: <<https://rbdcivil.emnuvens.com.br/rbdc/article/view/55>>. Acesso em: 22 jan. 2023.



LITMAN, Todd. **Autonomous vehicle implementation predictions**: implications for transport planning. Victoria Transport Policy Institute, 2019. Disponível em: <<https://bit.ly/2tiSqIH>>. Acesso em: 05 dez. 2022.

MEIRELLES, Arthur; SANTANA, Agatha Gonçalves. A responsabilidade civil envolvendo inteligências artificiais em carros autônomos: repercussões no Código de Defesa do Consumidor. **Civilistica.com**. Rio de Janeiro, a. 11, n. 2, 2022. Disponível em: <<http://civilistica.com/a-responsabilidade-civil-envolvendo/>>. Acesso em: 14 jan. 2022.

NORVIG, Peter; RUSSEL, Stuart J. Agentes inteligentes. *In: Inteligência artificial: uma abordagem moderna*. 4. ed. Rio de Janeiro: GEN, 2022, p. 33-56. *E-book*.

_____. Filosofia, ética e segurança da IA. *In: Inteligência artificial: uma abordagem moderna*. 4. ed. Rio de Janeiro: GEN, 2022, p. 890-917. *E-book*.

_____. O que é IA? *In: Inteligência artificial: uma abordagem moderna*. 4. ed. Rio de Janeiro: GEN, 2022, p. 01-05. *E-book*.

PIRES, Thatiane Cristina Fontão; SILVA, Rafael Peteffi da. A responsabilidade civil pelos atos autônomos da inteligência artificial: notas iniciais sobre a Resolução do Parlamento Europeu. **Revista Brasileira de Políticas Públicas**, v. 7, n. 3, p. 239-254, dez. 2017. Disponível em: <<https://www.arqcom.uniceub.br/RBPP/article/viewFile/4951/3643>>. Acesso em: 15 out. 2021.

RODRIGUES, JC. **O carro que dirige decide por mim**: estamos preparados para a era dos carros autônomos? 2. ed. São Paulo: Ed. do Autor, 2020.

SCHWAB, Klaus. **A quarta revolução industrial**. São Paulo: Edipro, 2019. *E-book*.

SENADO FEDERAL (Brasil). **Relatório final**: comissão de juristas responsável por subsidiar elaboração de substitutivo sobre inteligência artificial no Brasil. Disponível em: <<https://legis.senado.leg.br/comissoes/mnas?codcol=2504&tp=4>>. Acesso em: 01 fev. 2023.

SILVA, Rodrigo da Guia; TEPEDINO, Gustavo. Desafios da inteligência artificial em matéria de responsabilidade civil. **Revista Brasileira de Direito Civil**, Belo Horizonte, v. 21, p. 61-86, jul./set. 2019.

SILVA, Rodrigo da Guia; TEPEDINO, Gustavo. Inteligência Artificial e elementos da responsabilidade civil. *In: FRAZÃO, Ana; MULHOLLAND, Catlin. Inteligência Artificial e Direito: ética, regulação e responsabilidade*. São Paulo: Thomson Reuters Brasil, 2020, p. RB-1.1-RB-1.8. *E-book*.

SOCIETY OF AUTOMOTIVE ENGINEERS. **Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles**. 2021. Disponível em: <https://www.sae.org/standards/content/j3016_202104/>. Acesso em: 16 jan. 2023.

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

WESENDONCK, Tula. A responsabilidade civil pelos riscos do desenvolvimento: evolução histórica e disciplina no Direito Comparado. *In: Direito e Justiça*, v. 38, n. 2, p. 212/227,



jul./dez.

2012.

Disponível

em:

<<https://revistaseletronicas.pucrs.br/ojs/index.php/fadir/article/view/12549>>.

Acesso

em:

12.06.2022.



RISK OF DEVELOPMENT: THE CHALLENGE OF VALUE ALIGNMENT AND THE UNFORESEEABLE ACTS OF THE ARTIFICIAL INTELLIGENCE

ABSTRACT

The present article aims at exploring the legal classification possibilities of the unforeseeable acts of the artificial intelligence, in the light of the Consumer Protection Code. The used methodology is based on qualitative, bibliographic and documental research, and makes use of inductive, deductive and hypothetical-deductive methods. The study will initially present general aspects of the artificial intelligence, showing the new risks originated from this technology and emphasizing the unforeseeable acts and the difficulty to harmonize its goals with human interests. For better understanding of the problematic, self-driving cars will be used as examples of this technology. In the second part of the article, the reader will be introduced to the development risks concept, as well to legal aspects that allow us to qualify it as conception defects, in the light of the Consumer Protection Code. At the end, the article makes it possible to qualify the unforeseeable acts of this technology as a development risk and one which causes damage. The article concludes thus, that the unforeseeable acts of the artificial intelligence are determined by the goals programmed in the machine, which is covered by the theory of risk, so, if they cause damages, they should be classified as conception defects.

Keywords: Artificial intelligence. Self-driving cars. Unforeseeable acts. Risk of developments. Civil liability.

