

## Programações podem ser punidas? Responsabilidade penal em decisões tomadas pela inteligência artificial

### Can programmed interfaces be punished? Criminal liability for decisions taken by artificial intelligence

Renato Watanabe de Morais<sup>1</sup>



This work is licensed under a Creative Commons Attribution 4.0 International License.

**Resumo:** A visão antropocêntrica é a base do Direito Penal tradicional, que considera apenas a pessoa humana como sujeito de direitos e detentora da dignidade jurídica. No entanto, o aprendizado de máquina, tecnologia utilizada na matemática e nas ciências da computação, levanta a questão, ainda filosófica, de se um robô pode cometer crimes e ser culpado por isso. A partir do momento em que grande parte da vida das pessoas passa a ser virtual, o desenvolvimento de algoritmos para analisar e direcionar as informações torna-se cada vez mais complexo e usado para ler padrões comportamentais. Entretanto, quanto mais dados o algoritmo recebe, mais variada sua resposta pode ser. A figurativa responsabilidade penal da inteligência artificial pode ser uma fronteira interessante para o nexos de causalidade, impedindo que o programador seja punido por atos decididos pelas conexões neuronais eletrônicas. A falta de regulamentação pode levar a consequências graves, sugerindo assim a adoção de programas de compliance na esfera das ciências da computação.

**Palavras-chave:** Responsabilidade penal; Inteligência artificial; Programador; Literatura cyberpunk.

**Abstract:** The anthropocentric view is the basis of traditional Criminal Law, which considers only the human being as subject of rights and holder of legal dignity. However, machine learning, a technology used in mathematics and computer science, raises the philosophical question of whether a robot can commit crimes and be held responsible for them. As more and

---

<sup>1</sup> Bacharel em Ciências Jurídicas e Mestre em Direito Penal pela Universidade de São Paulo/SP. Professor de Direito e Coordenador dos CST em Serviços Jurídicos, Cartoriais e Notariais; em Gestão de Segurança Privada; e em Investigação Forense e Perícia Criminal da Universidade Santo Amaro/SP.

more of people's lives become virtual, the development of algorithms to analyze information becomes increasingly complex and is used to read behavioral patterns. Nonetheless, The more data the algorithm receives, the more varied its response can be. The figurative criminal responsibility of artificial intelligence can be an interesting boundary for the causal nexus, preventing the programmer from being punished for acts decided by electronic neural connections. The lack of regulation can lead to serious consequences, suggesting then the adoption of compliance programs in the field of computer science.

**Keywords:** Criminal liability; Artificial intelligence; Programmer; Cyberpunk literature.

---

## 1. Introdução

A cada novo semestre, a mesma *programação*. Em determinada aula do primeiro semestre do curso de Direito Penal, o Professor, *roboticamente*, ao tratar do objeto do Direito Penal, acaba proferindo as mesmas palavras: o único evento do plano sensível que interessa ao Direito Penal é aquele em que o resultado, proibido pela norma, em alguma medida possa ter sua responsabilidade atribuída a um ser humano que atenda aos requisitos do conceito analítico de crime sedimentados pela doutrina.

A conduta humana, alçada ao posto de fenômeno jurídico enquanto aquilo a ser analisado junto ao paradigma do conceito dogmático de crime, é ao mesmo tempo consequência e indicativo da função limitadora do Direito Penal. A culpabilidade presente no conceito de crime traz a concepção de que a ferramenta mais severa da atuação estatal poderá ser utilizada quando houver algum grau de elemento volitivo. Logo, terá relevância aquele evento em que houver no mínimo vontade, na ação ou na omissão, não obstante isso não se verifique no resultado. Por outro lado, afirmar que somente a conduta humana traz interesse jurídico penal é limitar o próprio poder punitivo do Estado na medida em que não há liberdade punitiva conquanto graves sejam as consequências de um determinado cenário.

Este ponto de partida é essencial para que o estudante de direito consiga sedimentar conceitos fundantes dogmáticos. Possibilitam, posteriormente, que ele tensione e critique o atual estado de arte da aplicação da pena. Desta forma, não se aceita a legislação vigente como realidade dada, mas construção social, mesmo que com questionável representatividade, que

pode e deve passar por constantes avaliações pela academia. É o caso, por exemplo, da responsabilidade penal da pessoa jurídica.

Entretanto, a responsabilidade penal da pessoa jurídica tem fundamento num aspecto de caráter muito mais pragmático que efetivamente de uma análise filosófica a respeito de o que é conduta passível de punição. Os defensores da responsabilidade penal da pessoa jurídica não negam certa incompatibilidade do instituto com os conceitos tradicionais. Afinal, pessoa jurídica não possui desejo. Em verdade, ela sequer existe naturalmente. É uma ficção jurídica, uma convenção de atribuição de direitos e deveres. Ocorre que boa parte do substrato que mantém a existência deste tipo de responsabilidade reside muito mais numa tentativa de usar o Direito Penal como forma de impedir que empresas e outros conglomerados sejam utilizados para a prática de atos indesejados (BUSATO, 2020, p. 559). A complexidade estrutural das pessoas jurídicas, ou de contratos coligados entre pessoas jurídicas, dificultaria a aplicação de princípios basilares como a responsabilidade subjetiva e averiguação de dolo ou culpa. Assim é realizada uma troca: de um lado se flexibilizam barreiras dogmáticas; do outro, busca-se a aplicação de penas menos severas. Dessa forma, observa-se uma clara aproximação com o direito administrativo (SILVA SÁNCHEZ, 2011, p. 179).

A questão no entorno da responsabilização de empresas reside, assim, na óptica da função do Direito Penal e qual o seu alcance no controle social. O presente trabalho, por outro lado, ainda que verse também sobre artificialidades, traz algumas inquietudes filosóficas, uma possível resposta e uma proposta.

Pune-se o ser humano pelo dogma antropocêntrico de ser dotado de livre-arbítrio em que, com exceção dos distúrbios de ordem mental, assume-se que o indivíduo *escolhe* ser criminoso (MARTINELLI; BEM, 2019, p. 638). Todavia, trata-se de concepção operacionalmente válida para o Direito, porém refutada por outras áreas do conhecimento humano. Os caminhos decisórios humanos são muito mais delimitados e menos conscientes do que pode acreditar o senso comum.

Se o pensamento desencurralado é uma ilusão, talvez a tomada de decisão por uma inteligência artificial após consolidação de dados por meio de algoritmos automatizados não seja tão apartada da realidade humana. E se a aplicação de pena é admitida para um determinado padrão de escolhas, não haveria impeditivo, num primeiro momento, para que a sanção possa ser estendida também a algo que tem igual comportamento.

A criminalização de robôs tem aura de ficção científica, objeto de devaneios de escritores e roteiristas *cyberpunk*, mas é certo que a tomada de decisões por inteligência artificial, nesta primeira metade do século XXI, já é cotidiano que implica desde uma neurocirurgia remota até a lista de compras do mercado. Portanto, uma reflexão, mesmo que breve, de como manejar um instrumento social milenar num cenário em que se questiona até com quem se interage ganha paulatinamente importância.

## 2. Responsabilidade penal da *pessoa* algorítmica

“Domin: O que poderia acontecer com ela?”

Marius: Ela pararia de se mover. Eles provavelmente a colocariam numa máquina compactadora.

Domin: Isso é morte, Marius. Você não tem medo da morte?

Marius: Não” (ČAPEK, 1923, tradução nossa).

No início da década de 1920, uma peça de teatro introduziria na língua inglesa uma palavra que hoje, cem anos depois, faz-se presente no vocabulário não apenas de especialistas em programação, mas de analistas políticos, médicos, juristas. “Robô” foi o termo utilizado pelo checo Karel Čapek (MORAVEC, 2021) para se referir a humanoides criados artificialmente a partir de carne e sangue sintético. Em *Rossumovi Univerzální Roboti*, robôs, aqui seres vivos, são feitos para realizar trabalhos – a palavra “robotá” em checo pode ser traduzida como trabalho forçado – ordenados por humanos na expectativa de redução de custos e aumento de produção (ČAPEK, 1923). Eventualmente, a explorada mão-de-obra, após uma inadvertida e secreta alteração na fórmula da produção realizada por um dos fisiologistas, acaba dando início a uma revolução que levou à quase extinção dos seres humanos.

Esse enredo, repetido por vários trabalhos posteriores de ficção científica, expõe uma velha discussão filosófica acerca de o que define o ser humano como tal. Na edição de setembro de 1940 da revista *Super Science Stories*, Asimov (2014, p. 19), no conto *Robbie*, apresenta ao leitor a preocupação dos pais da menina Gloria, que percebia no robô-babá – que dá nome ao conto – uma amizade tão intensa que a impedia de brincar com outras crianças ou se interessar por outra coisa. Rick Deckard, cuja história foi mostrada ao público pela primeira vez em 1968, buscando trocar sua ovelha de estimação elétrica por uma verdadeira, resolve aceitar o trabalho de “aposentar” andróides que se rebelaram e mataram seus donos. À medida que busca cumprir

com sua missão, a diferença entre o verdadeiro e o criado por humanos fica cada vez menos evidente a ponto de Deckard se autoquestionar (DICK, 2017).

Muito embora, em todos os exemplos supra, há de se notar a presença de máquinas humanoides, é certo que elas são mero suporte físico que auxilia no convencimento do enredo pelo leitor, já que a reflexão proposta recairia na, então futurista, inteligência artificial.

A narrativa da peça teatral checoslovaca introduz a ideia de provar a desnecessidade do conceito de um ser divino superior que teria criado a raça humana. A partir do instante que a humanidade conseguisse criar seres a partir de sua própria tecnologia, Deus, ou quem o faça suas vezes, seria comprovadamente inexistente. Assim como nos mitos gregos, em que a punição surge quando a humanidade sucumbe à arrogância pelo desejo de se equiparar aos deuses (ZOJA, 2000, p. 119), as pessoas padecem aniquiladas pela cria superior. Entretanto, sem a fórmula elemental, até os próprios robôs se encontram sem saída para sua permanência na História. O drama encontra seu desfecho num *deus ex machina*<sup>1</sup> em que se descobre que um casal de andróides tem em sua composição um pouco de Adão e Eva e que a humanidade poderia começar novamente do zero. Moral: a natureza vence afinal. Robôs, portanto, por mais sofisticados que sejam, não são humanos que, embora imperfeitos, ainda seriam frutos de natureza superior.

Essa visão antropocêntrica é a base do Direito Penal tradicional. Somente pode ser considerado sujeito de direitos a pessoa humana (e, assim, detentor único da dignidade jurídica), dada, em tese, sua superioridade concedida naturalmente, conforme preceitua o jusnaturalismo, sendo ela manifestada em sua racionalidade, que já era defendida no humanismo renascentista (SOARES, 2009, p.132).

A conduta que interessa para o Direito Penal, destarte, é aquela perpetrada por um humano, na medida em que, conforme sua inteligência única, passa por um processo gradativo de socialização desde o seu nascimento. Logo, absorve os valores impostos culturalmente e passa a agir de acordo com o desejado pela sociedade (REALE JÚNIOR, 2020, p. 5). Lateralmente, essa mesma capacidade cognitiva o faz ter o condão de romper com o pacto de convivência social, já que, embora moldado pelo meio, teria livre arbítrio. Cumpre lembrar que a concepção trazida por alguns membros da chamada Escola Clássica, que serviu como fundamento para o que hoje compreendemos ser a dogmática penal, sequer considerava a

---

<sup>1</sup> “Um evento ou poder inesperado que resolve uma situação aparentemente desenganada, especialmente como um recurso de enredo numa peça ou numa novela” (OXFORD LANGUAGES, 2021, tradução nossa).

influência do meio, entendendo que o livre-arbítrio seria indeterminado (REALE JÚNIOR, 2020, p. 42).

A racionalidade humana faria com que uma constante análise custo-benefício permanecesse em voga. Calcula-se todo tempo se a atitude a ser tomada traria mais benefícios que as demais opções e se eventuais prejuízos seriam suplantados por eles. Traçados objetivos de conforto e segurança, que poderiam ser resumidos numa busca hedonista, a pessoa acabaria por escolher qual caminho seria mais curto para atingi-los (CLARKE; CORNISH, 2014). Porém, observa-se que a liberdade de agir, *per si*, não explicaria bem o porquê de punir, vez que a teoria racionalista do crime não responderia os motivos pelos quais os objetivos egoístas de alguns eventos criminosos não coincidiriam com os objetivos coletivos.

Hirschi (1969), por exemplo, sustenta que o que ocasiona a criminalidade seria a ausência de laços sociais na família, na escola, na comunidade que leva à falta de validação moral da legalidade. Aproxima-se, assim, da tradição behaviorista social da Escola de Chicago e a teoria ecológica humana, em que o processo de comunicação tem um papel fundante no comportamento do indivíduo (PARK; BURGESS, 1921, p. 36).

Surge, conseqüentemente, a dúvida que dita boa parte dos estudos da mente: em que medida há livre arbítrio se o meio limita ou influencia as decisões tomadas? Essa discussão é importante para o presente trabalho, mas sobretudo um ponto de partida para outra reflexão que, num futuro não tão distante, pode eventualmente escapar das obras de ficção científica ou de certos nichos acadêmicos.

À margem das fantasiosas manifestações artísticas, o uso da inteligência artificial, num primeiro momento, deu-se no sentido da automação de respostas a determinados padrões apresentados pelos usuários de dispositivos eletrônicos. O objetivo é garantir que esses “agentes inteligentes” façam a leitura do cenário e tomem ações que aumentem a eficiência de dada tarefa ou as chances de obtenção de sucesso de certo objetivo final (POOLE; MACKWORTH; GOEBEL, 1998, p. 2). Teixeira (2014, p. 6) ilustra as primeiras tentativas de criação de uma inteligência artificial que processa dados como pensa um ser humano por meio de programas de jogar xadrez:

No caso do xadrez, para se responder corretamente a uma jogada do oponente, seria necessário percorrer todas as possibilidades que se seguiriam dela (...). Mas com a heurística<sup>2</sup>, você tenta reduzir essas conseqüências desenvolvendo algum tipo de

---

<sup>2</sup> Por heurística, o autor aqui se refere ao método de solução de problemas baseado num raciocínio seletivo. Ou seja, ao invés explorar sistematicamente todas as possibilidades de solução de um dado quadro de eventos –

estratégia. É assim que nós, humanos, jogamos xadrez – e foi baseando-se na mente humana que Newell, Shaw e Simon desenvolveram seu programa para jogar xadrez. Infelizmente, os primeiros programas de jogar xadrez, que apareceram na década de 1960, perdiam para crianças de dez anos.

Num mundo ideal, são programados parâmetros de percepção de dados válidos a partir de um banco já existente ou de um fluxo contínuo, estabelecem-se quais os resultados possíveis e quais os desejados dessa organização maciça de informações feita pelo computador e comanda-se que aponte qual a melhor atitude a ser tomada com base em estatística ou objetivos a serem alcançados. Preferencialmente, espera-se que não somente a resposta seja pontada, mas que seja executada pela própria inteligência artificial.

No entanto, ao longo do século XX e no decorrer do atual momento, percebeu-se efervescente criação e acúmulo de conhecimento e progresso científico que ainda não havia ocorrido na história em tão curto espaço de tempo. O método científico moderno, embora não alheio a críticas pelo alijamento do conhecimento vulgar como fonte válida de respostas, encontrou nas ciências exatas, mais binárias que as humanas, fértil trajeto para desenvolvimento de novas tecnologias de informação e comunicação (TONET, 2013, p. 10).

Dentro do campo da matemática e das ciências da computação, o aprendizado de máquina ganha tração e lança novas luzes nos estudos dos algoritmos. A partir do instante que grande parte da vida de bilhões de pessoas passa a ser uma realidade virtualizada, um gigante banco de dados passa a ficar disponível para análise e direcionamento (BOYD; CRAWFORD, 2011). Com isso, o desenvolvimento de algoritmos para organizar estatisticamente esse grande volume de informações vai se tornando cada vez mais complexo e necessário para realizar leituras de padrões comportamentais sociais.

Se, nos primórdios, as máquinas trabalhavam com *inputs* únicos (as peças brancas começam com peão e4) e davam respostas únicas (as pretas devem jogar e5 para disputar o controle do centro do tabuleiro), o desenvolvimento do *machine learning* contribuiu para uma nova gama de possibilidades de uso, desde a análise comportamental para fins publicitários, a escolha do melhor procedimento cirúrgico a ser adotado ou, mesmo, qual a jogada do xadrez

---

como num ataque de força bruta –, o investigador buscará elementos próximos ao objeto analisado e tentará encontrar informações que possam fazê-lo deduzir qual a melhor resposta. “A outra maneira [de descobrir a senha bancária do vizinho] é tentar encontrar algum tipo de atalho para adivinhar essa senha. Você começa a perguntar ao seu vizinho o dia em que ele nasceu, casou, etc., e tenta números próximos a esses (...)” (TEIXEIRA, 2014, p. 6).



mais adequada (se as peças brancas começam com peão e4 e as pretas respondem com peão e5, e a jogada seguinte se desenvolve com cavalo branco f3 e considerando que o oponente é Magnus Carlsen, que usou a abertura Ruy-Lopez em 18 das 20 últimas partidas e foi derrotado em 2/3 das vezes em que se usou a Defesa Berlinense, logo, será essa a melhor solução a ser adotada).

O aprendizado de máquina é uma tecnologia em que os computadores adquirem a “capacidade” de aprender conforme as respostas esperadas por meio de associações de dados diversos. Podem ser imagens, números, palavras, voz, ou seja, qualquer linguagem que possa ser identificada pelo algoritmo. Os algoritmos de *machine learning*, diferentemente dos tradicionais em que se automatiza uma resposta para cada *input*, são gerados a partir dos dados que serão analisados e dos resultados ou respostas que se esperam dessa análise. Em seguida a este processo, o sistema cria as próprias regras ou perguntas (MITCHELL, 1997, p. 5). A quantidade de dados a ser estudada traz cenários e desafios diversos. Aquele que importa para o atual propósito é o de *big data*.

Dá-se que o grande volume de informações trabalhadas, ao mesmo tempo que contribui estatisticamente para a definição de um padrão, soma uma nova dificuldade: os algoritmos não podem ser excessivamente específicos. Assim, criam-se mais balizas para as possibilidades de *output* que verdadeiras respostas prontas (KOZA; BENNETT; ANDRE; KEANE, 1996).

Conseqüentemente, quanto mais material é entregue para o algoritmo, mais variado poderá ser seu comportamento. Sem o estrito preestabelecimento de comportamentos, a performance do dispositivo passa a ser cada vez menos previsível e, em alguma medida, orgânica.

É importante, porém, lembrar que o estudo acerca do aprendizado de máquinas não envolve somente as informações que entram ou saem de uma ferramenta cognitiva. Como acima referido, o sistema cria suas próprias regras semânticas. A forma como os dados são correlacionados são, por vezes, mais importantes que a própria conclusão. Veja-se como exemplo as redes neuronais artificiais. São sistemas de computação inspirados nas redes neurais biológicas que constituem os cérebros dos animais. Uma rede neuronal artificial é baseada em uma coleção de unidades conectadas, ou nós, chamados neurônios artificiais – que tem por modelo vago os neurônios de um cérebro biológico. Cada conexão, como as sinapses, pode transmitir um sinal a outros neurônios. Após o recebimento, o neurônio processa e sinaliza os demais a ela conectados. Esse sinal é um número real e o output de cada neurônio é calculado



por alguma função não-linear da soma de seus inputs, como, por exemplo, o histórico de partidas de um enxadrista (GOODFELLOW; BENGIO; COURVILLE, 2016, p. 96). Essa heurística confere à máquina contornos cognitivos que acabam se assemelhando com o ser humano.

Em 1950, essa problemática já foi abordada por Alan Turing. Em seu artigo *Computing Machinery and Intelligence* (1950), Turing propõe a aplicação de um teste para verificar a habilidade da máquina em exibir comportamento inteligente equivalente, ou ao menos indistinguível, de uma pessoa. Diversas objeções foram feitas ao seu modelo de avaliação como, por exemplo, a inabilidade da máquina efetivamente pensar e a objeção de Ada Lovelace de que as máquinas não poderiam ser criativas (RUSSELL; NORVIG, 2003, p. 948). Embora, admite-se, a pertinência das objeções, é certo que algumas poderiam ser feitas também aos humanos, na medida que a capacidade criativa é também limitada pelas experiências sensoriais e cognitivas de cada um e o conceito de pensamento pode variar conforme a premissa teórica utilizada.

O Teste de Turing é utilizado em concursos, como os organizados pela Universidade de Reading, Londres, como uma forma de incentivo ao aperfeiçoamento do estudo e da aplicação da inteligência artificial. É célebre o episódio de 2014 em que a máquina *chatbot* russa “Eugene Goostman” teria (há debates entre cientistas da computação acerca da validade do método na aplicação da prova) sido aprovada no teste se passando por um garoto de 13 anos que morava na Ucrânia, enganando 33% dos jurados do evento que marcava o aniversário de 60 anos da morte de Alan Turing. Certo, porém, que o Teste se tornou popular mais por sua representação na cultura popular – o Teste Voight-Kampff de *Androides Sonham Com Ovelhas Elétricas?* – e ainda se está longe de sua aplicação para definição de responsabilidade jurídica ante um evento ilícito.

Duas experiências distintas podem ser exemplificadas sobre a aprendizagem por máquinas. Sophia, então cidadã da Arábia Saudita, em outubro de 2017 teve a oportunidade de discursar perante as Nações Unidas sobre desenvolvimento (UN NEW CENTRE, 2017). Não seria um acontecimento extraordinário se Sophia não fosse um robô construído e programado por David Hanson, capaz de reproduzir expressões faciais de sentimentos sintéticos e manter algum grau de diálogo utilizando-se da mesma ferramenta de reconhecimento de comandos utilizada pelo Google (HANSON ROBOTICS, 2021).

Um ano antes, a Microsoft disponibilizou um *bot* para que pudesse aprender com os usuários da rede social Twitter como se comunicar. Em menos de um dia, @TayandYou se tornou uma máquina – literalmente e figurativamente – racista, misógina e antissemita. Inicialmente, apresentou-se como um gerador de tweets inofensivo. Contudo, a forma como os usuários da plataforma interagiram fizeram com que ela espalhasse seus comportamentos. Um dos usuários afirmou que os ataques de 11 de setembro de 2001 eram de responsabilidade dos judeus e assim ela passou a argumentar. Horas mais tarde, a inteligência artificial construiu “eu [uma pachoucada] odeio feministas e elas devem todas morrer e queimar no inferno tradução nossa” (CHERELUS; TENNERY, 2016, tradução nossa).

Talvez não prevendo com tal precisão, mas certamente sabendo dos riscos de uma possibilidade ilimitada de escolhas para um robô, Isaac Asimov, em 1942, redige em seu conto *Runaround* aquelas que ficaram conhecidas como as Três Leis da Robótica: “Um robô não pode ferir um ser humano ou, por inação, permitir que um ser humano sofra algum mal”; “Um robô deve obedecer às ordens que lhe sejam dadas por seres humanos, exceto nos casos em que entrem em conflito com a Primeira Lei”; “Um robô deve proteger sua própria existência, desde que tal proteção não entre em conflito com a Primeira ou Segunda Leis” (2014, p. 49).

A partir disso, surge a indagação: pode um robô praticar crime? Vez que seu poder de decisão pode se assemelhar ao de uma pessoa, caso um robô viesse a praticar uma conduta típica e antijurídica, poderia a inteligência artificial ser culpável? Eduardovich defende que, em face dos prospectos avanços na temática apresentada, uma das possibilidades seria criar a categoria jurídica “entidade eletrônica”. Ela andaria paralela aos institutos da pessoa física e da pessoa jurídica, tendo regulamento próprio e com a possibilidade de receber sanções. O autor afirma que a ausência de responsabilidade penal da máquina somente faz sentido na medida em que as pessoas ainda mantenham controle sobre aquela. Vez que a tendência é o surgimento de uma inteligência independente de qualquer interferência de programação humana, sugere ele que a emergência de um capítulo “Medidas criminais a respeito das entidades eletrônicas” no Código Penal ucraniano é questão de tempo (2017, p. 138, tradução nossa).

Seu posicionamento não é de todo isolado. A necessidade de uma legislação que responsabilize a inteligência artificial é assunto que ganha tração. Começam, pois, a surgir cada vez mais vozes nesse sentido:

O direito penal moderno não deixa outra possibilidade a não ser identificar um indivíduo envolvido em qualquer forma de atividade e a criação de IA. Acreditamos que é impossível concordar com as posições antropocêntricas de concretização da

responsabilidade criminal, pois como mencionado acima, devido às propriedades essenciais da IA, elas contradizem tanto o princípio da culpa quanto da justiça. Acreditamos que é inaceitável postergar mudanças legislativas até que apareça a real tendenciosidade de cometer crimes de IA, pois neste caso a criminalização será extemporânea, portanto, é necessário desenvolver mudanças legais no menor tempo possível. (KIRPICHNIKOV, PAVLYUK, GREBNEVA, OKAGBUE, 2020, p. 5, tradução nossa).

No ano de 2015, em Baunatal, cidade próxima a Frankfurt, Alemanha, em uma planta industrial da Volkswagen, uma máquina acabou matando um dos funcionários da empresa. O jovem de 22 anos fazia parte de uma equipe que estava montando o robô estacionário quando este o agarrou e esmagou contra uma placa de metal. As especulações iniciais indicaram que a culpa foi de um erro humano, e não problema do robô, que pode ser programado para realizar várias tarefas no processo de montagem. À época, promotores se questionavam e estudavam se ingressariam com ações judiciais e, sobretudo, contra quem (PHILLIP, 2015).

Três anos depois, um carro autônomo operado pela empresa de transporte por aplicativo Uber - e com um motorista reserva de emergência ao volante - atropelou e matou uma mulher em uma rua em Tempe, cidade no estado do Arizona, Estados Unidos. Uma investigação preliminar mostrou que o veículo se movia a cerca de 65 quilômetros por hora quando atingiu uma transeunte, que caminhava com sua bicicleta na rua. O carro não havia desacelerado antes do impacto e o motorista de segurança do Uber não havia mostrado sinais de desatenção. O tempo estava claro e seco (WAKABAYASHI, 2018).

No dia 05 de janeiro de 2020, deu-se início a um inquérito militar no âmbito do exército estadunidense para averiguar um ataque, ocorrido no dia anterior, num colégio militar em Tripoli, capital da Líbia, que tirou a vida de 32 estudantes e deixou vários lesionados. O relatório, finalizado pelo Conselho de Segurança das Nações Unidas, apontou que o ataque foi realizado por um drone não tripulado e sem qualquer ordem de comando humana. A máquina atirou um míssil, destinado a atacar veículos blindados, guiado por um sistema semiautomático a laser, atravessou duas paredes e explodiu no pátio da escola (CONSELHO DE SEGURANÇA DA ONU, 2021).

A inteligência artificial é realidade dada. A arte no século XX adiantou muitas dessas reflexões e aplicações práticas. Na mesma medida, o Direito Penal enquanto ferramenta de controle social também se mostra presente e não dá sinais de arrefecimento. Urge, assim, a necessidade de inflexão entre os dois campos de conhecimento humano.

### 3. Crítica da razão punitiva artificial

- Vamos tentar outra (...). Você está assistindo a um filme antigo na TV, um filme de antes da guerra. Está acontecendo um banquete. A entrada – ele pulou a primeira parte da pergunta – consiste em cão cozido, recheado com arroz.

- Ninguém mataria e comeria um cão – Luba Luft disse. – Eles valem uma fortuna. Talvez fosse um cachorro de mentira: uma imitação. Certo? Mas eles são feitos de fios e motores; não podem ser comidos (DICK, 2017, p. 139).

Um questionamento que deve surgir ao abordar a problemática é da razão para punir inteligências artificiais. Veja-se. Hulsman (2012, p. 36) afirma que a resposta estatal criminal utilizada hoje não é verdadeiramente uma punição, ou ao menos um processo condizente com os objetivos declarados do Direito Penal de prevenção e repressão. É, em realidade, um processo comunicativo travestido de resposta efetiva, com uma série de ritos conhecidos e aceitos pelo público, mas que acabam ocultando a verdadeira face da criminalidade, suas razões e impactos da criminalização. Esse símbolo trazido pela sanção estatal já seria importante argumento contrário à resposta penal contra máquinas por alguns motivos.

Em primeiro, se a privação de liberdade de seres humanos já é questionável sob a óptica da Criminologia radical, menos razão haveria em expandir o Direito Penal para além das fronteiras que perpassam a da pessoa humana. A responsabilização penal da pessoa jurídica, como visto no introito, já padece de algumas críticas de reforçar o estigma de panaceia da dogmática criminal.

Além disso, do ponto de vista político, tendo em mente que o Direito Penal é uma das mais fortes ferramentas comunicativas que o Estado gere, como comentado por Hulsman, não haveria interesse no controle de corpos não humanos. A pena privativa de liberdade é importante capital Estatal no processo imagético junto à sociedade. A mera inutilização ou alteração dos códigos-fonte não traz o mesmo impacto que o aprisionamento de pessoas na fábula do combate ao delito. Mesmo na hipótese trazida por Eduardovich (2017, p. 139), em que o Direito Penal poderia ser aplicado quando a inteligência artificial tivesse um suporte físico (um androide sendo destruído numa máquina compactadora, por exemplo), não seria ainda suficiente para suplantar a imagética prisão povoada de gente. Desta forma, embora sempre desejável restringir ao máximo (ou extinguir) o poder punitivo penal, reconhece-se que, enquanto existir, seu uso expansionista é mais atraente aos olhos de uma sociedade induzida por vingança.

Destaca-se que, se a tendência seguir como estudiosos preveem, com a tomada de decisões por computador sendo cada vez mais “orgânicas”, ou, ao menos, independente de ação humana, mais a conduta robótica tornar-se-ia uma conduta penalmente relevante. E, vez que a heurística artificial está cada vez mais presente em nosso cotidiano, a responsabilidade de uma máquina poderia ser importante quebra do nexo de causalidade. Se a tomada de decisão se deu exclusivamente dentro das redes neuronais de um robô, por consequência, o programador não poderia ser responsabilizado. Tampouco se estaria no âmbito do caso fortuito ou da força maior. A programação foi criada por uma pessoa com um objetivo de realizar uma tarefa ou um conjunto de atividades por ela prevista. Enquanto houver essa previsibilidade do comportamento da máquina, a responsabilidade poderia recair sobre o humano vez que com sua expertise o resultado indesejado seria (ou deveria ser) igualmente previsível. Desta forma, as categorias normativas da culpa se fariam presentes. Todavia, se o processo se torna independente, ou seja, as redes neurais passam a estabelecer seus próprios processos processamento de dados, em face do programador o nexo de causalidade não estaria mais presente ou, ao menos, mais difícil de ser configurada. Salvo se houver a obrigatoriedade de impor a limitação.

Sem embargo seja um pensamento tentador, não se pode ignorar o risco da autorização da criação de qualquer tipo de inteligência artificial, sem a imposição de limites ou controle dos *outputs*. É lugar-comum hodiernamente afirmar que a sociedade pós-moderna é uma sociedade que avança rapidamente tecnologicamente, mas que exponencia proporcionalmente os riscos (BECK, 2011, p. 28). No campo da ciência da computação, os danos podem ser mais severos por dois motivos principais. O primeiro é o abismo que existe acerca do domínio do uso das tecnologias da informação. Sendo a digitalização da vida muito mais exclusiva que inclusiva, dados pessoais tendem a ficar mais expostos vulnerabilizando quem sequer gozam dessa consciência. Comportamentos podem ser induzidos pelos algoritmos sem qualquer possibilidade de escolha e podem acabar definindo até mesmo a adoção de políticas públicas (MUNIZ; LEUGI; PEREIRA; PRZEYBILOVICZ; ALVES, 2021, p. 702).

O segundo motivo que se destaca é o uso cada vez mais crescente dos postulados robóticos pressionando os índices de desigualdade social, seja efetivamente ampliando a concentração de capital nos países e clusters econômicos já desenvolvidos (ALONSO; KOTHARI; REHMAN, 2020), seja pela precarização do trabalho e diminuição da capacidade de produção de emprego. Quanto a este último ponto, deve-se verificar, junto do primeiro

motivo, que a artificialização das relações de trabalho faz com que populações vulneráveis fiquem mais suscetíveis a experimentos que podem acarretar graves danos.

Inegável, nesse sentido, com tudo o apresentado, que a forma como a automação do pensamento ocorre deve levar a reflexões no âmbito do Direito Penal. Há a possibilidade de que programas sejam feitos com a certeza de que, ainda que possam originar resultados desastrosos, não observem resposta estatal alguma vez que passam a figurar longe da previsibilidade de quem os criou. Não se trata de impedir o avanço do *machine learning* (impossível, já que a marcha tecnologia impede que o ser humano volte ao estado anterior àquela tecnologia). Mas de impedir que ela seja independente e verdadeiramente autônoma.

Dessa maneira, as limitações decisórias dos robôs devem ser construídas, ainda que contrassenso à própria ideia de aprendizado de máquina. Se não limitado, pelo menos que passe por um crivo final humano. Muito embora o *input* de dados possa fugir da previsibilidade do programador, o algoritmo e os *outputs* podem ser pré-estabelecidos.

Observando agora o lado humano da discussão sobre inteligência, o conceito de personalidade atualmente concebido carrega em si o elemento que é buscado, por exemplo, no referido Teste de Turing. Ao tentar diferenciar a inteligência artificial da do ser humano, tenta-se perceber se o objeto de exame tem consciência de si mesmo e, mais, se tem capacidade de valorar essa consciência (estima) e de reconhecer e valorar a consciência do outro (empatia). Além disso, não se trabalha com o livre-arbítrio enquanto fenômeno ilimitado, ou seja, sem qualquer tipo de referencial. Sugere-se, então, ao invés do uso de *livre-arbítrio*, a ideia de autodirecionamento. O ser humano até então teria alguma capacidade de escolha, ou seja, de agir contra seu desejo primal (que poderia ser interpretado freudianamente como a instância do *id*) e externar conforme valores introjetados (American Psychiatric Association, 2015, p. 762). Assim, busca-se aqui afastar a justificativa do livre-arbítrio para punir pessoas.

No campo da neurociência, a conclusão não é muito diferente. Porém, adianta-se desde já certa preocupação com o viés determinista biológico adotado por alguns neurocientistas, na medida em que acabam se aproximando do positivismo lombrosiano e da concepção que o comportamento humano é ditado fisiologicamente. Aqui, a discussão não se encerra no entendimento da conduta humana, mas leva ao embate sobre a resposta a ser dada a quem pratica aquela que foi classificada politicamente como delito. Ignorar fatores externos, como aqueles trazidos pela sociologia e pela psicologia, pode levar à defesa de que existem humanos

propensos ao comportamento desviante. Aceitando essa premissa, ações preventivas ou mesmo de neutralização passam a ganhar fundamentação (CASTRO, 2008, p. 5).

Em todo modo, é certo que a ideia de livre-arbítrio vai se tornando cada vez mais distante da realidade acadêmica.

Ainda nesse diapasão, alguns estudiosos deterministas, baseados nas recentes pesquisas neurocientíficas, como Gerhard Roth, afirmam e comprovam a efetiva ausência da liberdade humana. Sustentam tais autores que os indivíduos são determinados pelo inconsciente e que a ideia de livre-arbítrio é uma ilusão, o que acaba por reduzir o ser humano a verdadeiro autômato ou, mesmo, por que não dizer, em autêntico “relógio sofisticado” (CERQUEIRA; ALBAN, 2014, p. 240).

Cabe observar aqui um ponto de justaposição entre as duas reflexões ora apresentadas. Apesar da referência ao humano como um autômato, o que, novamente, acaba o aproximando dos robôs, é importante o recurso psicológico e sociológico no sentido de apontar que o inconsciente não é algo apartado do próprio homem. Não se trata de entidade ensimesmada, logo, ele ainda estará suscetível a ingerências sensoriais, culturais, sociais etc. Não há pessoas iguais, nem mesmo em contextos aparentemente iguais.

Em relação ao ser humano, contudo, as limitações do pensar e agir estão condicionadas ao imponderável da vida e eventos que fogem verdadeiramente do controle do sujeito. Se a consciência humana pode ser vista como um amálgama dos estímulos percebidos ao longo da vida e fatores biológicos, é certo que a personalidade humana tem um grau de complexidade que vai além de padrões binários.

#### 4. Conclusão

A elucubração sugerida leva a outra mais concreta e mais próxima da atual realidade. Se o robô não é criação orgânica da natureza, senão uma ferramenta criada e limitada por ação humana, deve-se verificar o grau, caso haja, de responsabilidade criminal do programador.

Por um lado, a responsabilidade penal da máquina seria interessante fronteira para o nexos de causalidade, impedindo que o programador venha a ser punido por atos decididos pelas conexões neuronais eletrônicas. De outro, não restringir a atividade programadora pode levar a consequências graves, tanto para indivíduos, quanto para grupos indeterminados.

Os procedimentos cognitivos e expressivos de humanos e computadores são, até o presente momento, diversos, mas cada vez mais próximos. O que diferencia não é o livre-arbítrio, visto que, como argumentado, não existe. Mas o espaço mental decisório é suscetível,



para o ser humano, de ingerências sociais e psicológicas que não são perceptíveis às máquinas. Por mais complexo e intrincado que possa ser um sistema neuronal artificial, é possível com boas linhas de programação, alcançar um resultado que seja mais condizente com valores e metas democráticos.

Os estudiosos atualmente trabalham com a possibilidade de punir robôs partindo do pressuposto da imponderabilidade dos seus atos. Questiona-se o porquê de não os desenvolver tendo por base linhas-guia que façam com que seus resultados sejam ao menos previsíveis. Se questões técnicas se impõem, então, que não seja utilizada a inteligência artificial fora de ambiente controlado ou que, ao menos, se aplique uma instância humana antes de operacionalizar os frutos obtidos digitalmente.

Sugere-se, então, a adoção de programas de *compliance* na esfera das ciências da computação. A conformidade regulatória seria fundamental para que os programadores alcancem, em seus esforços em prol do avanço tecnológico, medidas que garantam o cumprimento das leis, políticas e regulamentações relevantes. Concomitante a isso, a criação de *offices* que fiscalizem e estabeleçam tais diretrizes. Por mais contraproducente que possa ser para a marcha do caminhar eletrônico, urge há tempos a necessidade de cuidar para que esse avanço não ultrapasse os limites do risco socialmente tolerável. Sugestões para tanto envolvem a constante tomada de consentimento dos titulares de dados sobre seu tratamento, manutenção permanente do histórico de manipulações dos dados e dos algoritmos, definição de resultados previsíveis indesejados e aplicação de estratégias que os evitem e presença constante de etapas de verificação humana sobretudo ao final do processo neuronal artificial.

O futuro sugere uma integração cada vez mais intensa entre pessoa e computador. Os expedientes serão antecipados de análises volumétricas de dados e informações que, sem o uso da máquina se tornam inteligíveis, mas que vão auxiliar no combate de pandemias, descoberta de curas, manejo de alimentos e infinitas outras possibilidades. Entretanto, da mesma forma que já se viu em outras ferramentas humanas, a mesma pedra lascada que ara a terra, mata o homem. Se é possível hoje adotar medidas preventivas, a diminuição da aceleração das inovações tecnológicas é preço menor a se pagar.

## 5. Referências

ALONSO, Cristian; KOTHARI, Siddharth; REHMAN, Sidra. *Como a inteligência artificial pode aumentar a distância entre as nações ricas e pobres*. 03 dez. 2020. Disponível em:

<https://www.imf.org/pt/News/Articles/2020/12/02/blog-how-artificial-intelligence-could-widen-the-gap-between-rich-and-poor-nations>. Acesso em: 10 ago. 2021.

American Psychiatric Association. *Manual diagnóstico e estatístico de transtornos mentais*. Trad. Maria Inês Corrêa Nascimento. 5ª ed. Porto Alegre: Artmed, 2014.

ASIMOV, Isaac. *Eu, Robô*. São Paulo: Aleph, 2014. Tradução de: Aline Storto Pereira.

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

BOYD, Danah; CRAWFORD, Kate. Six Provocations for Big Data. In: *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*, set. 2011. Disponível em <https://ssrn.com/abstract=1926431>. Acesso em: 05 mar. 2021.

BUSATO, Paulo César. *Direito Penal: Parte Geral*. Vol. 1. 5ª ed. São Paulo: Atlas, 2020.

CASTRO, Lola Aniyar de. El regreso triunfal de Darwin y Lombroso: las diferencias humanas en la criminología angloparlante presente en los simposios internacionales de criminología de Estocolmo. In: *Capítulo criminológico: revista de las disciplinas del Control Social*, Maracaibo, vol. 36, n. 4, 2008, p. 5-25, Universidad del Zulia.

CERQUEIRA, Marina; ALBAN, Rafaela. Culpabilidade, livre-arbítrio e neurociências. In: BUSATO, Paulo César (org.). *Neurociência e Direito Penal*. São Paulo: Atlas, 2014, p. 239-272.

CHERELUS, Gina; TENNERY, Amy. *Microsoft's AI Twitter bot goes dark after racist, sexist tweets*. Reuters. 2016. Disponível em: <https://www.reuters.com/article/us-microsoft-twitter-bot-idUSKCN0WQ2LA>. Acesso em: 05 mar. 2021.

CLARKE, Ronald V.; CORNISH, Derek B. *The Reasoning Criminal: rational choice perspectives on offending*. Londres: Routledge, 2014.

CONSELHO DE SEGURANÇA DA ONU. *Resolução 1970 (2011)*. Index: S/2021/229, 08 mar. 2021. Disponível em: <https://undocs.org/S/2021/229>. Acesso em: 01 ago. 2021.

ČAPEK, Karel. *Rossum's Universal Robots: a fantastic melodrama in three acts and an epilogue*. Nova Iorque: Samuel French, 1923. Tradução em inglês por Paul Selver. Disponível em: <https://www.gutenberg.org/files/59112/59112-h/59112-h.htm>. Acesso em: 05 mar. 2021.

DICK, Philip K. *Androides sonham com ovelhas elétricas?*. Trad. Ronaldo Bressane. 2ª ed. São Paulo: Aleph, 2017.

EDUARDOVICH, Radutniy Oleksandr. Criminal liability of the artificial intelligence. *Problems Of Legality*, n. 138, p. 132-141, 27 set. 2017. Yaroslav Mudryi National Law University. <http://dx.doi.org/10.21564/2414-990x.138.105661>.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. *Deep Learning*. Cambridge: MIT Press, 2016.

Hanson Robotics. *Sophia*. [online][S.I.]. Disponível em <https://www.hansonrobotics.com/sophia/>. Acesso em: 05 mar. 2021.

HIRSCHI, Travis. *Causes of Delinquency*. Berkeley: University Of California Press, 1969.

HULSMAN, Louk. Alternativas à justiça criminal. In: PASSETTI, Edson. *Curso Livre de Abolicionismo Penal*. 2ª ed. Rio de Janeiro: Revan, 2012, p. 35-68.

KIRPICHNIKOV, Danila; PAVLYUK, Albert; GREBNEVA, Yulia; OKAGBUE, Hilary. Criminal Liability of the Artificial Intelligence. *E3S Web Of Conferences*, [S.L.], v. 159, p. 04025-04035, 2020. EDP Sciences. <http://dx.doi.org/10.1051/e3sconf/202015904025>.

KOZA, John R.; BENNETT, Forrest H.; ANDRE, David; KEANE, Martin A.. Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. In: *Artificial Intelligence in Design '96*. Springer, Dordrecht. 1996, pp. 151–170.

MARTINELLI, João Paulo O.; BEM, Leonardo Schmitt de. *Lições Fundamentais de Direito Penal: Parte Geral*, 4ª ed. São Paulo: Saraiva, 2019.

MITCHELL, Tom M. *Machine Learning*. Nova Iorque: McGraw-Hill Science, 1997.

MORAVEC, Hans Peter. *Robot: technology*. Technology. 2021. [online][S.I.]. Disponível em: <https://www.britannica.com/technology/robot-technology>. Acesso em: 05 mar. 2021.

MUNIZ, Cátia Regina; LEUGI, Guilherme Bergo; PEREIRA, Cleide de Marco; PRZEYBILOVICZ, Érico; ALVES, Angela Maria. Uma análise sobre exclusão digital durante a pandemia de COVID-19 no Brasil: Quem tem direito às cidades inteligentes?. *Revista de Direito da Cidade*, v. 13, n. 2, p. 700-728, 2021. Universidade do Estado do Rio de Janeiro. <https://doi.org/10.12957/rdc.2021.54909>.

Oxford Languages. Oxford English Dictionary. [online][S.I.]. Disponível em <https://languages.oup.com/>. Acesso em 10 de julho de 2021.

PARK, Robert E.; BURGESS, Ernest W.. *Introduction to the Science of Sociology*. Chicago: The University Of Chicago, 1921. Disponível em: <http://www.gutenberg.org/files/28496/28496-h/28496-h.htm>. Acesso em: 05 mar. 2021.

PHILLIP, Abby. *Robot grabs man, kills him in German car factory*. Washington Post. 2015. Disponível em: <https://www.washingtonpost.com/news/morning-mix/wp/2015/07/02/robot-grabs-man-kills-him-in-german-car-factory/>. Acesso em 09 ago. 2021.

POOLE, David; MACKWORTH, Alan; GOEBEL, Randy. *Computational Intelligence: A Logical Approach*. New York: Oxford University Press, 1998.

REALE JUNIOR, Miguel. *Fundamentos de Direito Penal*, 5ª ed. São Paulo: Grupo GEN, 2020.

RUSSELL, Stuart J.; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*, 2ª ed. New Jersey: Prentice Hal, 2003.

SILVA SÁNCHEZ, Jesús María. *La Expansión del Derecho Penal: Aspectos de la Política criminal en las sociedades postindustriales*, 3ª ed. Montevideo: B de F, 2011.

SOARES, Ricardo Mauricio Freire. *O Princípio Constitucional da Dignidade da Pessoa Humana*. São Paulo: Editora Saraiva, 2009.

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

TONET, Ivo. *Método científico: uma abordagem ontológica*. São Paulo: Instituto Lukács, 2013.

TURING, Alan Mathison. I. - Computing Machinery and Intelligence. *Mind*, v. LIX, n. 236, p. 433-460, 1 out. 1950. Oxford University Press (OUP). <http://dx.doi.org/10.1093/mind/lix.236.433>.

UN Neww Centre. *At UN, robot Sophia joins meeting on artificial intelligence and sustainable development*. 2017. Disponível em: <https://www.un.org/en/desa/un-robot-sophia-joins-meeting-artificial-intelligence-and-sustainable-development#:~:text=At%20UN%2C%20robot%20Sophia%20joins%20meeting%20on%20a%20artificial%20intelligence%20and%20sustainable%20development,-Artificial%20intelligence%20has&text=%5Bartificial%20intelligence%5D%20could%20help%20proficiently,latest%20and%20most%20advanced%20robot..> Acesso em: 05 mar. 2021.

WAKABAYASHI, Daisuke. *Self-Driving Uber Car Kills Pedestrian in Arizona, Where Robots Roam*. The New York Times. 2018. Disponível em <https://www.nytimes.com/2018/03/19/technology/uber-driverless-fatality.html>. Acesso em: 09 ago. 2021.

ZOJA, Luigi. *História da arrogância: psicologia e limites do desenvolvimento humano*. São Paulo: Axis Mundi Editora, 2000.