

DE MÁQUINAS E SERES VIVOS: QUEM AGE EM RESULTADOS CRIMINAIS DECORRENTES DE DECISÕES CIBERNÉTICAS¹

*Of machines and living things:
who acts in criminal results
derived from cybernetic decisions?*

Paulo César Busato²

RESUMO:

O objetivo deste trabalho é abordar o problema da relação entre o direito penal e as tecnologias de inteligência artificial (doravante I.A.) a partir de uma metodologia que consiste na exposição de alguns problemas relevantes da dogmática penal diante de tais pressupostos, seguida de uma proposta de recolocação desses problemas a partir de uma abordagem do sistema de imputação baseada na filosofia da linguagem. Espera-se que essa *abordagem* ofereça alternativas para o debate sobre o tema.

Palavras-chave: Inteligência Artificial. Direito Penal. Filosofia da Linguagem.

ABSTRACT:

The objective of this work is to address the problem of the relationship between criminal law and artificial intelligence technologies (hereinafter A.I.) based on a methodology that consists of exposing some relevant problems of criminal dogmatics in light of such assumptions, followed by a proposal for replacement of these problems based on an approach to the imputation system based on the philosophy of language. It is hoped that this approach will offer alternatives for the debate on the topic.

Keywords: Artificial Intelligence. Criminal Law. Philosophy of Language

¹ Tradução do artigo originalmente em espanhol, “De máquinas y seres vivos: ¿Quién actúa en los resultados delictivos derivados de decisiones cibernéticas?” realizada por Marcos Wachowicz, Professor de Direito do Curso de Graduação e docente do Programa de Pós-Graduação de Direito pela Universidade Federal do Paraná, coordenador líder do Grupo de Estudos em Direito Autoral e Industrial – GEDAI/UFPR, coordenador da Rede Ibero Americana de Propriedade Intelectual – RIAPI e presidente do Instituto Observatório do Direito Autoral - IODA.

² O autor é Professor de Direito Penal da Universidade Federal do Paraná, Brasil, Doutor pela Universidade Pablo de Olavide, Sevilha, Espanha e Promotor de Justiça do Ministério Público do Paraná, Brasil.

SUMÁRIO

1. O PROBLEMA; 2. TENTATIVAS DE SOLUÇÃO; 2.1. TENTATIVAS DE ESTABELECEER RESPONSABILIDADE PELO MESMO I.A; 2.1.1. INICIATIVAS DE POLÍTICAS; 2.1.2. EQUIVALÊNCIA FUNCIONAL ENTRE HUMANOS E MÁQUINAS; 2.2. A FRAGMENTAÇÃO DAS PROPOSIÇÕES NA DOGMÁTICA FUNCIONALISTA; 3. REVISAR O PROBLEMA DE UM PONTO DE VISTA ALTERNATIVO; 3.1. OS LIMITES DO PROBLEMA: AS CLASSES DE I.A; 3.2. RESULTADOS IMPREVISÍVEIS E O ELEMENTO SUBJETIVO; 3.3. A ABORDAGEM EQUIVOCADA DA DOCTRINA CRIMINAL CONTEMPORÂNEA; 3.4. UM PASSO ATRÁS ATÉ A RAIZ DO PROBLEMA; 3.5. REPENSANDO O PROBLEMA; 3.5.1. SOBRE O QUÊ. AÇÃO COMO UMA EXPRESSÃO DE SIGNIFICADO; 3.5.2 EM QUEM; REFERÊNCIAS.

1 O PROBLEMA

Atualmente, foram identificados diferentes casos em que a causa de resultados juridicamente desvalorizados - incluindo lesões a direitos jurídicos fundamentais, como a vida e a integridade física - é realizada com a intervenção do I.A.. Isso tem chamado a atenção para a necessidade de se estabelecer algum tipo de controle social penal³. São frequentes os relatos de lesões resultantes da intervenção de robôs em cirurgias⁴ ou em processos industriais⁵, de armas letais autônomas conduzidas por drones⁶, dos chamados ‘carros autônomos’, de operações autônomas no

³ Sobre essa necessidade de identificar “*quem é o culpado*” com o advento dos carros autônomos, consulte (LOH & LOH, 2017, pp. 35 et seq.). E também (GLESS & WEIGEND, 2019, pp. 41-42).

⁴ Cf. (SULLIVAN & SCHWEIKART, 2019, p. 162).

⁵ Hallevy relata que, já em 1981, um trabalhador japonês de 37 anos, em uma fábrica de motocicletas, foi morto por um robô equipado com I.A. que, equivocadamente, identificou o trabalhador como uma ameaça à sua missão e calculou a forma mais eficiente de eliminar essa ameaça, que foi empurrá-lo, com seu potente braço mecânico, na direção de outra máquina em funcionamento nas proximidades, matando-o instantaneamente, seguindo a programação para evitar perturbações que pudessem interferir em sua missão. Cf. (HALLEVY, 2016, p. 171-172).

⁶ Cf. (GLESS & WEIGEND, 2019, p. 38).

mercado de móveis que produzem *colisões relâmpago*⁷ e muitos outros. E tudo isso ainda se refere ao que é frequentemente definido como a primeira geração de I.A.⁸, em um fenômeno que pode ser descrito como uma verdadeira metamorfose⁹.

No entanto, se é verdade que a tecnologia evolui rapidamente, também é verdade que seus condutores são justamente as pessoas, e os avanços geralmente consistem em produzir resultados que, embora acarretem riscos, são, em regra, computados atuarialmente como aceitáveis em comparação com os benefícios que acarretam¹⁰. Essa dinâmica com relação aos riscos foi incorporada à dogmática jurídica penal por meio do chamado funcionalismo¹¹. Mas parece que esse paradigma está começando a desmoronar.

Embora seja normal que novos desafios dogmáticos produzam discrepâncias de opinião, parece que em casos envolvendo I.A. parece haver um espanto especial com os resultados oferecidos pelo sistema de imputação, que não parecem alcançar os resultados esperados.

⁷ Exemplos são dados em (WHITE & BAUM, 2020, pp. 66 e ss.) e (SOUSA, 2020, p. 60). Para detalhes sobre os casos dos carros autônomos Uber e Tesla, consulte (JANUÁRIO, 2020, p. 97 e nota p. 97, nota 6).

⁸ Em relação aos estratos com os quais a I.A. é classificada, temos a primeira geração identificada com o termo *inteligência artificial restrita*, que define a capacidade de realizar tarefas específicas, mas já há o prognóstico em uma segunda geração, denominada *inteligência geral artificial*, na qual se supõe que as máquinas serão capazes de raciocinar, planejar e resolver autonomamente problemas gerais além de sua programação e uma terceira, denominada *superinteligência artificial*, em que serão capazes de alcançar a autoconsciência e o desenvolvimento de habilidades científicas e sociais. Veja (KAPLAN & HAENLEIN, 2019, p. 16). Veja também (TURNER, 2019, especialmente pp. 28 e seguintes). Para uma ampla discussão sobre as perspectivas da superinteligência, consulte (BOSTROM, 2014).

⁹ Beck dedicou um de seus últimos escritos para explicar isso (BECK, 2018, pp. 30-39). O termo “metamorfose digital do mundo” é usado na análise criminal por Hilgendorf, referindo-se a Di Fabio e Schwab em (HILGENDORF, 2020, p. 43).

¹⁰ O argumento se refere ao argumento seminal (BECK, 2011).

¹¹ É importante observar que o funcionalismo em matéria penal não deve ser entendido como tendo o mesmo significado que o funcionalismo em termos filosóficos e sociológicos. Embora apareça alguma conexão, não há, de fato, uma transposição direta. Nesse sentido, ver (VIVES ANTÓN, 2011, pp. 442-443).

Se já estamos convivendo com a revolução tecnológica e com vários avanços científicos há algum tempo, o que há de tão diferente nesses casos que levou a uma reação tão particular na doutrina criminal?

Parece que o problema tem a ver com o fato de que a intervenção da I.A. produz resultados que nem sempre são previstos por seus programadores ou usuários¹². A partir disso, surge uma primeira dificuldade para o direito penal, derivada de sua característica de exigir um elemento subjetivo para a atribuição de responsabilidade.

Se houver resultados significativos desvalorizados fora do escopo da previsibilidade, ainda há espaço para a responsabilidade criminal? É possível estabelecer a responsabilidade criminal pelos resultados se, no momento da intervenção, sua existência era completamente incerta? De fato, o elemento subjetivo é frequentemente considerado um obstáculo à responsabilização em casos de danos causados por sistemas autônomos de I.A.¹³ e o problema é delimitado com base na consideração de que o dano atribuído à máquina não está incluído na representação da pessoa que a dispõe ou opera, escapando da intenção e até mesmo da previsibilidade. Como os resultados podem estar fora do escopo da intenção ou da previsibilidade do programador ou do usuário da máquina, falamos de uma *lacuna de responsabilidade*¹⁴.

2 TENTATIVAS DE SOLUÇÃO

Para superar a *lacuna de responsabilidade*, a doutrina criminal majoritária - ainda ancorada no paradigma do risco funcionalista - debateu a questão, mas chegou a um beco sem saída.

Por um lado, tentou-se criar uma responsabilidade penal para a própria I.A., reconhecendo-a como uma pessoa para fins penais¹⁵; por outro lado, tentou-se resolver os problemas com base em estruturas dog-

¹² Nesse sentido, (HALLEVY, 2016, p. 172) e (JANUÁRIO, 2020, pp. 95-96).

¹³ Assim, em (SOUSA, 2020, pp. 70-72).

¹⁴ Assim, em (SOUSA, 2020, pp. 74).

¹⁵ (HALLEVY, 2016, pp. 186-193).

máticas derivadas da matriz funcionalista do risco, com resultados não apenas insuficientes, mas também sem um mínimo de uniformidade.

2.1 Tentativas de estabelecer responsabilidade pelo mesmo I.A

A ideia de uma *personalidade jurídica eletrônica* pode ser identificada como um produto de dois impulsos distintos. O primeiro é o impulso legislativo que buscou estabelecer uma dinâmica de controle jurídico da evolução científica no campo da robótica; o segundo é o argumento da equivalência funcional entre a personalidade humana e a personalidade eletrônica, que depende de se tomar como ponto de partida para essa equivalência a concepção de ação ancorada na mecânica newtoniana e a expressão da subjetividade baseada no modelo cartesiano do *cogito ergo sum*.

Afirma-se que “a responsabilidade criminal de um I.L.A. de acordo com um modelo de responsabilidade direta não é diferente da responsabilidade criminal de um ser humano. Em alguns casos, certos ajustes são necessários, mas, substantivamente, é a mesma responsabilidade criminal, que se baseia nos mesmos elementos e é avaliada da mesma forma”¹⁶.

2.1.1 Iniciativas de políticas

O impulso normativo de atribuir personalidade aos IAs¹⁷, com sua conversão em um centro jurídico de imputação, já era esperado pela doutrina¹⁸ há algum tempo.

De fato, o Parlamento Europeu, na resolução de 16 de fevereiro de 2017, referente a recomendações à Comissão Europeia sobre legislação civil sobre robótica e I.A., apontou para a criação de um status legal para

¹⁶ (HALLEVY, 2016, p. 193).

¹⁷ Para uma análise geral da questão, ver (BARBOSA, 2021, pp. 97 e ss.).

¹⁸ Cf. (GIUFRIDA, 2019, p. 444).

robôs¹⁹, mesmo que isso tenha provocado alguma controvérsia com uma reação significativa.

2.1.2 Equivalência funcional entre humanos e máquinas

De um ponto de vista funcionalista, argumenta-se que as I.A. atingiram níveis de desenvolvimento que permitem identificá-las como sub-sistemas sociais capazes de promover a estabilização das expectativas da comunidade, constituindo, assim, equivalentes funcionais capazes de reunir os elementos objetivos e subjetivos para realizar a imputação jurídico-penal.

Afirma-se que a evolução das I.A. gerou nelas características humanas como a comunicação, a internalização do conhecimento, a capacidade de aprendizado, a tomada de decisões e até mesmo a criatividade e a espontaneidade²⁰. Há ainda aqueles que se referem à necessidade de autoconsciência da I.A. para a afirmação de sua responsabilidade²¹, mas ainda reconhecem que a evolução tecnológica já está na eminência de produzi-la²².

Nesse caso, seria possível admitir que as máquinas raciocinam exatamente como os seres humanos e, portanto, deveriam merecer tratamento jurídico igual.

¹⁹ Resolução do Parlamento Europeu de 16 de fevereiro de 2017. “59. Insta a Comissão a explorar, analisar e considerar, na sua avaliação de impacto do seu futuro instrumento legislativo, as implicações de todas as soluções jurídicas possíveis, tais como: [...] f) Criar um *estatuto* jurídico específico para os robôs a longo prazo, de modo que pelo menos os robôs autônomos mais sofisticados possam ser identificados como tendo o estatuto de pessoas eletrônicas responsáveis pela reparação dos danos que possam causar e, eventualmente, afirmar a personalidade eletrônica nos casos em que os robôs tomem decisões autônomas ou interajam de outra forma com terceiros de forma independente”.

²⁰ Cf. (HALLEVY, 2016, p. 176) e (LOH & LOH, 2017, p. 38). Portanto, as críticas ao modelo de autorresponsabilidade da I.A. baseadas em argumentos relacionados à consciência, em um molde cartesiano, como os desenvolvidos pelo mesmo autor (SOUSA, 2020, p. 77), não são suficientes para distanciar tais tentativas.

²¹ (GLESS & WEIGEND, 2019, pp. 44-45).

²² (GLESS & WEIGEND, 2019, p. 45).

Desse ponto de vista, é feita uma comparação indevida entre máquinas, pessoas físicas e pessoas jurídicas²³, concluindo que os robôs equipados com inteligência artificial são capazes de realizar tanto o *actus reus* quanto o *mens rea* necessários para a responsabilidade criminal, ou seja, são capazes de atender aos requisitos objetivos e subjetivos da imputação²⁴.

Para apoiar isso, do ponto de vista do desempenho do ato (*actus reus*), uma capacidade de ação é identificada no I.A. que corresponde à capacidade de produzir um movimento físico que produz um resultado, referindo-se assim à ação como um dado ontológico²⁵.

Por outro lado, no que diz respeito à *mens rea*, uma vez aceito o pressuposto cartesiano de que a identidade humana é estabelecida com base no ato de pensar ou raciocinar, houve muitos argumentos a favor do desenvolvimento de uma responsabilidade jurídica penal das próprias máquinas com base no reconhecimento nelas de uma suposta “capacidade de pensar”. A capacidade de aprender de forma autônoma com o ambiente e de fazer escolhas diante dele tem sido um argumento frequente na doutrina para justificar o desenvolvimento de uma capacidade “pensante” da I.A. que a aproximaria do humano, o que, afinal, é o objetivo de todas as formas de I.A.

²³ Apontando para uma semelhança equivocada (JANUÁRIO, 2020, p. 100) e (HALLEVY, 2016, p. 173-174). Embora o melhor lugar para o desenvolvimento do tema fosse um estudo à parte, é claro que é possível apontar que a responsabilidade penal das pessoas jurídicas é exatamente o oposto da responsabilidade penal da I.A.. Basta salientar que, enquanto as pessoas jurídicas podem expressar linguisticamente suas intenções em suas ações (ver, a esse respeito, BUSATO, 2019, p. 98), embora nunca consigam produzir, por si só, o movimento físico newtoniano, essa é a única coisa que os I.A. conseguem, os quais não passam de replicadores de uma fórmula logarítmica completamente incapaz de expressar linguisticamente intenções. Com menos sucesso, também se busca diferenciar a RPPJ da responsabilidade penal dos I.A. a partir de um suposto “controle humano” (SOUSA, 2020, p. 78).

²⁴ Nesse sentido (HALLEVY, 2016, pp. 177-178).

²⁵ A doutrina continua a considerar as “condutas” como dados ontológicos. Vide os exemplos de (HALLEVY, 2016, p. 182-183 e 187). E repetem os mesmos erros básicos mesmo quando supõem “avançar” pelo acréscimo inútil de uma ou outra exigência valorativa aos mesmos dados ontológicos. Por exemplo, em (GLESS & WEIGEND, 2019, pp. 46-48).

De fato, os autores que defendem a capacidade subjetiva da I.A. tendem a reduzir o requisito da dimensão subjetiva da capacidade de responsabilidade penal ao “conhecimento”, definido como “a recepção sensorial de dados factuais e a compreensão de tais dados”²⁶ e exatamente a ideia de que robôs equipados com I.A. são capazes de aprender com o ambiente e reagir de forma a evitar determinados comportamentos que faz com que a doutrina²⁷ aponte para uma equivalência dessa reação ao fundamento da culpabilidade, baseada na capacidade de agir de outra forma. Assim, afirma-se que “sob uma perspectiva funcional, a culpa é atribuída à pessoa que é identificada como a causa responsável pelo conflito social sedimentado no fato punível”²⁸.

O problema muito claro aqui é de substância e tem a ver com o entendimento do próprio termo *inteligência*. Em primeiro lugar, como Max Tegmark²⁹ afirmou com razão, é curioso ver como os mesmos pesquisadores inteligentes de IA não conseguem chegar a um consenso sobre o que define a inteligência. A armadilha é cair no reducionismo ao equiparar a inteligência à capacidade de lógica, compreensão, planejamento, consciência emocional, autoconsciência, criatividade, solução de problemas ou aprendizado. Nenhuma delas pode traduzir totalmente a inteligência, na medida em que a linguagem social reconheceu outra dimensão da inte-

²⁶ Assim (HALLEVY, 2016, p. 188).

²⁷ Assim, por exemplo, (GLESS & WEIGEND, 2019, p. 42) afirmam que “os seres humanos assumem a culpabilidade criminal quando intencionalmente ou de forma imprudente cometem um delito criminal e, assim, reconhecem a injustiça de seu comportamento e são capazes de evitá-lo de forma executória. No entanto, como vimos, os agentes inteligentes também são capazes de aprender e, pelo menos em um sentido figurado, agir de forma autônoma”.

²⁸ (GLESS & WEIGEND, 2019, p. 50).

²⁹ “Minha esposa e eu recentemente tivemos a sorte de participar de um simpósio sobre inteligência artificial organizado pela Fundação Nobel da Suécia e, quando um painel dos principais pesquisadores de IA foi solicitado a definir inteligência, eles discutiram sem chegar a um consenso. Nós nos divertimos com isso: não havia acordo sobre o que é inteligência, nem mesmo entre os pesquisadores de inteligência inteligente! Assim, ficou claro que não havia uma definição “correta” predominante de inteligência. Em vez disso, muitas outras competiam, incluindo a capacidade de lógica, compreensão, planejamento, consciência emocional, autoconsciência, criatividade, resolução de problemas e aprendizagem”. (TEGMARK, 2017, p. 49). As dificuldades com relação a isso são descritas de forma semelhante em (TURNER, pp. 7-22).

ligência, chamada de inteligência emocional³⁰. De fato, o próprio termo *inteligência artificial*, em uma inspeção mais minuciosa, deve ser suficiente para que se perceba que é um engano, um artifício, uma emulação incapaz de alcançar a equivalência.

2.2 A fragmentação das proposições na dogmática funcionalista

Muitos têm optado por uma solução para os problemas sem recorrer a uma base teórica que ancora a responsabilidade penal em uma reprovação moral e em um sistema de imputação que dá prevalência às estruturas lógico-objetivas sobre os princípios gerais do direito penal. Quase que imediatamente, ao se perceber o resultado desvalioso - que, nesses casos, os mais marcantes tendem a ser a morte ou a lesão -, a discussão se volta para os critérios de distribuição da responsabilidade.

Em nosso ambiente dogmático, os dilemas entre salvar a vida ou a integridade física dos passageiros ou de terceiros envolvidos no trânsito³¹; se o programador e os participantes da produção do veículo podem se valer de causas de justificação³²; casos de conflito de deveres³³; responsabilidade pelo produto³⁴ e os limites do risco permitido³⁵.

No ambiente *da common law*, a característica utilitarista dos *dilemas do caso do carrinho* também é discutida³⁶, onde se discute o que se encontra no programa I.A. de veículos autônomos em situações de deveres conflitantes e suas variantes.

³⁰ Veja, sobre isso, (GOLEMAN, 1995); (BRANCO, 2004) e (KENDEL & SCHWARTS, 1997).

³¹ Isso é discutido, por exemplo, por (HÖRNLE & WOHLERS, 2019, pp. 123-153) e (JANUÁRIO, 2020, pp. 102-117).

³² Cf. (WEIGEND, 2019, pp. 109-122) e (COCA VILA, 2019, pp. 155-189).

³³ Consulte (HILGENDORF, 2019, pp. 65-86) e (GRECO, 2019, pp. 191-201).

³⁴ (ESTELLITA, 2019, p. 33).

³⁵ Por exemplo, (JANUÁRIO, 2020, pp. 120 e ss.).

³⁶ Consulte (LOH & LOH, 2017, pp. 38-42).

Há duas notas distintivas muito claras nessas lucubrações teóricas: a primeira é a absoluta artificialidade (perdoe o trocadilho) das primeiras proposições, que não se encaixam perfeitamente na estrutura desprovida do status “*personae*” do I.A.; a segunda é a arrogância de insistir na preservação de um esquema dogmático herdado de estruturas fracassadas, completamente incapaz de resolver os casos com um mínimo de uniformidade e segurança jurídica.

3 REVISAR O PROBLEMA DE UM PONTO DE VISTA ALTERNATIVO

A crítica à proposta de estabelecer a responsabilidade criminal da própria I.A. baseia-se no argumento utilitarista muito fraco de que sua admissibilidade levaria à falta de responsabilização das pessoas envolvidas na produção, distribuição e uso das máquinas³⁷ ou porque não parece possível estabelecer algum tipo de punição criminal³⁸.

Na tentativa de melhor situar a discussão do problema, propõe-se um caminho alternativo: Delimitar o âmbito de discussão dos resultados criminalmente desvalorizados produzidos com a intervenção das I.A.s. Repensar os limites da contribuição subjetiva necessária para a atribuição de responsabilidade penal, atualizando a discussão sobre o elemento subjetivo da imputação e os pressupostos de produção de resultados desconhecidos *ex ante* e, por fim, estruturando a teoria do delito a partir da filosofia da linguagem, demonstrar a inviabilidade das proposições sobre a responsabilidade penal dos próprios I.A. em razão de sua incapacidade de ação.

3.1 Os limites do problema: as classes de I.A

Se o uso de sistemas de I.A. não é propriamente uma novidade³⁹ e há instrumentos de sanção em vigor para isso, então por que alguns casos

³⁷ Nesse sentido (SOUSA, 2020, pp. 79-80).

³⁸ Assim (JANUÁRIO, 2020, pp. 119 e ss.).

³⁹ Nesse sentido (SOUSA, 2020, p. 61).

envolvendo I.A. levantam um nível adicional de preocupação e parecem escapar do raciocínio do sistema de justiça criminal?

Ocorre que existem modelos computacionais de I.A. cuja base algorítmica é organizada para receber *inputs* do ambiente em que se encontra e, a partir disso, produz novas informações, não previsíveis nem programáveis, que são incorporadas ao funcionamento da mesma I.A. A mesma máquina é programada para descobrir padrões a partir dos dados que recebe do ambiente - ou mesmo de outras máquinas, no caso da chamada *internet das coisas* ou *big data* - e, com isso, consegue produzir modelos e regras de funcionamento⁴⁰.

A questão é que, obviamente, o cálculo atuarial do risco e seu limite de permissibilidade - típicos do esquema funcionalista - não podem mais alcançar os pressupostos em que o progresso tecnológico chega ao campo do *aprendizado de máquina*. Máquinas que produzem resultados a partir de intervenções impulsionadas pelo encontro entre um algoritmo aberto e dados externos aleatórios, cuja combinação é o que impulsiona sua operação⁴¹, escapam do campo da análise de risco simplesmente porque não se sabe nem o *sim* nem o *como* do risco futuro envolvido.

⁴⁰ Cf. (JANUÁRIO, 2020, p. 99) e (SOUSA, 2020, pp. 61-62).

⁴¹ Uma ideia melhor desenvolvida do que é o *aprendizado de máquina* é oferecida por Banteka quando ele afirma que «a IA difere de outros softwares automatizados convencionais e apresenta novos desafios para o direito devido à sua capacidade de autoaprendizagem, acumulando experiência pessoal e gerando soluções para problemas com base em uma análise independente de vários cenários sem o envolvimento de um programador. Isso é chamado de *autotreinamento*. Para a I.A., não há regras pré-programadas para a participação humana e a solução de problemas. Em vez disso, as I.A. seguem instruções sobre como aprender com os dados que encontram à medida que operam. Esse é um processo que começou a imitar a experiência humana e que, em termos de I.A., chamamos de *aprendizado de máquina*. O aprendizado de máquina se baseia em ferramentas e processos estatísticos que partem de um banco de dados e de um conjunto de algoritmos que, em seguida, projetam uma regra ou um procedimento para dar sentido a esses dados ou prever dados futuros. Os algoritmos de aprendizado de máquina fazem uso extensivo de ferramentas de inferência estatística para identificar riscos, prever e minimizar erros, atribuir pesos a variáveis e, por fim, otimizar resultados. (BANTEKA, 2021, p. 544). Uma explicação resumida do tópico também aparece em (KAPLAN & HAENLEIN, 2019, p. 17); (SOUSA, 2020, p. 66) e (JANUÁRIO, 2020, p. 99).

Portanto, o primeiro passo é reconhecer que nem todos os resultados criminalmente desvalorizados envolvendo sistemas tecnológicos constituem, de fato, novos desafios para o direito penal, mas apenas os chamados *robôs cognitivos* ou sistemas autônomos cuja intervenção específica não está completamente programada e não pode ser sequer previsível ou mesmo acessível ao programador, o que é conhecido na literatura como *sistema de caixa preta*⁴².

A *lacuna da responsabilidade* hipotética começa a aparecer nos casos em que se dá à própria máquina a possibilidade de evoluir por si mesma, de acordo com sua experiência no ambiente real⁴³. Nesses casos, aparece um aspecto decisivo, que é a incerteza dos resultados, o que constitui um importante problema dogmático, pois se a atribuição de responsabilidade penal depende da contribuição subjetiva, ou seja, da presença de dolo ou negligência, a falta absoluta não só de previsão, mas também de previsibilidade do resultado levaria permanentemente, desde o início, à impunidade.

3.2 Resultados imprevisíveis e o elemento subjetivo

É exatamente aí que reside o problema: o resultado decorrente da intervenção da I.A. não está incluído na previsibilidade humana *ex ante*, o que induz - como se verá a seguir, erroneamente - ao resultado que, em certo sentido, é considerado *próprio*, ou seja, o da própria máquina⁴⁴.

A questão não se limita ao elemento subjetivo. Não se trata simplesmente de dizer que não é possível falar de dolo, porque não há vontade ou conhecimento, ou imprudência, porque não há previsibilidade do que está por vir. É uma questão de saber se a responsabilidade criminal pode ser imposta para desencadear a incerteza dos resultados, porque

⁴² Portanto, consulte (ABBOTT & SARCH, 2019, p. 331). Para obter detalhes sobre o chamado *sistema de caixa preta*, especialmente com relação a aplicações médicas de I.A., consulte (SULLIVAN & SCHWEIKART, 2019, pp. 161-166).

⁴³ Nesse sentido (ABBOTT & SARCH, 2019, pp. 369 e ss.) e (SOUSA, 2020, p. 64).

⁴⁴ Ele se refere a ela, de passagem, (SOUSA, 2020, p. 62).

seria ilógico pretender reconhecer o surgimento da intenção ou um *dever de cuidado em um* momento posterior à ação.

Não é novidade que a doutrina criminal lide com casos de causalidade de risco desconhecido.

Casos semelhantes já foram enfrentados anteriormente, por exemplo, na manipulação genética, gerando fortes reações doutrinárias⁴⁵ e até mesmo legislativas, com a proibição penal com figuras típicas de perigo abstrato⁴⁶, ainda que o direito legal estivesse perfeitamente identificado⁴⁷ e as condutas em si fossem simplesmente orientadas para um futuro incerto⁴⁸, ainda não alcançado pela tecnologia do momento de sua edição. A própria ONU, após um longo debate, somente em 2005 veio a emitir uma Declaração propondo que os Estados-membros proibissem todas as formas de clonagem humana “incompatíveis com a dignidade humana e a proteção da vida humana”, sem esclarecer que tipos de conduta essa proibição deveria abranger⁴⁹.

As perguntas que surgem são: qualquer coisa que produza resultados desvalorizados graves, mas não previsíveis, mas possíveis, deve dar origem à responsabilidade criminal? E, em caso afirmativo, que tipo de

⁴⁵ O exemplo é o XIV Congresso da Associação Internacional de Direito Penal, realizado em Viena, em 1988, no qual foi proposta (resolução 6.9) a criminalização da clonagem de seres humanos. Confira-se (ROMEO CASABONA, 2000, p. 160).

⁴⁶ Exemplos disso são o artigo 160.3 do atual Código Penal espanhol, redigido pela primeira vez em 1995, que prevê a punição da clonagem para fins eugênicos, independentemente de isso ser realmente possível, e a extensa Gesetz zum Schutz von Embryonen (Embryonenschutzgesetz - ESchG) alemã, que já em 1990 estabeleceu uma infinidade de proibições puníveis com prisão para possíveis manipulações genéticas, independentemente dos resultados realmente produzidos a partir delas.

⁴⁷ O trabalho de (ROMEO CASABONA, 2000, pp. 145-168) é seminal sobre as dificuldades de identificar o direito legal em casos de manipulação genética. O debate, que remonta à virada do século, mistura direitos legais individuais fundamentais, como liberdade ou identidade, com direitos legais coletivos, como evitar práticas eugênicas, todos baseados em hipóteses que estão longe de oferecer qualquer tipo de certeza.

⁴⁸ A expressão “comportamento orientado para o futuro” para se referir a esse tipo de situação em que o risco é incerto enquanto se aguarda a evolução tecnológica aparece em (ROMEO CASABONA, 2007, pp. 16-23).

⁴⁹ Essa é a Declaração da Assembleia Geral de 8 de março de 2005, art. 8.b). A questão é discutida em (ROMEO CASABONA, 2006, pp. 79 e segs.). Para detalhes específicos, no mesmo volume, consulte (ISASI & ANNAS, 2006, pp. 13-26).

responsabilidade, quem pode ser responsabilizado e por que tipo de conduta?

As questões se referem exatamente à reconstrução da categoria de crimes de perigo abstrato com base na restrição de sua legitimidade aos casos de conduta perigosa ou casos de “perigo hipotético”, ou seja, aqueles que não alegam um perigo real, mas uma ação capaz de produzir um perigo⁵⁰.

Se - como mencionado acima - há resultados de danos a importantes bens jurídicos produzidos pelo I.A., um campo de regulação criminal está perfeitamente estabelecido com infrações penais nas quais a técnica do perigo abstrato é utilizada, o que é interpretado, neste caso, no sentido referido por Gallas⁵¹, como ‘crimes de perigo possível’.

Com base nessa legitimação, a tarefa passa a ser a de estabelecer em que consiste a conduta perigosa a ser definida nas infrações penais. Como se pode ver, o ponto-chave é a determinação da conduta, de modo que não é surpreendente a miserável contribuição das teorias dogmáticas que desconsideram o problema da ação.

3.3 A abordagem equivocada da doutrina criminal contemporânea

O título deste artigo, “*Das máquinas e dos seres vivos*”, é uma referência deliberada às origens da tese da *autopoiese* de Maturana e Varela⁵² que, por meio de Luhmann⁵³, chegou ao direito penal⁵⁴ por representar a forma mais radical entre as tendências funcionalistas. Conforme observado acima, a dogmática está preocupada em discutir questões normativas, considerando como certa a existência de conduta criminosa. Pelo menos

⁵⁰ A esse respeito, é fundamental a delimitação feita por Torío López com base na interpretação da teoria da norma de Binding (TORÍO LÓPEZ, 1997, pp. 830 e ss.).

⁵¹ Veja (GALLAS, 1972, p. 179).

⁵² (MATURANA ROMESÍN & VARELA GARCÍA, 2003).

⁵³ Veja (LUHMANN, 1972).

⁵⁴ Veja (JAKOBS, 1997, pp. 9-11).

no que diz respeito ao nosso ambiente, isso deriva de uma *abordagem* dominante da teoria do crime ancorada em modelos funcionalistas, que se referem diretamente à discussão da imputação, sem antes levantar questões sobre a delimitação negativa fornecida pelo *requisito do ato*. Deveria ser óbvio que uma abordagem dogmática organizada com base no raciocínio do cálculo de risco e seus limites de permissibilidade (funcionalista) não seria capaz de oferecer soluções coerentes para casos que vão muito além das fórmulas atuariais.

Quando se começa a discutir a ideia de responsabilidade, vai-se direto aos conceitos normativos, sem tentar - *a priori* - separar a conduta de meras ocorrências, o que parece ser uma questão decisiva quando se trata das questões nas quais a I.A. está envolvida.

Deve-se observar que, ao afirmar linguisticamente - no sentido comum - a atribuição de um resultado a algo - em que a expressão responsabilidade é usada em algum sentido -, não estamos necessariamente falando de conduta. Por exemplo: pode-se dizer que a chuva foi responsável pela enchente ou que o sol foi responsável pelo derretimento da neve, sem que isso represente alguma forma de ação. Obviamente, isso não é o mesmo que dizer que uma pessoa foi responsável pela morte de outra⁵⁵. Em todos os exemplos, é possível falar sobre *quem* é o responsável e *qual* foi o resultado, mas somente no último exemplo é possível perguntar *o que* a pessoa responsável fez. Não se pergunta *o que* o sol ou a chuva fizeram, mas *o que* o assassino fez.

Isso enfatiza que o fator decisivo na responsabilidade requer uma discussão prévia da ação, que é proposta a seguir.

3.4 Um passo atrás até a raiz do problema.

Não é possível tentar discutir a questão da atribuição do resultado criminoso sem antes expor um problema fundamental que tem a ver com a preservação de uma garantia básica do direito penal: o problema da ação e do direito penal do ato.

⁵⁵ Sobre essa diferença (LOH & LOH, 2017, p. 37).

A. A ação não tem nada a dizer no direito penal? A. Uma morte preterdolosa.

De fato, desde seu início, a dogmática funcionalista se distanciou deliberadamente da discussão da ação⁵⁶. Sua crítica correta e contundente, que teve como alvo as concepções ontologicamente baseadas da teoria do crime, levou a um beco sem saída diante dos problemas ditados pela inteligência artificial. Isso porque, ao distanciar a discussão sobre a existência de uma ação do plano do *ser*, a busca por um conceito descritivo de ação, condicionando o *dever-ser* jurídico-penal, jogou fora, mais do que um simples conceito ontológico, o próprio conceito de ação, subordinando-o às discussões sobre criminalidade⁵⁷.

Ocorre que, em se tratando de casos envolvendo morte ou lesão envolvendo I.A., antes de se discutir os limites da responsabilidade penal dos envolvidos, uma questão prévia deve ser esclarecida: a responsabilidade eventualmente existente deve decorrer de *algo* que *foi feito* ou *omitido*, o que leva à discussão sobre a capacidade de ação e omissão, mais especificamente, a capacidade de ser o autor do fato. Em suma, é necessário saber *quem fez o quê* antes de procurar saber se o resultado está dentro ou fora do risco permitido, dentro ou fora da esfera de organização do perpetrador, se o resultado é ou não justificado ou exculpado, quais são os limites do dever e como a responsabilidade é distribuí-

⁵⁶ O ponto de referência sobre esse assunto é o conhecido artigo crítico de Roxin dirigido à tese de Welzel (ROXIN, 1962, p. 515-561). A versão em português foi utilizada no texto (ROXIN, 1998, p. 91-143). O autor afirma que “Um conceito ontológico de ação final pleno de conteúdo e ‘independente de qualquer relação com o direito’, como constituindo a base do sistema de direito penal, deve ser renunciado, porque tal conceito não existe e não pode existir” (p. 126) e conclui: “[...] é melhor inverter a relação entre o conceito de ação final e o conceito de direito penal” (p. 126). O melhor é inverter a relação entre ação, finalidade e criminalidade: não é correto afirmar que uma estrutura pré-jurídica e vinculante da ação para o legislador determina o conteúdo da finalidade e da criminalidade, mas, ao contrário, o tipo nasce de considerações valorativas jurídicas e está orientado para os conteúdos sociais de sentido que determinam quais são as condutas finais em sentido jurídico e quais são os fatores a serem incluídos no tipo subjetivo. Nesse sentido, a teoria final *da ação* deve ser substituída por uma teoria final *do delito*” (p. 127).

⁵⁷ Embora bastante clara em toda a dogmática funcionalista, essa subordinação encontra forma expressa em (ROXIN, 2000, p. 218): “Uma pessoa agiu se certos efeitos, originados ou não por ela, podem ser atribuídos a ela como pessoa”.

da. Se partirmos de um direito penal do ato, primeiro é necessário saber *quem fez (ou deixou de fazer) o quê*, ou seja, é importante discutir o que é conduta e o que é um simples ato, quais são os limites que separam a ação e a ocorrência e que permitem que um autor seja reconhecido como o centro passivo da imputação penal, afinal, a responsabilidade penal é basicamente e acima de tudo uma forma de responsabilização por algo que foi feito⁵⁸. Somente depois que isso for estabelecido é possível discutir qualquer tipo de responsabilidade e até que ponto o status do perpetrador pode ser ampliado.

B. Uma garantia inquestionável: a função negativa do conceito de ação no direito penal.

Parece haver um consenso entre os penalistas - pelo menos em um ambiente acadêmico supostamente democrático - de que um direito penal do ato deve ser preservado, em oposição a um direito penal do autor⁵⁹; assim, a ideia de ação continua a cumprir uma função delimitadora no direito penal, expressa pelo *requisito do ato* (requisito da ação). Em outras palavras, só é permitido impor responsabilidade criminal por ações ou omissões.

Portanto, para sustentar um direito penal minimamente democrático, é necessário separar ações e omissões, de um lado, de simples ocorrências, de outro. Separar o que se faz do que simplesmente acontece parece ser tão básico quanto essencial para que se possa discutir a atribuição de responsabilidade criminal⁶⁰.

⁶¹Por outro lado, não se deve esquecer que a polêmica sobre a I.A. convida a revisar os pontos de vista sobre o escopo da consciência moral, por meio da qual o interesse filosófico tende a levar naturalmente ao campo da filosofia da mente e, portanto, de volta ao beco sem saída

⁵⁸ Há referências semelhantes (*para prestar contas de suas ações*) especialmente na literatura anglo-saxônica, por causa de seu conhecido pragmatismo. Veja, como exemplos, (DUFE, 1998, p. 290); (LOH & LOH, 2017, p. 37).

⁵⁹ Até o próprio Roxin, principal crítico do conceito de ação na dogmática, alerta para a importância dessa delimitação em (ROXIN, 2000, p. 176 e ss.).

⁶⁰ É isso que Vives Antón propõe como ponto de partida em (2011, pp. 221 e segs.).

⁶¹ Veja, sobre as conexões entre a I.A. e a filosofia, (DENNET, 1993, pp. 320-334).

dos conceitos de ação como movimento físico impulsionado pelo pensamento, com todas as oposições que podem ser apresentadas a ele, desde inconsistências filosóficas até inadequações dogmáticas.

Com relação à primeira, basta citar a incapacidade do Direito de lidar com uma ideia de ação que decorre da concepção cartesiana - ou de suas derivações - da mente⁶².

Com relação a este último, é necessário estar ciente do erro de buscar um retorno às concepções de ação baseadas em dados ontológico-empíricos⁶³, ou mesmo identificá-lo em associação com critérios de personalidade⁶⁴.

Assim, se por um lado é verdade que é necessário dar um passo atrás na discussão dogmática, reenquadrando o problema em termos de ação, por outro lado, é importante não voltar ao passado de concepções de ação há muito superadas, nem ceder à artificialidade de trocar a discussão da ação por uma discussão de normas. Por essa razão, propõe-se aqui entrar nas bases teóricas da filosofia da linguagem com base no fato de que a diferença entre ações e fatos “é que atribuímos significados às primeiras e não aos segundos”⁶⁵.

3.5 Repensando o problema

Se é impossível discutir a responsabilidade penal sem antes pensar em *quem fez o quê*, e o *fazer* ou *omitir* que consiste em uma conduta se ba-

⁶² Veja, em detalhes, a crítica de Vives Antón em (2011, pp. 161-200).

⁶³ A deficiência do modelo dogmático ancorado no conceito ontológico de ação tem sido amplamente debatida. Cf. (MARINUCCL, 1998, especialmente pp. 136 e ss.) e (BUSATO, 2007, especialmente pp. 87 e ss.). A doutrina mais recente - apesar de se afirmar livre do roteiro ontológico - continua a utilizar uma linguagem a respeito da ação que a identifica com o movimento físico de natureza newtoniana. O próprio Roxin admite isso quando afirma que “O conceito de manifestação da personalidade [...] está ligado ao modo pré-jurídico de entender a ação, sem se distanciar da concepção ordinária da vida de forma naturalista (como ‘movimento muscular’) [...]” (ROXIN, 2000, p. 256). (ROXIN, 2000, p. 256).

⁶⁴ Sobre as dificuldades de identificar a responsabilidade em casos de A.I. baseados em personalidade, consulte (WHITE & BAUM, 2020, pp. 70-7) e, em detalhes, (HUBBARD, 2011, 405-441).

⁶⁵ (VIVES ANTÓN, 2011, p. 222).

seia em um significado, isso implica, por um lado, pensar no que significa a ação (*o quê*) e, em seguida, com base nisso, identificar quem é o autor (*o quem*) para descobrir se há ou não um espaço de responsabilidade penal no qual o mesmo I.A. pode ser incluído.

3.5.1 Sobre o quê. Ação como uma expressão de significado

Se a identificação do que tem sido feito é uma pesquisa sobre ação, é importante, em primeiro lugar, remover precisamente os defeitos que levaram ao abandono da discussão sobre ação⁶⁶, que foram precisamente três: a *sobrecarga de funções dogmáticas*, a *pretensão conceitual* e a *associação com o insider*.

A. A sobrecarga de funções dogmáticas.

No que diz respeito à *sobrecarga de funções dogmáticas*, é importante enfatizar que não se pode exigir mais da ideia de ação do que ela pode oferecer. Não se pode esperar que a ação funcione como um conceito coordenador, definidor e classificador⁶⁷, como tem sido a intenção nas concepções ontologicamente baseadas da teoria do crime. Essa sobrecarga tem consistido em exigir uma existência ontológica e pré-jurídica da ação, concomitante a uma capacidade de absorver os qualificadores normativos derivados da teoria do delito e, inclusive, de ser capaz de abarcar tanto as formas positivas (comissão) quanto as negativas (omissão) de manifestação, além das formas subjetivas de orientação intencional e negligente.

O fracasso retumbante de tais alegações é bem conhecido⁶⁸. Assim, o que se deve ter em mente é que a capacidade de atuação da ideia de ação no direito penal se reduz a possibilitar a separação dos casos em que há ação daqueles em que ela não está presente, por se tratar de simples ocorrências. A questão é tão básica quanto decisiva⁶⁹. Embora possa

⁶⁶ O labirinto mencionado por Vives em (2011, p. 219).

⁶⁷ Sobre o assunto, veja comentários críticos em (VIVES ANTÓN, 2011, p. 123) e (BUSATO, 2007, pp. 71-78).

⁶⁸ (BUSATO, 2007, pp. 125-133).

⁶⁹ Cf. (VIVES ANTÓN, 2011, p. 220).

parecer irrelevante, essa é a expressão de um direito penal do ato, uma garantia fundamental que não pode ser simplesmente abandonada.

B. A alegação conceitual.

Mas dizer que a identificação de ações em oposição a eventos é necessária não significa exigir a presença de um *conceito* de ação, longe disso. A tarefa de identificação mencionada acima não depende de uma *afirmação conceitual*.

Deve-se enfatizar que não há nada que possa, com relação a infrações penais, delimitar o *ser* da ação a ponto de permitir que ela seja contida em um conceito. Isso ocorre simplesmente porque não há nada em comum que possa unificar todas as ações⁷⁰. Enquanto os fatos podem ser descritos de acordo com regras naturais, as ações são interpretadas de acordo com *contextos de uso*, ou seja, são reconhecidas como tal em um determinado contexto.

No entanto, tanto a tentativa de capturar a essência da ação por um método substantivo - ou seja, defini-la como algo no mundo - ou por um método atributivo - concebê-la como o substrato para a imputação - resulta em fracassos igualmente retumbantes, simplesmente porque “a ação não é, portanto, um fato *específico*; nem pode ser *definida* como o *substrato para a imputação* jurídico-penal”⁷¹ mas as ações consistem simplesmente em “interpretações que, de acordo com diferentes tipos de regras sociais, podem ser dadas ao comportamento humano”⁷².

Assim, é possível separar ações e fatos com base no fato de que, enquanto os fatos simplesmente acontecem, as ações *fazem sentido*.

Esse *sentido* ou *significado* também é identificado não como algo no mundo, mas como o *uso* da própria linguagem para significar a ação, na

⁷⁰ Na verdade, como Wittgenstein se refere à ação, bem como à linguagem, com a qual ela está entrelaçada, “não há nada de comum nesses fenômenos para os quais usamos a mesma palavra para todos - mas eles estão *relacionados* uns aos outros de muitas maneiras diferentes. E por causa desse parentesco, ou desses parentescos, chamamos todos eles de ‘linguagem’”. (WITTGENSTEIN, 2002, p. 87).

⁷¹ (VIVES ANTÓN, 2011, pp. 221).

⁷² (VIVES ANTÓN, 2011, p. 221).

forma de um jogo de linguagem⁷³. Identificamos algo como tendo a *aparência de ação* quando, de acordo com nosso modo de vida, algo expressa um significado que pode ser identificado, a partir de nossas práticas sociais, como uma ação. Assim, por exemplo, se observo alguém passando na rua, penso que essa pessoa está agindo, mesmo que eu não saiba dizer se isso significa ir a um compromisso, fugir de uma ameaça ou até mesmo ser sonâmbulo⁷⁴. O que se consegue, em princípio, é identificar algo que tenha a *aparência de ação*, porque, se for confirmado que se trata de sonambulismo, a própria existência da ação se torna discutível.

Por outro lado, se for uma ação, em uma segunda etapa é possível identificar que tipo de ação é, dependendo se ela pode ser identificada como a expressão de uma intenção específica⁷⁵.

C. O abandono do preso.

Mas essa *intenção* não é algo que está dentro do sujeito, nem é algo que se refere à mente. É precisamente nesse ponto que os penalistas mais atentos à evolução do sistema de imputação compreenderam o enorme potencial de solução de problemas de uma teoria do crime ancorada na filosofia da linguagem⁷⁶.

É importante entender que «as regras *que regem* o significado das ações [...] servem [...] para explicar a ação, avaliá-la, justificá-la e *defini-la* e interpretá-la”⁷⁷. Mas as regras só podem ser seguidas ou contrariadas

⁷³ Consulte (VIVES ANTÓN, 2011, pp. 224-231).

⁷⁴ O exemplo é de Vives Antón (2011, pp. 228-229).

⁷⁵ Como esclarece Vives Antón, “[...] a atribuição de intenções ao sujeito, ou a classificação de sua conduta como não intencional, não desempenha necessariamente um papel na delimitação conceitual da ação, mas muitas vezes se limita a determinar se o autor incorre ou não em mérito ou demérito, se é responsável pelo que fez e em que medida é responsável por isso. O papel *definidor* (conceitual) é, portanto, em princípio, desempenhado pela *intencionalidade objetiva*. A *intenção subjetiva* (a atribuição concreta de intenções ao sujeito), por outro lado, tem a função *substantiva* de tornar possível o julgamento da conduta”. (2011, p. 239).

⁷⁶ Como Muñoz Conde reconheceu em seu prólogo à obra seminal de Vives Antón, “é no campo da determinação do conceito de ‘intenção’, que no direito penal pode ser entendido como dolo, que a teoria da ação significativa proposta por Tomás Vives tem a maior relevância” (Muñoz Conde, 2011, p. 32). (MUÑOZ CONDE, 2011, p. 32).

⁷⁷ (VIVES ANTÓN, 2011, p. 229).

pela existência de um *uso* estabelecido que obviamente está *fora do assunto*. Assim, o *tipo de ação em questão* é identificado de acordo com as regras que são socialmente relevantes para sua identificação de acordo com o uso social.

Portanto, embora seja verdade que, para falar de ações, os sujeitos que agem devem formar *intenções*, as ações que eles realizam não dependem das intenções que formam internamente (na mente), mas do significado atribuído a suas ações.

Assim, o significado de uma ação não é projetado de dentro para fora, mas o contrário⁷⁸. O propósito interno, em primeira pessoa, subjetivo - que não se nega que exista - é insuficiente para a determinação do significado social intersubjetivo das ações⁷⁹ e, como tal, é uma maneira pouco promissora de resolver o problema da ação. A simples produção de resultados indesejados é uma ampla prova disso: no final, é perfeitamente possível e até mesmo comum ofender, desapontar, enganar, sem nunca ter a intenção de fazê-lo⁸⁰.

A intenção é simplesmente “incorporada à ação”⁸¹ de acordo com as regras e práticas da própria linguagem. Assim, pode-se dizer, com Vives Antón, que há “uma ‘*intencionalidade*’ externa e *objetiva*, uma prática social constituinte de significado [...] na qual as *intenções do sujeito* se baseiam, e sem a qual elas nem mesmo são identificáveis como intenções”⁸².

A ação, como expressão do significado de uma intencionalidade, representa apenas que o agente é capaz de comunicar uma intenção autônoma, independentemente do que acontece com a dimensão interna do que se passa em sua mente. Ou seja: o papel definidor de uma ação como tal depende da *intencionalidade objetiva* e não de uma *intenção subjetiva*, e é isso que pode servir de filtro interpretativo para entender se é possível falar de ações com a mesma I.A.

⁷⁸ (VIVES ANTÓN, 2011, p. 219).

⁷⁹ Da mesma forma (VIVES ANTÓN, 2011, p. 219).

⁸⁰ Cf. (VIVES ANTÓN, 2011, p. 231).

⁸¹ (VIVES ANTÓN, 2011, p. 234).

⁸² (VIVES ANTÓN, 2011, p. 234).

3.5.2 Em quem

Se a ação é a expressão do significado de uma intenção, então surge a questão de quem é capaz de expressar intenções por meio de ações ou omissões. Se for verdade que a única coisa que o direito penal pode fazer é exercer controle social sobre condutas intoleráveis⁸³, ele encontra seus limites de aplicação *ratione personae* com relação àqueles que são capazes de realizar ações ou omissões proibidas pelo direito substantivo.

Surge então a questão de saber se o I.A. pode ser um sujeito passivo de intervenção criminal, ou seja, se é capaz de fazer algo que possa ter um *sentido de ação*, algo que possa *expressar objetivamente intenções*.

Devemos começar por ter em mente que o movimento físico não é suficiente para determinar a existência da ação. Tampouco é suficiente que o verbo correspondente a determinada ação esteja expresso no tipo penal para superar a barreira da lei penal do fato.

Devemos nos perguntar se podemos falar de uma intencionalidade objetiva, uma expressão de significado da linguagem humana no que uma máquina “faz” (entre aspas).

Quando uma pessoa age, percebemos, no próprio *desempenho* da ação, uma *intenção objetivamente expressa* que, de acordo com regras sociais consistentes com o uso comum da linguagem, pode traduzir a *intenção subjetiva*. Quando a interação é com um robô ou uma máquina, o uso comum da linguagem não procura na suposta “conduta” qualquer *intenção objetivamente expressa*, mas simplesmente o resultado - mais ou menos complexo - de uma programação. Em suma, como refere Susana Aires de Sousa, “a máquina decide um resultado objetivamente, submetendo a informação ao algoritmo, segundo um determinado modelo, sem que haja qualquer ‘intenção’ ou ‘descuido’ materializado num resultado juridicamente desvalorizado”⁸⁴.

A questão não tem nada a ver com a capacidade cognitiva ou psíquica ou com o fato de os I.A. poderem ou não expressar sua vontade,

⁸³ Nesse sentido (BUSATO, 2020, pp. 623-628).

⁸⁴ (SOUSA, 2020, p. 69).

mas sim com o entendimento de que sua suposta “vontade” não é, no final das contas, algo como uma expressão significativa de uma intenção humana. Isso é o que está presente tanto na pessoa física quanto na pessoa jurídica (como resultado de intervenções exclusivamente humanas)⁸⁵ e não é encontrado nem em animais (devido a seu modo de vida diferente⁸⁶) nem em robôs (que não são capazes de expressar o significado de uma intenção, mas apenas uma derivação pré-programada).⁸⁷

Não faz sentido comum afirmar que o *software* ou o *sistema operacional* - de um carro autônomo, por exemplo - é capaz de expressar intenções⁸⁸, independentemente da quantidade de cálculos logarítmicos que ele consiga processar. O que acontece é que o programa simplesmente deflagra comandos eletrônicos que são obrigados a seguir o que foi progra-

⁸⁵ Não é o objetivo aqui entrar na longa discussão sobre os fundamentos da responsabilidade criminal das pessoas jurídicas, mas simplesmente enfatizar que a questão de sua inclusão no direito penal é bem diferente da questão dos casos que envolvem decisões decorrentes do P.I. Não é possível reconhecer no P.I. uma manifestação de uma «forma de vida humana», como podemos fazer no caso das pessoas jurídicas. A pessoa jurídica é algo que se desenvolve como uma forma de vida humana, pois não pode existir ou ser criada a não ser no ambiente jurídico, portanto, ela “vive” ou “existe” apenas dentro da esfera jurídica, que é nada menos que uma forma de linguagem humana. Assim, a existência de pessoas jurídicas - e de sua própria capacidade de ação - ocorre dentro do ambiente jurídico, ou seja, dentro de algo peculiar à forma humana de vida. O ambiente operacional da A. I. não é um “ambiente” exclusivamente humano, porque em sua existência e funcionalidade ele pode operar completamente à parte do ser humano. Portanto, não faz sentido equiparar essas duas questões, que são completamente diferentes.

⁸⁶ Como disse Wittgenstein: “Mesmo que um leão pudesse falar, nós não o entenderíamos”. Cf. (Wittgenstein, 1971, p. 260). No original: „Wenn ein Löwe sprechen könnte, wir könnten ihn nicht verstehen“.

⁸⁷ Não é desconhecido no campo da neurociência o argumento de que os seres humanos são programados por impulsos elétricos predeterminados (ver ROTH, 2003, p. 251); (SINGER, 2003, pp. 12, 20, 42, 58 et seq.). Se esse ponto de vista for aceito - com o qual não se concorda em absoluto - (manifestei-me especificamente contra ele em (BUSATO, 2018, pp. 149-184), seria possível equipará-los à predeterminação da I.A. e, conseqüentemente, admitir sua responsabilidade penal direta.

⁸⁸ As pesquisas sobre o desenvolvimento da I.A. passaram a compor sistemas capazes de emular o comportamento humano, escolhendo e interrompendo tarefas de acordo com a urgência, aprendendo posturas mais agressivas ou mais sutis, até mesmo cometendo erros deliberados, mas sempre sob uma espécie de mimetismo do comportamento humano e não de acordo com qualquer tipo de emoção. Sobre o desenvolvimento atual desse tipo de imitação, consulte (DAUGHERTY & WILSON, 2018, pp. 113-133).

mado antes⁸⁹. Enquanto afirmamos, por exemplo, a responsabilidade de um motorista de carro porque dizemos que ele *dirigiu seu carro com pressa* (imprudentemente) ou afirmamos que o assassino atirou em alguém por *vingança* (maliciosamente), não creio que possamos dizer que o sistema de um carro autônomo expressou sua intenção de chegar mais cedo ao seu destino ou que o algoritmo de um drone armado queria se vingar.

É irrelevante procurar a *expressão objetiva das intenções* em qualquer coisa que não sejam as próprias ações ou omissões, ou seja, nas realizações como elas são apresentadas.

Não há dúvida de que a I.A. pode determinar o movimento do carro e que pode disparar uma arma, em ambos os casos obtendo o resultado de *matar*. Mas, da mesma forma, não há dúvida de que é impossível dizer que a I.A. *queria* vingança ou estava com *pressa*, simplesmente porque é impossível para a inteligência artificial expressar objetivamente as intenções.

De fato, a ação é concebida “não *como um substrato* comportamental capaz de receber um significado, mas *como um significado* que, de acordo com um sistema de regras, pode ser atribuído a determinados comportamentos humanos”⁹⁰, que, no máximo, uma máquina pode emular, mas nunca expressar.

E se o que a I.A. faz é mesmo no sentido de ação, não faria sentido discutir a responsabilidade penal além dos limites da conduta do usuário da máquina e de seus programadores no âmbito dos crimes de perigo

⁸⁹ O mesmo Hallevy, que defende a composição de uma responsabilidade penal própria para os I.A., trai seu próprio raciocínio quando aponta como inadequado o chamado *modelo de responsabilidade perpetration-via-another*, sob o argumento de que os I.A. decidem cometer um crime com base em sua própria experiência e conhecimento acumulado. Cf. (HALLEVY, 2016, p. 181). É importante perceber que se trata de um condicionamento determinado pelo programa, pois, em nenhuma hipótese, as máquinas comandadas por I.A. - diferentemente dos humanos - podem optar por agir contra o **que sua experiência e conhecimento acumulados recomendam**, simplesmente por diversão ou risco. Isso mostra como, mesmo que se possa admitir o uso da palavra *decisão para fins* de produção de resultados por uma I.A., não se pode de forma alguma dizer que, com essa produção de resultados, ela tenha expressado uma intenção.

⁹⁰ (VIVES ANTÓN, 2011, p. 221).

abstrato ligados ao princípio da precaução⁹¹, dentro dos limites da responsabilidade já conhecida e que pode ser perfeitamente estabelecida, independentemente de os resultados produzidos não terem sido previamente desejados ou representados por qualquer pessoa.

Feito o corte, não é objetivo deste trabalho descrever detalhadamente a distribuição da responsabilidade criminal entre as pessoas que gerenciam o I.A.; para isso, seria necessário outro estudo. Ainda assim, não é demais apontar, em linhas gerais, as razões pelas quais se entende que, reconhecida a incapacidade de atuação do I.A., todos os casos de resultados criminalmente inválidos são cobertos pela responsabilidade das pessoas, sem subtrair qualquer *lacuna de responsabilidade*.

De fato, deve-se presumir que o *software* que controla o movimento de um dispositivo eletrônico foi programado para manter uma rotina de normalidade e evitar opções arriscadas dentro do previsível. Se, em tais casos, os resultados forem prejudiciais aos bens jurídicos ou se forem devidos a erros de projeto, programação, construção, uso de material inadequado, etc., nesse caso a responsabilidade é estabelecida de acordo com as regras de responsabilidade pelo produto⁹².

Se, por outro lado, os resultados forem decorrentes de um uso indevido do dispositivo, devido a um desvio de cuidado em seu uso, a responsabilidade pelo resultado deve pertencer ao uso indevido ou arriscado, pois é a pessoa e não o dispositivo que age⁹³.

⁹¹ Veja, resumidamente, em (SOUSA, 2020, pp. 85-86 e 89). Para obter detalhes sobre o princípio da precaução, consulte (ROMEO CASABONA, 2002, pp. 36 e seguintes).

⁹² Seja por defeitos de fabricação, de projeto, seja pelo devido cuidado na elaboração do próprio programa de controle. Nesse sentido, embora limitado nas hipóteses, (JANUÁRIO, 2020, p. 120). Sobre responsabilidade penal do produto, ver (HASSEMER & MUÑOZ CONDE, 1995.) No ambiente anglo-saxônico, ver (OWEN & DAVIES, 2016). Resumido em (SOUSA, 2020, pp. 80-83). e resumido em (Gurney, 2020, principalmente pp. 53-56).

⁹³ Por exemplo: quando ligamos para outra pessoa no WhatsApp para falar com ela, dizemos que fomos nós e não o telefone que ligou, embora seja indubitável que alguém o tenha programado de tal forma que, com um simples toque na tela, uma pessoa específica com quem queremos falar receba um alerta de chamada e não saibamos quem e como ela foi feita.

Os resultados da opção aqui proposta são equivalentes às fórmulas do *modelo de responsabilidade perpetração-via-anterior* apresentado como uma das possíveis soluções para o problema por Hallevy⁹⁴. Entretanto, o conteúdo e a base dos argumentos são completamente diferentes dessa proposição, pois não se trata de considerar os I.A. como pessoas incapazes de culpa, mas sim de reconhecê-los como incapazes de ação.

REFERÊNCIAS

- ABBOTT, Ryan e SARCH, Alex. "Punishing Artificial Intelligence: Legal Fiction or Science Fiction". **UC Davis Law Review** - Vol. 53. Davis: University of California, 2019.
- BANTEKA, Nadia. "Artificially Intelligent Persons". **Houston Law Review** **58**, Houston: Universidade de Houston, 2021.
- BARBOSA, Mafalda Miranda. "Nas fronteiras de um admirável mundo novo? O problema da personificação de entes dotados de inteligência artificial". In: BARBOSA; Mafalda Miranda, BRAGA NETO, Felipe; SILVA, Michel César; FALEIROS JÚNIOR, José Luiz de Moura (Coords.), **Direito digital e inteligência artificial. Diálogos entre Brasil e Europa**. São Paulo: Foco, 2021.
- BECK, Ulrich. **Sociedade de risco. Rumo a outra modernidade**. Trans. de Sebastião Nascimento, São Paulo: Editora 34, 2011.
- BECK, Ulrich. **A metamorfose do mundo. Novos conceitos para uma nova realidade**. Trans. de Maria Luíza X. de A. Borges, Rio de Janeiro: Zahar, 2018.
- BOSTROM, Nick. **Superintelligence [Superinteligência]**. Oxford: oxford University Press, 2014.
- BRANCO, Alexandra. **Para Além do QI: Uma Perspetiva Mais Ampla de Inteligência**. Coimbra: Quarteto, 2004.
- BUSATO, Paulo César. "Liberdade de ação versus Neurociências". In: **A linguagem do Direito penal**. São Paulo: Tirant lo Blanch, 2018.
- BUSATO, Paulo César. **Derecho penal y acción significativa**. Valencia: Tirant lo Blanch, 2007.

⁹⁴ Veja (HALLEVY, 2016, pp. 179-181), que aponta que tal modelo de responsabilidade não reconhece a "capacidade mental" em A.I. Em nossa opinião, o problema não reside em qualquer tipo de requisito mental, mas na capacidade de expressar linguisticamente intenções em ações.

BUSATO, Paulo César. **Direito penal. Parte Geral**. 5ª ed., São Paulo: GEN-Atlas, 2020.

BUSATO, Paulo César. **Três teses sobre a responsabilidade penal das pessoas jurídicas**. Valência: Tirant lo Blanch, 2019.

COCA VILA, Ivó. “Coche autopilotados en situaciones de necesidad. Una aproximación desde la teoría de la justificación penal”. *In*: ESTELLITA, Heloísa; LEITE, Alaor (Org.). **Veículos autônomos e Direito penal**. . São Paulo: Marcial Pons, 2019.

DAUGHERTY, Paul R. e WILSON, H. James. **Human + Machine. Reimagining Work in the Age of AI [Reimaginando o trabalho na era da IA]**. Boston: Harvard Business Review Press, 2018.

DENNET, Daniel C. “Cuando los filósofos se encuentran con la inteligencia artificial”, *em*: GRAUBARD, Stephen R. (Comp.) **The New Debate on Artificial Intelligence**. . Madri: Gedisa, 1993.

DUFF, R. A. “Responsibility.” *in*: **Routledge Encyclopedia of Philosophy**, editado por Edward Craig. Londres: Routledge, 1998.

ESTELLITA, Heloísa e LEITE, Alaor. “Veículos Autônomos e Direito penal: uma introdução”. *In*: ESTELLITA , Heloísa; LEITE, Alaor (Org.). **Veículos autônomos e Direito penal**. São Paulo: Marcial Pons, 2019.

GALLAS, Wilhelm. „Abstrakte und konkrete Gefährdung“. *In*: LÜTTGER, Hans; BLEI, Hermann; HANAU, Peter (org.). **Festschrift für Ernst Heintz zum 70 Geburtstag**. Berlin: Walter de Gruyter, 1972.

GIUFRIDA, Iria. “Liability for AI Decision-Making: Some Legal and Ethical Considerations”, *em* **Fordham Law Review**, Volume 88, Edição 2, 2019.

GLESS, Sabine e WEIGEND, Thomas. “Agentes inteligentes e o Direito penal”. *In*: ESTELLITA , Heloísa; LEITE, Alaor (Org.). **Veículos autônomos e Direito penal**. São Paulo: Marcial Pons, 2019.

GOLEMAN, Daniel. **Emotional Intelligence (Inteligência emocional)**. Nova York: Bantam Books, 1995.

GRECO, Luís. “Veículos autônomos e situações de colisão”. *In*: ESTELLITA , Heloísa; LEITE, Alaor (Org.). **Veículos autônomos e Direito penal**. São Paulo: Marcial Pons, 2019.

GURNEY, Jeffrey K.. “Imputing driverhood: applying a reasonable driver standard to accidents caused by autonomous vehicles”. *In*: LIN, Patrick, JENKINS,

Ryan; ABNEY, Keith (Ed.). *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence*. Nova York: Oxford University Press, 2020.

HALLEVY, Gabriel. “The Criminal Liability of Artificial Intelligence Entities - from Science Fiction to Legal Social Control”, in **Akron Intellectual Property Journal** - March 2016, Akron: The University of Akron, 2016.

HASSEMER, Winfried e MUÑOZ CONDE, Francisco. **La responsabilidad por el producto en Derecho penal**. Valencia: Tirant lo Blanch, 1995.

HILGENDORF, Eric. “Direito e máquinas autônomas. Um esboço do problema”. In: ESTELLITA, Heloísa; LEITE, Alaor (Org.). **Veículos autônomos e Direito penal**. São Paulo: Marcial Pons, 2019.

HILGENDORF, Eric. “Sistemas autônomos, inteligência artificial e robótica: uma orientação a partir da perspectiva jurídico-penal”, in **Digitalização e Direito**. Trans. de Orlandino Gleizer, São Paulo: Marcial Pons, 2020.

HÖRNLE, Tatjana e WOHLERS, Wolfgang. “Trolley Problem revisitado: como devem ser programados os veículos autônomos no dilema vida-contravida?”. In: ESTELLITA, Heloísa; LEITE, Alaor (Org.). **Veículos autônomos e Direito penal**. São Paulo: Marcial Pons, 2019.

HUBBARD, E. Patrick. “‘Do Androids Dream?’ Personhood and Intelligent Artifacts”. **Temple Law Review** 83, Filadélfia: Temple University Beasley School of Law, 2011.

ISASI, Rosario M. e ANNAS, George J. “To clone alone: the United Nations Human Cloning Declaration”. **Revista de derecho y genoma humano**, nº 24, Biscaia: Universidad del País Vasco, 2006.

JAKOBS, Günther. **Derecho penal. Parte Geral. Fundamentos y teoría de la imputación**. Trad. de Joaquin Cuello Contreras e Jose Luis Serrano González de Murillo, 2ª ed., Madri: Marcial Pons, 1997.

JANUÁRIO, Túlio Xavier. “Veículos autônomos e imputação de responsabilidades por acidentes” in: **A inteligência artificial no Direito penal**. Lisboa: Almedina, 2020.

KAPLAN, Andreas e HAENLEIN, Michael. “Siri, Siri, em minha mão: Quem é o mais justo da terra? On the interpretations, illustrations, and implications of artificial intelligence”, em **Business Horizons** 62, no. 1. Indiana: Kelley School of Business-Elsevier, jan./fev. 2019.

KENDEL, Eric R.; SCHWARTS, James H.; JESSEL, Tomas M. **Fundamentos da Neurociência e do Comportamento**. Rio de Janeiro: Guanabara Koogan, 1997.

LOH, Wulf e LOH, Janina. “Autonomia e responsabilidade em sistemas híbridos: O exemplo dos carros autônomos”, em **Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence (De carros autônomos à inteligência artificial)**. Nova York: Oxford University Press, 2017.

ONA, Niklas. **Rechtssoziologie**. Volumes I e II, Hamburgo: Rowohlt, 1972.

MARINUCCI, Giorgio. **El delito como acción. Crítica de un dogma**. Tradução de José Eduardo Sainz-Cantero Caparrós, Madri: Marcial Pons, 1998.

MATURANA ROMESÍN, Humberto e VARELA GARCÍA, Francisco. **De máquinas e seres vivos. Autopoiesis: la organización de lo vivo**. 6ª ed. Buenos Aires: Lumen, 2003.

MUÑOZ CONDE, Francisco. “Prólogo” em **Fundamentos del Sistema Penal**. 2ª ed., Valencia: Tirant lo Blanch, 2011.

OWEN, David G. e DAVIES, Mary J.. **Owen & Davis on Products Liability**. 4ª ed., Eagen: Thomson West, 2016.

ROMEO CASABONA, Carlos María. “Bienes Jurídicos implicados en la clonación”. **Revista de Derecho, Criminología y Ciencias Penales** nº 2, Santiago: Universidad San Sebastián, 2000.

ROMEO CASABONA, Carlos María. “Contribuições do princípio de precaução ao Direito penal”. **Revista de Estudos Criminais** nº 5. Porto Alegre: ITEC, 2002.

ROMEO CASABONA, Carlos María. “Hacia un Derecho transcultural para la genética y la biotecnología humanas”. **Anuario Jurídico de La Rioja** nº 12, La Rioja: Universidad de la Rioja, 2007.

ROMEO CASABONA, Carlos María. “La cuestión jurídica de la obtención de células troncales embrionarias humanas con fines de investigación biomédica. Consideraciones de política legislativa”. **Revista de derecho y genoma humano**, nº 24, Biscaia: Universidad del País Vasco, 2006.

ROTH, Gerhard. **Fühlen, Denken, Handeln. Wir das Gehirn unser Verhalten steuert**. Frankfurt: Suhrkamp, 2003.

ROXIN, Claus. “Contribuição para a crítica da teoria finalista da ação”, in: **Problemas fundamentais de Direito penal**. Trad. de Ana Paula dos Santos Luís Natscheradetz, Lisboa: Vega, 1998.

ROXIN, Claus. „Zur Kritik der finaler Handlungslehre“, em **Zeitschrift für die gesamte Strafrechtswissenschaft**, volume 74, número 4. Berlim: De Gruyter, 1962.

ROXIN, Claus. **Derecho penal. Parte General. Tomo I. Fundamentos. La estructura de la teoría del delito.** Trad. de Diego Manuel Luzón Peña, Miguel Díaz y García Conlledo e Javier de Vicente Remesal, Madri: Civitas, 2000, p. 218.

ROXIN, Claus. **Derecho penal. Parte General. Tomo I. Fundamentos. La estructura de la teoría del delito.** Trad. de Diego Manuel Luzón Peña, Miguel Díaz y García Conlledo e Javier de Vicente Remesal, Madri: Civitas, 2000.

SINGER, Wolf. Ein neues Menschenbild? Gespräche über Hirnforschung. Frankfurt: Suhrkamp, 2003.

SOUSA, Susana Aires de. „Não fui eu, foi a máquina: teoria do crime, responsabilidade e inteligência artificial“ in: **A inteligência artificial no Direito penal.** Lisboa: Almedina, 2020.

SULLIVAN, Hannah R. e SCHWEIKART, Scott J.. “As atuais doutrinas de responsabilidade civil são adequadas para lidar com os danos causados pela IA?” **AMA Journal of Ethics**, Volume 21, Número 2, fevereiro de 2019.

TORÍO LÓPEZ, Ángel. “Los delitos de peligro hipotético (contribución al estudio diferencial de los delitos de peligro abstracto). in **Anuario de Derecho penal y ciencias penales**, Madrid: CGPJ, 1997.

TURNER, Jacob. **Robot Rules. Regulating Artificial Intelligence [Regulando a Inteligência Artificial]**. Londres: Palgrave-McMillan, 2019.

VIVES ANTÓN, Tomás S. **Fundamentos del Sistema Penal.** 2ª ed., Valencia: Tirant lo Blanch, 2011.

WEIGAND, Thomas. “Direito de necessidade para carros autônomos?”. In: ESTELLITA, Heloísa; LEITE, Alaor (Org.). **Veículos autônomos e Direito penal.** São Paulo: Marcial Pons, 2019.

WHITE, Trevor N. e BAUM, Seth D.. Liability for present and future robotics technology”. In: LIN, Patrick; JENKINS, Ryan; ABNEY, Keith (Ed.). *Robot Ethics 2.0. From Autonomous Cars to Artificial Intelligence.* Nova York: Oxford University Press, 2020.

WITTGENSTEIN, Ludwig. **Investigaciones Filosóficas.** n° 65. Versão bilíngue, traduzida por Alfonso García Suárez e Ulises Moulines, Barcelona: Ed. Crítica, 2002.

WITTGENSTEIN, Ludwig. **Philosophische Untersuchungen.** Frankfurt-am-Main: Suhrkamp, 1971.

Recebido: 07/06/2023