

A inteligência artificial no banco dos réus

28/11/2023

A sociedade contemporânea é, conforme demonstramos em publicação anterior[1], fortemente marcada pelo uso de meios digitais. A inteligência artificial (IA) tem ampliado seu protagonismo, o que leva ao questionamento sobre eventuais reflexos no direito penal.

É possível dividir essa tecnologia em dois tipos. O primeiro é a IA fraca [2], em que uma máquina (ou sistema) é pré-programada para adotar determinado ato diante de uma informação recebida do meio. É uma forma de automatização, em que o ser humano realiza uma programação para que uma máquina (ou um sistema) faça um ato diante de um “gatilho” previamente estabelecido.

Exemplo da utilização de IA fraca com reflexos na esfera penal é extraído das ações, nas décadas de 1980 e 1990, do IRA (*Irish Republican Army*) nos notórios conflitos na Irlanda do Norte. Os terroristas[3], para atingirem agentes do exército britânico, desenvolveram uma espécie de IA em que bombas escondidas em determinados locais somente eram acionadas e detonadas quando as equipes de oficiais disparavam *flashes* de máquinas fotográficas para documentar cenas de crimes ou locais suspeitos. Os *flashes* das câmeras eram captados por um sistema que acionava a detonação dos explosivos. Tudo ocorria somente pelo mecanismo de IA, sem interferência humana ou controle remoto.

TJ-PE



Hipóteses como a acima relatada não apontam maiores problemas na medida em que, frente ao artigo 29 do Código Penal Brasileiro, a responsabilidade penal é daquele ser humano que realizou a programação, visto que praticou conduta para obtenção de um resultado ilícito (penal), ainda que mediante um sistema ou máquina. É a mesma consequência jurídica aplicada para um agente que, por exemplo, estabelece uma armadilha com uma arma de fogo diante de uma porta ou janela.

Conduta humana, nexos causal, resultado, elemento subjetivo e culpabilidade estão claramente delineados na hipótese da IA fraca, sendo plenamente verificável a ação humana (programação da IA) como elemento suficiente para caracterização do crime. A maior divergência estaria na prova da existência de dolo ou culpa na conduta de programar a IA. Mas essa dificuldade é inerente a qualquer tipo penal que admita a figura do dolo e da culpa, sendo tema que, como já exploramos[4], gera relevante debate sobre o critério objetivo para diferenciação.

A questão se torna tormentosa quando tratamos da IA de segunda geração ou IA de *inductive learning*[5]. Esse tipo, fortemente marcado pelo *machine learning* ou *deep learning*, tem como característica a possibilidade de aprendizagem autônoma. Trata-se, realmente, de criatividade da máquina ou sistema que, adotando uma premissa pré-programada (por exemplo, o aumento

de produtividade em uma fábrica), busca produzir novas soluções diante da análise conjunta da premissa programada, das respostas anteriormente produzidas e das novas informações disponíveis. Não se trata de aplicar uma solução única e pré-programada, mas sim de criar soluções inovadoras e inéditas. A resolução para cada caso concreto é produzida pela IA.

O foco do nosso estudo está exatamente na ocorrência de resultados não desejados pela norma penal em função da atuação de IA de segunda geração. Exemplo dessa situação pode ser retirado da obra de Gabriel Hallevy[6], quando aponta que, em 1981, um operário japonês de uma fábrica de motocicletas teve sua vida encerrada quando uma máquina o esmagou. A morte ocorreu quando a IA, cuja premissa era o aumento da produtividade, interpretou o empregado como uma ameaça a sua missão. Logo, a forma mais eficaz de eliminar a ameaça foi a retirada do empregado do cenário, o que se realizou, embora ao custo da extinção da vida humana.

Essa situação pode, nos dias de hoje, ser ampliada para uma infinidade de acontecimentos, tais como: IA que determina que o carro autônomo invada a calçada, atropеле e mate diversas pessoas, visto que seria uma forma mais eficaz de garantir a segurança dos ocupantes do veículo ou uma IA que, em caso de escassez de recursos médicos (oxigênio, por exemplo), decide calcular quais pacientes têm mais chances de sobrevivência e, simplesmente, desliga as máquinas daqueles que não oferecem boas chances.

Os exemplos são infindáveis e invadem também o setor público, como a discussão sobre o Compas (uma forma de IA com aplicação na justiça criminal dos Estados Unidos) e os algoritmos utilizados para determinar a pena, a possibilidade de reincidência e a necessidade ou não de prisão cautelar[7]. Questões como a possibilidade da IA praticar abuso de autoridade ou discriminação no momento de estabelecer o montante da pena ou a necessidade de prisão cautelar vêm à luz.

É preciso, portanto, regulamentar as situações em que a IA de segunda geração gera resultados tidos como crimes. Em outras palavras, as consequências das atuações que impliquem violações a bens jurídicos protegidos pelas normas penais.

Podemos vislumbrar três hipóteses para equacionar a questão. A primeira seria a aplicação da mesma resposta gerada diante da IA fraca, isto é, respondem criminalmente as pessoas físicas e jurídicas, caso esteja prevista a responsabilidade penal do ente coletivo (o que, no Brasil, apesar da viabilidade constitucional, somente é estabelecida para crimes ambientais[8]). Não acolhemos essa hipótese, na medida em que haveria, ao fim e ao cabo, responsabilidade penal objetiva, uma vez que a IA de segunda geração, diferentemente da IA fraca, cria respostas não previstas ou pré-programadas por pessoas físicas e jurídicas.

A única forma de viabilizar essa solução seria estabelecer qual o risco permitido na programação de IA de segunda geração e, rompido esse limite, responsabilizar criminalmente os programadores e a pessoa jurídica responsável. Trata-se da criação de um tipo penal de perigo abstrato. Embora essa solução seja juridicamente possível, há dúvidas sobre a viabilidade, em termos técnicos, de determinar o risco permitido e não permitido para programação de IA com capacidade criativa.

A segunda hipótese é aquela vislumbrada por Victor Augusto Estavam Valente[9], que defende o tratamento da situação em comento como um produto defeituoso. Pensamos que essa solução esbarra no mesmo problema da anterior. Se a IA tem capacidade criativa, no final, haveria uma responsabilidade penal objetiva das pessoas (físicas e jurídicas), pois o crime não seria resultado de um defeito da programação da IA (a atuação da capacidade criativa não pode ser previamente estabelecida e os resultados produzidos não são decorrentes de um sistema defeituoso). Em nosso entender, a única forma de adotar essa solução seria criar um tipo penal omissivo que obrigasse os responsáveis pela IA (ou seus criadores) a, diante de notificação de autoridade pública, cessar a atividade do sistema e realizar o *recall*, corrigindo o código-fonte.

Por fim, haveria a solução vislumbrada por Gabriel Hallevy[10] que afirma que a IA de segunda geração possui algumas capacidades análogas às humanas e que essas capacidades são suficientes para reconhecer personalidade jurídica e determinar a imposição de sanção penal. Trata-se de uma terceira fase da evolução da responsabilidade penal. A primeira fase é caracterizada pela aplicação exclusiva em face de seres humanos. A segunda fase abarca as pessoas jurídicas, especialmente em países de *common-law*[11]. A terceira etapa que se inicia deve englobar a IA de segunda geração que, diante de sua capacidade criativa (análoga a uma capacidade humana), detém os requisitos mínimos para ter sua personalidade jurídica reconhecida e ser considerada sujeito ativo de crimes.

São aplicáveis, para a responsabilidade penal da IA, as mesmas premissas fixadas em 1909, no caso *New York Central & Hudson River Railroad v. United States*[12], em que, ao tratar da responsabilidade penal de entes coletivos, restou decidido que “não é possível fechar os olhos para o fato de que a maioria das operações comerciais nos tempos modernos são realizadas por meio de tais entes e, particularmente, o comércio interestadual está praticamente todo em suas mãos” (tradução nossa).

Ressalte-se que o direito continental não é incompatível com a responsabilidade penal da IA. Conforme afirma Carlos Gómez-Jara Díez[13]: “a comunicação goza da mesma autorreferência que a consciência” (tradução nossa). Logo, o crime pode ser praticado por conduta humana ou por comunicação (produção de norma) com relevância jurídica. Pessoas jurídicas e IA não praticam condutas como humanos, mas produzem normas (comunicação) com relevância jurídica e, dessa forma, são consideradas como sujeitos ativos de crimes[14]. É por essa razão que a inexistência de dolo, culpa e de conduta, nos termos aplicados para seres humanos, não impede a responsabilidade penal de entes coletivos e da IA.

Não é possível ignorar os efeitos da IA de segunda geração em nossas vidas. Portanto, da mesma forma que a pessoa jurídica tornou-se sujeito ativo de crimes, é a vez da IA criativa. Concordamos com Gabriel Hallevy[15] quando afirma que: “ou nós impomos responsabilidade criminal sobre entidades de IA ou devemos mudar a definição básica de responsabilidade criminal que se desenvolveu ao longo de milhares de anos e abandonar os entendimentos tradicionais de responsabilidade criminal” (tradução nossa).

A inexistência de patrimônio própria da IA não é impeditivo para a solução vislumbrada por Gabriel Hallevy e por nós acolhida. Uma IA pode (ainda) não possuir bens (contas bancárias, veículos etc.). Mas, em muitos casos, possui capacidade de gerar valores (basta imaginar uma

IA que cria, por exemplo, fotos personalizadas na internet mediante pagamento), bem como pode possuir elevado valor de mercado. Essas circunstâncias são suficientes para que, além das sanções penais análogas àquelas aplicáveis às pessoas jurídicas e adaptáveis ao instituto, a IA também responda pela multa e outras indenizações decorrentes de atos próprios.

Ademais, nada impede que, além da responsabilidade penal da IA, seja estabelecida, de forma subsidiária, a responsabilidade das pessoas físicas e jurídicas criadoras ou gestoras da IA pelo ressarcimento de danos.

Por fim, ao Brasil, cabe ampliar a responsabilidade penal da pessoa jurídica, bem como estabelecer a responsabilidade da IA de segunda geração por atos próprios. Os brocados *societa delinquere non potest* e *machina delinquere non potest* estão em confronto com a realidade contemporânea e não são capazes de afastar a responsabilidade penal de pessoas jurídicas e de IA de segunda geração.

[1]ARRUDA, Eloisa de Sousa; CORRERA, Marcelo Carita. Utilização de Malware para Coleta de Provas no Processo Penal. **Conjur**, 22 de julho de 2023. Disponível em: <https://www.conjur.com.br/2023-jul-22/arrudae-correra-malware-coleta-provas-processo-penal>. Acesso em 13 ago 2023.

[2]VALENTE, Victor Augusto Estevam. Inteligência Artificial e o Direito Penal: O Propósito da responsabilidade criminal em decorrência de sistema tecnológicos altamente complexo nas empresas. Belo Horizonte: D'Plácido, 2023. P. 25.

[3]FULTON, Kevin. Double Agent. My Secret Life Undercover in The Ira. United Kingdom: John Blake. P. 129 – 130.

[4]PEREIRA, Claudio José Langroiva; IANNI, Gabriela de Castro; CORRERA, Marcelo Carita. Dolo Eventual e Culpa Consciente. Revista Magister de Direito Penal e Processual Penal. Porto Alegre, V. 19, nº 111, p. 33-50, dez. 2022/jan. 2023.

[5]VALENTE, Victor Augusto Estevam. Inteligência Artificial e o Direito Penal: O Propósito da responsabilidade criminal em decorrência de sistema tecnológicos altamente complexo nas empresas. Belo Horizonte: D'Plácido, 2023. P. 26-27.

[6]HALLEVY, Gabriel. When Robots Kill: Artificial Intelligence Under Criminal Law. Boston: Northeastern University Press, 2013. P. 38.

[7]DIETER, Maurício Stegemann. Política Criminal Atuarial. A Criminologia do Fim da História. Tese de Doutorado. Universidade Federal do Paraná. Curitiba: 2012. P. 185/186.

[8]CORRERA, Marcelo Carita. Responsabilidade Penal da Pessoa Jurídica: a Teoria do Delito em face da Autorresponsabilidade do Ente Coletivo. Belo Horizonte: D'Plácido, 2021. Passim.

[9]VALENTE, Victor Augusto Estevam. Inteligência Artificial e o Direito Penal: O Propósito da responsabilidade criminal em decorrência de sistema tecnológicos altamente complexo nas empresas. Belo Horizonte: D'Plácido, 2023. P. 129-130.



- [10] HALLEVY, Gabriel. When Robots Kill: Artificial Intelligence Under Criminal Law. Boston: Northeastern University Press, 2013. P. 177/178.
- [11] CORRERA, Marcelo Carita. Responsabilidade Penal da Pessoa Jurídica: a Teoria do Delito em face da Autorresponsabilidade do Ente Coletivo. Belo Horizonte: D'Plácido, 2021. Passim.
- [12] ESTADOS UNIDOS DA AMÉRICA. SUPREMA CORTE. New York Central & Hudson River Railroad v. United States 212 U.S. 481 (1909). Disponível em: <https://supreme.justia.com/cases/federal/us/212/481/>. Acesso em 16 nov. 2023.
- [13] DÍEZ, Carlos Gómez-Jara. Cuestiones Fundamentales de Derecho Penal Económico. Parte General y Parte Especial. Bogotá: Universidad Externado, 2012. posição 488.
- [14] CORRERA, Marcelo Carita. Responsabilidade Penal da Pessoa Jurídica: a Teoria do Delito em face da Autorresponsabilidade do Ente Coletivo. Belo Horizonte: D'Plácido, 2021. P. 70-75.
- [15] HALLEVY, Gabriel. When Robots Kill: Artificial Intelligence Under Criminal Law. Boston: Northeastern University Press, 2013. P. 178.

Fonte: <https://conjur.jumps.com.br/2023-nov-28/a-inteligencia-artificial-no-banco-dos-reus/>