

A RESPONSABILIDADE PENAL POR DECISÕES AUTOMATIZADAS: LIMITES E DESAFIOS DA IMPUTAÇÃO PENAL NA ERA DA INTELIGÊNCIA ARTIFICIAL

Mario Marrathma Lopes de Oliveira

Mestrando em Direito pelo Centro Universitário Christus. Bolsista da Fundação Cearense De Apoio Ao Desenvolvimento Científico E Tecnológico (FUNCAP). <http://lattes.cnpq.br/9268389328420260>.
<https://orcid.org/0000-0001-5938-3999>. E-mail: mario@marrathma.com

Rodston Ramos Mendes de Carvalho

Advogado. Professor Universitário. Pós-Doutorando em Ciências Jurídicas pelo Instituto Internacional de Educação e Pesquisa do Rio de Janeiro-RJ, em parceria com a Universidade UNIDA de Asunción – PY. <http://lattes.cnpq.br/2931337393815825>. <https://orcid.org/0009-0009-8115-9008>.
E-mail: rodstoncarvalho@gmail.com

Tátia de Jesus Alcântara Duarte

Mestre em Administração de Empresas pela Universidade Presbiteriana Mackenzie. <http://lattes.cnpq.br/5556182226113302>. <https://orcid.org/0000-0003-0232-7720>
E-mail: consultoria.mentebrilhante@gmail.com

Aimme Beatrice de Oliveira Dutra Cordeiro

Mestranda em Direito pelo Centro Universitário Christus (UNICHRISTUS). Advogada. <http://lattes.cnpq.br/3913654820228210>. <https://orcid.org/0009-0008-2369-1891>.
E-mail: aimmebeatrice@gmail.com

Sara Adelina de Oliveira

Estudante de Direito na Afya Centro Universitário São João Del-Rei, bolsista do Programa Universidade para Todos (ProUni). <https://orcid.org/0009-0008-4738-0246>. E-mail: saraadelina8081@gmail.com

DOI-Geral: <http://dx.doi.org/10.47538/RA-2026.V5N1>

DOI-Individual: <http://dx.doi.org/10.47538/RA-2026.V5N1-68>

RESUMO: O avanço da inteligência artificial e a crescente adoção de sistemas automáticos para decisões sensíveis têm gerado os mais relevantes efeitos para o mundo jurídico, especialmente no que se refere aos fundamentos e limites da responsabilidade penal. O artigo procura analisar quais são os desafios dogmáticos da atribuição de responsabilidade penal em cenários em que decisões relevantes são mediadas, influenciadas ou criadas por sistemas algorítmicos, à luz do ordenamento jurídico brasileiro e da teoria clássica do delito. A pesquisa percorre examinar os fundamentos constitucionais da responsabilidade penal, o conceito de culpabilidade e os requisitos clássicos da imputação subjetiva, confrontando tais elementos com as características das tecnologias contemporâneas, como opacidade algorítmica, autonomia, e a fragmentação das cadeias decisórias. A partir de pesquisa teórica e dogmática, elaborada a partir da Constituição Federal, do Código Penal, do Marco Civil da Internet, da Lei Geral de Proteção de Dados e da literatura jurídica nacional e estrangeira é discutido o descompasso da atribuição de responsabilidade penal direta aos sistemas de inteligência artificial e a necessidade de redirecionar a imputação aos sujeitos humanos e organizações que planejam, desenvolvem, implementam e supervisionam estas tecnologias. Também é feito o exame sobre o papel da transparência, da explicabilidade e, da intervenção humana

como condições técnicas e normativas para a legitimidade das decisões automatizadas e da retenção das garantias penais fundamentais. Conclui-se que, a despeito de a inteligência artificial trazer novos desafios para a teoria da imputação, o modelo penal em vigor deve subsistir, com adaptações interpretativas que garantam a centralidade da pessoa humana, a legalidade, a culpabilidade, a pessoalidade da pena e a proibição da responsabilidade penal objetiva, bem como devem traçar, em plano paralelo, a distribuição das responsabilidades civis e administrativas originadas das decisões automatizadas.

PALAVRAS-CHAVE: Inteligência Artificial. Responsabilidade Penal. Imputação Penal. Decisões Automatizadas. Culpabilidade.

CRIMINAL LIABILITY FOR AUTOMATED DECISIONS: LIMITS AND CHALLENGES OF CRIMINAL IMPUTATION IN THE ERA OF ARTIFICIAL INTELLIGENCE

ABSTRACT: The advancement of artificial intelligence and the growing adoption of automated systems for sensitive decisions have produced the most significant effects on the legal domain, particularly concerning the foundations and limits of criminal liability. This article aims to analyze the dogmatic challenges involved in attributing criminal liability in scenarios where relevant decisions are mediated, influenced, or generated by algorithmic systems, viewed through the lens of the Brazilian legal order and classical crime theory. The research examines the constitutional foundations of criminal liability, the concept of culpability, and the classical requirements of subjective imputation, juxtaposing these elements against the features of contemporary technologies, such as algorithmic opacity, autonomy, and the fragmentation of decision-making chains. From theoretical and dogmatic research grounded in the Federal Constitution, the Penal Code, the Internet Civil Framework, the General Data Protection Law, and national and foreign legal literature, the article discusses the mismatch of directly attributing criminal liability to artificial intelligence systems and the necessity of redirecting imputation to human agents and organizations that design, develop, implement, and oversee these technologies. It further assesses the role of transparency, explainability, and human intervention as technical and normative conditions for the legitimacy of automated decisions and the preservation of fundamental penal guarantees. In conclusion, despite artificial intelligence introducing new challenges to imputation theory, the prevailing penal model must endure, incorporating interpretive adaptations that uphold the centrality of the human person, legality, culpability, the personal nature of penalties, and the prohibition of objective criminal liability, while simultaneously delineating the allocation of civil and administrative responsibilities stemming from automated decisions.

KEYWORDS: Artificial Intelligence. Criminal Liability. Penal Imputation. Automated Decisions. Culpability.

INTRODUÇÃO

Estudam-se os limites e desafios da responsabilidade penal em circunstâncias em que decisões relevantes são mediadas, influenciadas ou produzidas por sistemas de

inteligência artificial. O foco consiste na posição do agente humano que decide com o apoio de modelos de algoritmos opacos. Analisa-se em que medida a teoria clássica da imputação penal, que tem como fundamentos a ação humana, a culpabilidade e o domínio do fato, é tensionada por características como opacidade algorítmica, autonomia operacional e desagregação das cadeias decisórias; e se é juridicamente admissível atribuir responsabilidade penal direta a sistemas de IA ou se a imputação deve ficar adstrita aos sujeitos humanos e organizações que concebem, desenvolvem, contratam e operam essas tecnologias

A pesquisa justifica-se, em primeiro lugar, pela crescente implementação de sistemas de inteligência artificial em processos decisórios sensíveis do setor público e privado, como: justiça criminal, saúde, concessão de benefícios e administração de recursos escassos. Ferramentas de *Machine Learning* e de *Big Data* passaram a atuar em espaços que eram tradicionalmente da deliberação humana, com efeitos diretos sobre direitos fundamentais - como dignidade, igualdade, privacidade e devido processo legal. Assim como o risco à legitimidade das decisões estatais e empresariais diante de erros relevantes decorrente de recomendações automatizadas.

Em segundo lugar, a expansão das decisões automatizadas acrescenta tensões específicas à dogmática penal. O modelo clássico de responsabilidade (baseada no comportamento consciente e voluntário de um sujeito individual) encontra limites evidentes no confronto com sistemas marcados pela opacidade técnica. Problematicam ainda mais a questão o comportamento probabilístico e as cadeias decisórias distribuídas entre desenvolvedores, fornecedores, gestores e operadores. Nessas condições, faz-se progressivamente mais difícil reconstruir o nexo, entre o resultado lesivo e a vontade de um agente humano determinado. Dificulta também identificar quem tinha domínio relevante sobre o risco e quais deveres de cuidado foram violados.

Em terceiro lugar, a relevância do estudo emana da discussão atual sobre a viabilidade da responsabilização penal (imediatamente ou de forma compartilhada) de sistemas de inteligência artificial. Se insere na equação: a criação de novas classes de sujeitos de imputação. Essa movimentação parece chocar-se com os fundamentos constitucionais do Direito Penal pátrio, mormente, a pessoalidade da pena, a

culpabilidade e a proibição da responsabilidade penal objetiva. O sistema criminal carece de uma sistematização dogmática aplicável a casos concretos de decisão automatizada.

Ao delimitar os contornos desta discussão, o artigo tem como objetivo estabelecer critérios para diferenciar, de um lado, as situações nas quais o erro decisório pode ser atribuído a falhas humanas na concepção, contratação ou supervisão; e, de outro, aquelas em que os resultados são fruto de riscos tecnológicos que não podem ser geridos à luz do conhecimento atual.

Este estudo possui uma abordagem teórica e dogmática, fundamentando-se em uma pesquisa bibliográfica que inclui fontes tanto nacionais quanto internacionais sobre questões relativas à responsabilidade penal, inteligência artificial e decisões automatizadas. O arcabouço normativo examinado é centrado, em grande parte, na Constituição Federal, no Código Penal, no Marco Civil da Internet e na Lei Geral de Proteção de Dados, bem como em documentos e literatura especializada elaborados, em particular, entre os anos de 2017 e 2025. A revisão da literatura é de caráter narrativo, sem protocolo sistematizado formal, sendo, portanto, passível de vieses de “*selection bias*” em relação à seleção das fontes bibliográficas; por isso suas conclusões têm viés predominantemente jurídico-dogmático e limitado ao cenário brasileiro, somente servindo de referência comparativa para outros ordenamentos.

Este artigo é organizado em quatro tópicos, além desta introdução e das considerações finais: O tópico segundo examina os fundamentos dogmáticos da imputação penal e os limites do modelo clássico sob a perspectiva das tecnologias de inteligência artificial. O terceiro examina como decisões automatizadas, opacidade algorítmica e fragmentação da cadeia decisória fornecem um quadro de crise de adaptação da imputação penal. O quarto analisa a distribuição da responsabilidade penal na cadeia de desenvolvimento, contratação e utilização dos sistemas de inteligência artificial. Por fim, o quinto analisa o papel da transparência, da explicabilidade e da intervenção humana, como condições técnicas e normativas para a legitimidade das decisões automatizadas e para a possibilidade de responsabilização penal.

FUNDAMENTOS DOGMÁTICOS DA IMPUTAÇÃO PENAL E OS LIMITES DO MODELO CLÁSSICO

O Direito Penal moderno estrutura-se a partir da ideia de que somente condutas humanas conscientes e voluntárias podem ser objeto de imputação penal. No ordenamento jurídico brasileiro, esse pressuposto decorre diretamente da Constituição da República, que consagra a dignidade da pessoa humana e a pessoalidade da pena, bem como do Código Penal, que constrói a teoria do delito sobre a ação humana dolosa ou culposa, afastando qualquer forma de responsabilidade penal objetiva (Brasil, 1988).

Esse modelo dogmático pressupõe a existência de um sujeito capaz de compreender o caráter ilícito do fato e de determinar-se de acordo com esse entendimento. A imputação penal exige, portanto, não apenas a produção de um resultado lesivo, mas a possibilidade de reconstrução normativa da conduta, do nexos causal e do elemento subjetivo. Trata-se de uma arquitetura conceitual que foi concebida para regular comportamentos humanos, em contextos nos quais o controle da ação e a previsibilidade do resultado se encontram, ao menos em tese, sob domínio do agente (Ortiz, 2023).

A introdução de sistemas de inteligência artificial nesse cenário rompe com diversos desses pressupostos. Conforme observa Ortiz (2023), os sistemas algorítmicos operam com graus variados de autonomia e opacidade, o que dificulta a identificação do processo decisório que conduziu a determinado resultado.

Sachoulidou (2024) acrescenta que os modelos clássicos de responsabilidade penal não foram projetados para lidar com estruturas decisórias que não possuem consciência, vontade ou intencionalidade, mas que, ainda assim, produzem efeitos socialmente relevantes e potencialmente lesivos.

Dessa forma, a hipótese de atribuir responsabilidade penal direta a sistemas de inteligência artificial mostra-se incompatível com os fundamentos dogmáticos do Direito Penal. Sistemas algorítmicos não são sujeitos de direito, não possuem capacidade de culpabilidade e não podem ser destinatários de pena em sentido próprio. O problema jurídico desloca-se, portanto, para a definição de como e em que medida os sujeitos humanos envolvidos no desenvolvimento, implementação, operação e supervisão dessas tecnologias podem ser responsabilizados (Sachoulidou, 2024 p. 49)

Nesse sentido, na configuração atual da dogmática penal e da legislação do Brasil, não há condições para o reconhecimento da responsabilidade penal direta para sistemas de inteligência artificial. Estes não são sujeitos de direito, não detêm capacidade para ser culpáveis e não podem ser destinatários de pena em sentido próprio (Sachoulidou, 2024; Ortiz, 2023). A tecnologia continua, portanto, a ser qualificada, do ponto de vista jurídico, como um instrumento agregado a cadeias decisórias humanas, na medida em que a imputação penal deve ser novamente atribuída a sujeitos humanos (e, nas situações autorizadas pela lei, a pessoas jurídicas) com base em deveres de cuidado, controle e supervisão quanto à concepção, implementação e utilização desses sistemas (Ávila; Corazza, 2022; Oliveira; Ávila, 2024).

O desafio central, portanto, não consiste em substituir o sujeito humano pelo artefato tecnológico no polo passivo da imputação penal, mas em repensar os critérios jurídicos de atribuição de responsabilidade em contextos nos quais a ação humana se encontra mediada por sistemas complexos. A tecnologia, ainda que dotada de elevado grau de autonomia funcional, não age por si mesma no plano normativo, mas integra cadeias decisórias concebidas, treinadas, parametrizadas e supervisionadas por pessoas naturais ou jurídicas.

Essa constatação desloca o debate para a análise dos deveres de cuidado, de controle e de vigilância que recaem sobre os diversos agentes envolvidos no ciclo de vida dos sistemas de inteligência artificial. Desenvolvedores, empresas fornecedoras, gestores e operadores não ocupam posições equivalentes na cadeia tecnológica, o que exige uma reconstrução mais refinada das esferas de responsabilidade, compatível com a complexidade técnica desses ambientes. Conforme observa Ortiz (2023), a dificuldade não está apenas em identificar o dano, mas em reconstruir juridicamente a contribuição relevante de cada agente para a produção do resultado.

Além disso, a fragmentação dos processos decisórios introduz um elemento adicional de incerteza na aplicação das categorias clássicas de autoria e participação. Em sistemas tradicionais, a imputação penal parte da identificação de uma conduta individual dotada de sentido jurídico-penal. Em ambientes mediados por algoritmos, contudo, o resultado pode emergir de uma combinação de decisões de projeto, escolhas de parametrização, dados de treinamento e interações operacionais, nenhuma das quais, isoladamente, explica por completo o evento lesivo (Ortiz, 2023, p. 89)

Sachoulidou (2024 p. 29) destaca que essa diluição causal desafia os modelos clássicos de atribuição de responsabilidade, sem, contudo, autorizar sua superação:

Diante desse cenário, a preservação do núcleo dogmático do Direito Penal impõe que a imputação continue ancorada na ideia de domínio humano sobre o risco. Ainda que esse domínio não se manifeste de forma direta e imediata no momento da produção do resultado, ele subsiste nas fases anteriores de concepção, implementação e controle do sistema. É nesse plano que se deve buscar a reconstrução normativa da responsabilidade, a partir da verificação de falhas relevantes nos deveres de cuidado, de supervisão ou de intervenção (Sachoulidou, 2024 p. 29)

Por essa razão, para Junqueira (2020) a inteligência artificial deve ser juridicamente compreendida como instrumento técnico inserido em contextos organizacionais e decisórios humanos, e não como centro autônomo de imputação. Essa abordagem permite preservar os princípios estruturantes do Direito Penal, especialmente a culpabilidade e a pessoalidade da pena, ao mesmo tempo em que abre espaço para uma adaptação criteriosa das categorias clássicas às novas realidades tecnológicas.

Assim, longe de representar uma crise terminal do modelo penal, a emergência de sistemas automatizados revela, na verdade, a necessidade de um refinamento interpretativo da teoria da imputação, capaz de lidar com ambientes de decisão complexos sem abdicar dos fundamentos normativos que garantem a racionalidade, a legitimidade e os limites do poder punitivo.

DECISÕES AUTOMATIZADAS, OPACIDADE ALGORÍTMICA E CRISE DA IMPUTAÇÃO PENAL

Os sistemas contemporâneos de inteligência artificial, especialmente aqueles baseados em aprendizagem de máquina, caracterizam-se por processos decisórios que nem sempre são plenamente compreensíveis, mesmo para seus desenvolvedores. Essa característica, frequentemente descrita como “opacidade algorítmica”, compromete a transparência e a rastreabilidade causal das decisões automatizadas (Ortiz, 2023).

No plano jurídico, essa opacidade gera um problema central: a dificuldade de reconstruir o itinerário decisório que levou ao resultado danoso. Se não é possível compreender como o sistema chegou à determinada conclusão, torna-se igualmente

complexo identificar qual agente humano contribuiu de modo juridicamente relevante para a produção do resultado. Sachoulidou (2024) aponta que esse fenômeno desafia diretamente as categorias tradicionais de autoria, participação e culpa.

A literatura técnica e jurídica tem indicado que, em muitos casos, os danos decorrentes de decisões automatizadas resultam de uma cadeia complexa de ações e omissões distribuídas entre diversos atores: programadores, empresas desenvolvedoras, gestores do sistema, operadores e usuários finais. Alves (2022), ao analisar o caso dos veículos autônomos, demonstra que essa fragmentação da cadeia decisória dificulta a individualização da responsabilidade penal, sobretudo quando não há violação clara e direta de um dever objetivo de cuidado por parte de um agente específico.

No campo da medicina, Junqueira (2020) observa que a utilização de sistemas automatizados de apoio à decisão clínica tensiona o conceito tradicional de culpa profissional, na medida em que o médico passa a compartilhar, ainda que parcialmente, o processo decisório com uma ferramenta tecnológica cujo funcionamento interno pode não ser plenamente compreendido. Situação semelhante é descrita por Carvalho e Giongo (2020) em relação a tecnologias aplicadas a contextos sensíveis de risco.

Diante desse quadro, a crise da imputação penal não decorre apenas da existência de novas tecnologias, mas da inadequação dos instrumentos clássicos para lidar com ambientes decisórios marcados por complexidade, opacidade e divisão funcional de tarefas.

Além disso, a própria estrutura probabilística de muitos sistemas baseados em aprendizagem de máquina introduz um elemento adicional de incerteza normativa. Diferentemente de programas tradicionais, cujo comportamento pode ser previsto a partir de regras previamente estabelecidas, esses sistemas ajustam seus parâmetros a partir de grandes volumes de dados, o que torna seus resultados, em certa medida, contingentes e dependentes de padrões estatísticos. Ortiz (2023) ressalta que essa característica compromete a previsibilidade ex ante do comportamento do sistema, elemento que é central para a construção jurídica do dever de cuidado e da culpa.

Sob a perspectiva da responsabilidade penal, essa imprevisibilidade coloca em tensão o próprio fundamento da censura penal, que pressupõe a possibilidade de exigir do agente uma conduta diversa. Quando o resultado decorre de uma interação complexa entre dados, modelos algorítmicos e contextos de uso, a avaliação da previsibilidade e da evitabilidade do dano deixa de ser um exercício simples de reconstrução causal e passa a exigir uma análise muito mais sofisticada das condições técnicas e organizacionais que envolveram a decisão automatizada (Sachoulidou, 2024).

Nesse cenário, a tentativa de resolver o problema por meio da simples ampliação das categorias tradicionais de autoria ou participação revela-se insuficiente. A fragmentação das funções e a distribuição de tarefas ao longo da cadeia tecnológica fazem com que nenhum dos agentes envolvidos detenha, isoladamente, o controle integral do processo decisório. Como observa Sachoulidou (2024), essa diluição do domínio sobre o fato desafia os modelos clássicos de imputação, sem, contudo, autorizar a renúncia aos seus pressupostos fundamentais.

Por essa razão, a crise da imputação penal em contextos de decisão automatizada deve ser compreendida, sobretudo, como uma crise de adaptação dos critérios normativos às novas formas de produção de riscos. Não se trata de negar a aplicabilidade do Direito Penal, mas de reconhecer que os instrumentos tradicionais de reconstrução da conduta, do nexa e da culpa precisam ser interpretados à luz de ambientes tecnicamente mediados e organizacionalmente complexos.

Assim, a opacidade algorítmica e a fragmentação das cadeias decisórias não eliminam a necessidade de responsabilização penal, mas impõem uma reformulação do modo como se identifica a contribuição juridicamente relevante de cada agente humano. A decisão pode ser automatizada; a responsabilidade penal, entretanto, permanece necessariamente humana e deve ser reconstruída a partir dos deveres de controle, supervisão e intervenção que recaem sobre aqueles que concebem, implementam e operam esses sistemas.

Se um juiz, um operador público ou um outro tipo de profissional decisor age por meio de recomendações de inteligência artificial, o resultado lesivo ainda será, em princípio, juridicamente imputado à decisão humana, cuja conduta está submetida às mesmas normas típicas sobre ação, vínculo e culpa. A responsabilidade penal não recai em artefatos tecnológicos, não sujeito de direitos nem capaz de culpa, mas nos agentes

que, dolosa ou culposamente, criam, aceitam os riscos ou não mitigam os riscos juridicamente relevantes associados ao uso da IA (Sapir, s.d.).

Distinga-se entre as situações em que o agente humano é somente um “executor acrítico” do algoritmo e aquelas em que o resultado danoso resulta de erro imprevisível e inevitável do sistema: Na primeira, há culpa do agente que confia plenamente na sugestão da IA. Delega a tomada de decisão para a máquina e falha nos deveres objetivos de cuidado, supervisão e validação humana incumbentes a sua função. Já na segunda hipótese, quando o erro pode ser atribuído a uma falha técnica robusta, cuja existência se torna invisível mesmo diante da diligência esperada do agente médio e se não vislumbra nenhum sinal de risco ou de falsificação que indicaria a intervenção, costuma não haver culpa do decisor. A discussão é deslocada para a eventual responsabilidade por outros dos elos da cadeia tecnológica, como desenvolvedores ou fornecedores ou para a responsabilidade não penal.

A possível responsabilização penal dependerá, assim, da averiguação de culpa, notadamente nas situações em que o agente confia cegamente num sistema intransparente, sem levar em conta suas limitações, vícios conhecidos ou condições de uso, infringindo o dever objetivo de cuidado a que está submetido em sua atividade profissional (Oliveira; Cury, 2025; Jatobá, 2021). Nesses cenários, se destacam figuras como a *culpa in eligendo* e *in vigilando* na escolha, parametrização e acompanhamento dos sistemas algorítmicos, e a exigência de diligência acrescida na verificação dos resultados pautados como anômalos ou claramente incongruentes apresentados pelo sistema.

Em outras palavras, podem ocorrer formas de imperícia, imprudência ou negligência ligadas a déficits variáveis de supervisão; à não observância de protocolos de validação humana; e à ausência de intervenções mínimas exigíveis diante de evidências de erro ou de vieses discriminatórios manifestos das respostas algorítmicas, na linha do que foi dito sobre a exigência de que o humano permaneça no “circuito” (*human in the loop*) e em posição de realizar crítico controle da decisão automatizada (Ávila; Corazza, 2022; Watanabe de Moraes, 2023).

Em termos dogmáticos, na imputação do resultado danoso ao decisor humano deve-se demonstrar que ele podia e tinha o dever de conhecer o risco que a utilização

daquela ferramenta específica criou. É preciso apontar que, independentemente disso, decidiu segui-la sem o grau de ceticismo, controle e transparência exigidos pelos padrões profissionais e normativos vigentes para o caso (Sapir, s.d.).

TRANSPARÊNCIA, EXPLICABILIDADE E A CENTRALIDADE DA INTERVENÇÃO HUMANA NA RECONSTRUÇÃO DA RESPONSABILIDADE PENAL

Diante dos desafios impostos pelos sistemas automatizados de decisão e por outras aplicações contemporâneas da inteligência artificial, parte significativa da doutrina tem defendido a necessidade de reforçar os requisitos de transparência e explicabilidade dos modelos algorítmicos como condição para a preservação da legitimidade jurídica desses sistemas. Em contextos nos quais decisões tecnicamente mediadas produzem efeitos relevantes sobre direitos fundamentais e podem ensejar consequências penais, a opacidade dos processos decisórios revela-se incompatível com as exigências mínimas de controle normativo e de responsabilização.

Nunes e Andrade (2023) sustentam que a chamada inteligência artificial explicável representa um possível caminho para aumentar a confiabilidade e a legitimidade dos sistemas algorítmicos, na medida em que permite maior compreensão acerca dos critérios e das variáveis que influenciam os resultados produzidos. Embora tal abordagem não elimine integralmente os problemas estruturais da imputação penal em ambientes tecnologicamente complexos, ela contribui para reduzir a assimetria informacional entre os agentes que desenvolvem e operam os sistemas e aqueles que sofrem os efeitos de suas decisões.

No plano normativo brasileiro, essa preocupação já encontra expressão na Lei Geral de Proteção de Dados Pessoais, que assegura ao titular o direito de obter informações claras e adequadas sobre os critérios e os procedimentos utilizados em decisões tomadas unicamente com base em tratamento automatizado de dados, bem como o direito de solicitar sua revisão por pessoa natural (Brasil, 2018). Essa exigência normativa revela uma opção legislativa clara pela preservação da intervenção humana como elemento de legitimação das decisões automatizadas.

Wimmer e Doneda (2025) aprofundam essa perspectiva ao sustentar que a chamada “humanização” das decisões automatizadas não deve ser compreendida como simples formalidade procedimental, mas como verdadeiro requisito de validade jurídica em contextos nos quais estão em jogo direitos fundamentais. A presença de um agente humano capaz de compreender, avaliar e, se necessário, corrigir o resultado produzido pelo sistema constitui condição indispensável para a manutenção de padrões mínimos de responsabilidade e controle.

Sob a ótica da responsabilidade penal, essa centralidade da intervenção humana assume papel ainda mais relevante. A possibilidade de reconstruir a imputação pressupõe a existência de sujeitos que detenham algum grau de domínio sobre o risco gerado pelo sistema e que possam ser chamados a responder por falhas relevantes nos deveres de cuidado, de supervisão ou de intervenção. (Sachoulidou, 2024).

Sem transparência mínima sobre o funcionamento do sistema e sem a possibilidade de interferência humana efetiva, a própria ideia de censura penal perde seu fundamento normativo, uma vez que se inviabiliza a identificação de um comportamento humano juridicamente relevante (Sachoulidou, 2024).

Nesse sentido, a exigência de explicabilidade não deve ser vista apenas como instrumento de proteção de direitos individuais no campo civil ou administrativo, mas também como pressuposto estrutural para a viabilidade da responsabilização penal em ambientes mediados por tecnologia. A opacidade absoluta dos sistemas algorítmicos não apenas compromete a confiança social em suas decisões, mas inviabiliza, do ponto de vista dogmático, a reconstrução da conduta, do nexo causal e do elemento subjetivo necessários à imputação penal, razão pela qual a inteligência artificial explicável tem sido apontada como requisito mínimo de legitimidade desses sistemas (Nunes e Andrade, 2023).

Desse modo, a transparência, a explicabilidade e a possibilidade de intervenção humana não são apenas boas práticas de governança tecnológica, mas sim verdadeiros pressupostos jurídicos para que a racionalidade do sistema penal permaneça preservada. Em contextos de decisão automatizada, a responsabilidade penal apenas poderá ser legitimamente afirmada se houver possibilidade de identificar sujeitos humanos que

exercçam, direta ou indiretamente, controle efetivo sobre o funcionamento e os impactos do sistema. Devem garantir a obediência às diretrizes que competem à Lei Geral de Proteção de Dados, a qual garante o direito de revisão das decisões automatizadas por pessoa natural (Brasil, 2018; Wimmer; Doneda, 2025).

Longe de configurar a superação do modelo clássico de responsabilidade penal, esse contexto faz sugerir a necessidade de reafirmar a sua fundamentação. Em chave adaptativa, vincular a legitimidade das decisões automatizadas à existência de mecanismos suficientemente robustos de controle e revisão humanos. Nas palavras de Wimmer e Doneda (2025, p. 14) “a chamada “humanização” das decisões automatizadas constitui condição de legitimidade jurídica, especialmente quando estão em jogo direitos fundamentais”.

Em termos diretos: a tecnologia pode mediar decisões; a responsabilidade penal, contudo, deve permanecer ancorada na conduta humana, na possibilidade de controle e na exigibilidade de comportamento conforme o Direito.

A NECESSÁRIA RECONDUÇÃO DA RESPONSABILIDADE PENAL AOS SUJEITOS HUMANOS

Os desafios impostos pelas decisões automatizadas e pelas tecnologias baseadas em inteligência artificial não conduzem, do ponto de vista dogmático, à necessidade de criação de novas categorias de sujeitos de responsabilidade penal.

Ao contrário, eles reforçam a centralidade do modelo clássico, segundo o qual somente pessoas naturais ou, nos limites admitidos pelo ordenamento, pessoas jurídicas podem figurar como destinatárias da sanção penal. Sistemas tecnológicos, ainda que dotados de elevada complexidade e autonomia operacional, não possuem consciência, vontade ou capacidade de culpabilidade, razão pela qual não podem ser considerados centros autônomos de imputação penal.

Nesse sentido, a resposta jurídica adequada ao problema da responsabilidade penal por decisões automatizadas não consiste em personificar a tecnologia, mas em

reconstruir os critérios de imputação aplicáveis aos sujeitos humanos que exercem domínio, ainda que mediato ou funcional, sobre o risco tecnológico.

Conforme observa Ortiz (2023), a dificuldade central não está em reconhecer a relevância penal dos resultados produzidos por sistemas automatizados, mas em identificar, no interior das cadeias técnicas e organizacionais, quais agentes violaram deveres juridicamente relevantes de cuidado, supervisão ou intervenção.

A literatura comparada aponta que os ordenamentos jurídicos têm convergido para a preservação da imputação centrada em sujeitos humanos, ainda que com a necessidade de refinamento dos conceitos de previsibilidade, evitabilidade e controle em contextos de elevada complexidade tecnológica. Essa perspectiva é compatível com os fundamentos constitucionais e dogmáticos do Direito Penal brasileiro, que permanecem ancorados na culpabilidade, na pessoalidade da pena e na vedação da responsabilidade penal objetiva (Brasil, 1988).

A recondução da responsabilidade penal aos sujeitos humanos exige, contudo, uma análise mais sofisticada das estruturas organizacionais e dos processos técnicos envolvidos na concepção, no treinamento, na implementação e na operação dos sistemas de inteligência artificial. Não se trata de imputar automaticamente o resultado ao elo mais próximo do dano, mas de identificar, à luz das circunstâncias do caso concreto, em que ponto da cadeia decisória ocorreu a violação relevante de um dever de cuidado ou de controle.

Nesse cenário, ganha relevo a ideia de domínio humano sobre o risco, ainda que exercido de forma mediata, distribuída ou institucional. A responsabilidade penal passa a ser reconstruída não apenas a partir da execução imediata do ato, mas também das decisões de projeto, das escolhas de parametrização, das políticas de uso e das falhas de supervisão que tornaram possível a produção do resultado lesivo.

Por essa razão, a emergência de decisões automatizadas não representa uma crise terminal da responsabilidade penal, mas um desafio de adaptação interpretativa de suas categorias clássicas. A tecnologia pode mediar ou influenciar decisões; a responsabilidade penal, entretanto, permanece necessariamente humana e deve continuar

fundada na possibilidade de controle, na exigibilidade de conduta diversa e na censura pessoal ao agente.

Assim, a reconstrução dogmática da imputação em contextos de inteligência artificial deve buscar preservar o núcleo normativo do Direito Penal, evitando tanto a tentação de atribuir responsabilidade a artefatos técnicos quanto o risco oposto de criar zonas de irresponsabilidade jurídica em razão da complexidade tecnológica.

CONSIDERAÇÕES FINAIS

O presente artigo teve como objetivo analisar os limites e os desafios da responsabilidade penal e da imputação penal em contextos nos quais decisões relevantes são mediadas ou produzidas por sistemas automatizados de inteligência artificial. Partindo dos fundamentos dogmáticos do Direito Penal e do ordenamento jurídico brasileiro, buscou-se demonstrar que a emergência dessas tecnologias não autoriza a superação dos princípios estruturantes da responsabilidade penal, mas exige uma adaptação criteriosa de suas categorias interpretativas.

Ao longo do trabalho, evidenciou-se que os sistemas algorítmicos, embora capazes de produzir efeitos socialmente relevantes e potencialmente lesivos, não podem ser considerados sujeitos de imputação penal. A ausência de consciência, vontade e capacidade de culpabilidade impede que tais sistemas figurem como destinatários legítimos da sanção penal, o que torna juridicamente inadequada qualquer tentativa de lhes atribuir responsabilidade direta. O problema jurídico desloca-se, assim, para a reconstrução da imputação em relação aos sujeitos humanos envolvidos nas cadeias de concepção, implementação e operação dessas tecnologias.

Demonstrou-se, ainda, que a opacidade algorítmica e a fragmentação das cadeias decisórias constituem obstáculos relevantes à aplicação dos modelos clássicos de imputação, sobretudo no que se refere à identificação da conduta juridicamente relevante, do nexo causal e do elemento subjetivo. Todavia, tais dificuldades não configuram razão suficiente para a renúncia aos pressupostos fundamentais do Direito Penal, mas indicam

a necessidade de um refinamento interpretativo que permita lidar com ambientes tecnicamente mediados e organizacionalmente complexos.

A análise dos sistemas de decisão automatizada evidenciou, em plano empírico, que os riscos tecnológicos contemporâneos tendem a diluir a autoria e a dificultar a individualização da responsabilidade. Ainda assim, reforçou-se a tese de que a resposta jurídico-penal adequada não consiste em personificar a tecnologia, mas em identificar, no interior das estruturas técnicas e organizacionais, as falhas relevantes nos deveres de cuidado, de supervisão e de intervenção imputáveis a agentes humanos determinados.

Em resumo, as decisões automatizadas mediadas por IA não fazem a máquina o centro de responsabilidade penal. Demandam, no entanto, a evolução dos critérios de imputação dos agentes humanos e das organizações que estão em domínio do risco criado. A questão jurídica central não é mais “se” a IA pode ser responsabilizada, mas “como” distribuir entre os decisores, desenvolvedores e instituições usuárias os deveres de cuidado, controle e supervisão (Ortiz, 2023; Sachoulidou, 2024).

Nesse contexto, destacou-se o papel central da transparência, da explicabilidade e da intervenção humana como condições de legitimidade das decisões automatizadas e como pressupostos estruturais para a própria viabilidade da responsabilidade penal. Sem um mínimo de controle humano efetivo e sem a possibilidade de reconstrução racional dos processos decisórios, a censura penal perde seu fundamento normativo e sua função garantista.

Conclui-se, portanto, que a inteligência artificial e os sistemas automatizados de decisão não impõem a superação do modelo penal vigente, mas reclamam sua reafirmação em chave adaptativa.

A tecnologia pode mediar decisões; a responsabilidade penal, contudo, deve permanecer necessariamente ancorada na conduta humana, na possibilidade de controle e na exigibilidade de comportamento conforme o Direito, sob pena de se comprometer a racionalidade, a legitimidade e os próprios limites do poder punitivo no Estado Democrático de Direito. Essa constatação se origina do fato de que os sistemas de inteligência artificial atuam como ferramentas inseridas em organizações humanas. Nesta hipótese, as decisões de projeto, parametrização, políticas de utilização e de supervisão

continuam condicionando o resultado. Portanto, a discussão se desloca da busca por uma subjetividade tecnológica para o aperfeiçoamento dos deveres de cuidado e controle a deles atribuídos (a quem pertence) o nome, respectivamente, desenvolvedores, fornecedores, gestores e operadores dos sistemas.

Adverte-se que tais conclusões devem ser examinadas sob a perspectiva de suas limitações. Fundamentalmente, a de que este estudo é dogmático e teórico, uma vez que não contém a coleta e análise de dados empíricos próprios, o que impossibilita a verificação direta das hipóteses e prejudica a avaliação da atuação institucional em casos concretos.

Essa limitação, no entanto, abre caminhos para novas pesquisas, ao menos, em duas direções. Em primeiro lugar, estudos dogmáticos comparados podem aprimorar os critérios clássicos de imputação penal (previsibilidade, evitabilidade, domínio do risco e culpa) para cenários mediados por sistemas de inteligência artificial. Bem como, pesquisas interdisciplinares entre Direito e Ciência da Computação podem analisar, por meio de testes das ferramentas de IA explicável e da análise dos mecanismos de revisão humana previstos na LGPD, em que medida essas ferramentas podem ser suficientes para viabilizar a reconstrução da conduta e da culpa no contexto da decisão automatizada.

Essas investigações, por sua vez, animam outras pesquisas empíricas, tais como estudos de caso forenses envolvendo decisões automatizadas na saúde, no trânsito ou na segurança pública. Podem também fomentar auditorias organizacionais em instituições que utilizam IA a fim de se investigar, exatamente, onde podem se encontrar possíveis pontos de melhorias de supervisão que podem ser relevantes para a responsabilização penal.

REFERÊNCIAS

ALVES, Natália Gomes. **Culpabilidade penal no âmbito da inteligência artificial: a responsabilização da pessoa jurídica no uso dos veículos autônomos.** *Virtuajus*, Belo Horizonte, v. 6, n. 11, p. 360–377, 2022. Disponível em: <https://periodicos.pucminas.br/virtuajus/article/view/28069>. Acesso em: 19 jan. 2026.

ÁVILA, G. N. de; CORAZZA, T. A. M. A hiperexposição pessoal e seus reflexos nos direitos da personalidade: necessidade de uma tutela transversal do direito à privacidade,

com enfoque no âmbito penal. **Juris Poiesis**, Rio de Janeiro, n. 37, p. 144-174, 27 abr. 2022. Disponível

em: <https://mestradoedoutoradoestacio.periodicoscientificos.com.br/index.php/jurispoiesis/article/view/10540/47968258>. Acesso em: 30 jan. 2026.

BRASIL, D. R.; BENTO, L. A. O direito fundamental à privacidade no contexto da Lei Geral de Proteção de Dados. **Revista de Direito Contemporâneo UNIDEP**, v. 1, n. 2, p. 7-24, 2023. Disponível em: <https://periodicos.uniddep.edu.br/rdc-u/article/view/162>. Acesso em: 20 set. 2023.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. *Diário Oficial da União*, Brasília, DF, 5 out. 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm. Acesso em: 19 jan. 2026.

BRASIL. Decreto-Lei nº 2.848, de 7 de dezembro de 1940. Código Penal. *Diário Oficial da União*, Rio de Janeiro, RJ, 31 dez. 1940. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del2848.htm. Acesso em: 19 jan. 2026.

BRASIL. Lei nº 12.965, de 23 de abril de 2014. Marco Civil da Internet. **Diário Oficial da União**, Brasília, DF, 23 abr. 2014. Disponível em: http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/112965.htm. Acesso em: 19 set. 2023.

BRASIL. Lei nº 13.709, de 14 de agosto de 2018. Lei Geral de Proteção de Dados Pessoais (LGPD). **Diário Oficial da União**, Brasília, DF, 15 ago. 2018. Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/113709.htm. Acesso em: 19 jan. 2026.

CARVALHO, Salo de; GIONGO, Pedro. Inteligência artificial e responsabilidade penal no contexto dos veículos autônomos. **Lex Medicinæ: Revista Portuguesa de Direito da Saúde**, Coimbra, ano 17, n. 34, p. 37-63, 2020. Disponível em: <https://revistas.usc.gal/index.php/epc/article/view/8902>. Acesso em: 19 jan. 2026.

JATOBÁ, G. C. Responsabilidade civil e a utilização de robôs de inteligência artificial na medicina. **Revista Jurídica da Seção Judiciária de Pernambuco**, Recife, 2021. Disponível em: <https://revista.jfpe.jus.br/index.php/RJSJPE/article/viewFile/244/227>. Acesso em: 30 jan. 2026

JUNQUEIRA, Felipe Xavier Januário. Inteligência artificial e responsabilidade penal no setor da medicina. **Lex Medicinæ: Revista Portuguesa de Direito da Saúde**, Coimbra, ano 17, n. 34, p. 37-63, 2020. Disponível em: <https://revistacienciasinacipe.fgr.org.mx/index.php/02/article/view/670>. Acesso em: 19 jan. 2026.

NUNES, D. J. C.; ANDRADE, O. M. O uso da inteligência artificial explicável enquanto ferramenta para compreender decisões automatizadas. **Revista Eletrônica do Curso de Direito da UFSM**, Santa Maria, v. 18, n. 1, e69329, 2023. DOI: <https://doi.org/10.5902/1981369469329>. Disponível em: <https://periodicos.ufsm.br/revistadireito/article/view/69329>. Acesso em: 25 set. 2023.

OLIVEIRA, Milla Christie Pereira de; CURY, Letícia Vivianne Miranda. Inteligência artificial e responsabilidade penal: quem responde por crimes praticados com algoritmos? **Revista ft.** 4 nov. 2025. Disponível em: <https://revistaft.com.br/inteligencia-artificial-e-responsabilidade-penal-quem-responde-por-crimes-praticados-co%D0%BC-algoritmos/>. Acesso em: 30 jan. 2026

ORTIZ, Mariana Fleming Soares. **Inteligência artificial e responsabilidade penal.** 2023. 115 f. Dissertação (Mestrado em Direito) – Universidade Nove de Julho, São Paulo, 2023. Disponível em: <https://bibliotecatede.uninove.br/bitstream/tede/3426/2/Mariana%20Fleming%20Soares%20Ortiz.pdf>. Acesso em: 19 jan. 2026.

SACHOULIDOU, Aikaterini. AI systems and criminal liability. **Oslo Law Review**, Oslo, v. 11, n. 1, p. 1–24, 2024. Disponível em: <https://www.scup.com/doi/10.18261/olr.11.1.3>. Acesso em: 19 jan. 2026.

SAPIR, Alan. Quando o algoritmo comete o crime. **JOTA**, [S. l.], [s. d.]. Disponível em: <https://www.jota.info/opiniao-e-analise/colunas/regulacao-e-novas-tecnologias/quando-o-algoritmo-comete-o-crime>. Acesso em: 30 jan. 2026.

WATANABE DE MORAIS, R. Programações podem ser punidas? Responsabilidade penal em decisões tomadas pela inteligência artificial. **Revista Científica do CPJM**, [S. l.], v. 2, n. 07, p. 130–148, 2023. Disponível em: <https://rcpjm.cpj.m.uerj.br/revista/article/view/217>. Acesso em: 30 jan. 2026.

WIMMER, Miriam; DONEDA, Danilo. “Falhas de IA” e a intervenção humana em decisões automatizadas: parâmetros para a legitimação pela humanização. **Revista de Direito Público**, Brasília, v. 18, n. 100, p. 1–30, 2025. DOI: 10.11117/rdp.v18i100.6119. Disponível em: <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/6119>. Acesso em: 19 jan. 2026.

Submissão: outubro de 2025. Aceite: novembro de 2025. Publicação: fevereiro de 2026.