MENS REA DIGITAL E A CULPABILIDADE ALGORÍTMICA: A RESPONSABILIDADE PENAL NA ERA DA INTELIGÊNCIA ARTIFICIAL DELITIVA

DIGITAL MENS REA AND ALGORITHMIC CULPABILITY: CRIMINAL LIABILITY IN THE ERA OF CRIMINAL ARTIFICIAL INTELLIGENCE

Ronny Max Machado[1]
Marcos Paulo Silva de Carvalho[2]

RESUMO: O presente artigo, por meio da metodologia da revisão bibliográfica especializada, analisará a possibilidade e os limites da responsabilidade penal no contexto da inteligência artificial, introduzindo os conceitos de mens rea digital e de culpabilidade algorítmica. Examina-se a impossibilidade de atribuir dolo ou culpa a sistemas autônomos destituídos de consciência e vontade, bem como a necessidade de reconfigurar a imputação penal humana frente à automação tecnológica. A partir do princípio nulla poena sine culpa e da teoria garantista, defende-se que a culpabilidade permanece atributo exclusivo da pessoa humana, recaindo sobre programadores, desenvolvedores e operadores que, por negligência, imperícia ou omissão, criam riscos juridicamente relevantes. Também se discute a interface entre a Lei Geral de Proteção de Dados (Lei nº 13.709/2018) e o Direito Penal, ressaltando a importância da ética digital e da prestação de contas (accountability) como elementos preventivos de responsabilidade. Por fim, propõe-se a consolidação do conceito de "culpa pela criação de risco algorítmico" e a formulação de critérios objetivos de imputação penal para condutas envolvendo inteligência artificial, em conformidade com os princípios da legalidade, da dignidade da pessoa humana e da segurança jurídica.

Palavras-chave: Inteligência Artificial; Responsabilidade Penal; Culpabilidade Algorítmica; Mens Rea Digital.

ABSTRACT: This article, using a specialized literature review methodology, analyzes the possibilities and limits of criminal liability in the context of artificial intelligence, introducing the concepts of digital mens rea and algorithmic culpability. It examines the impossibility of attributing intent or fault to autonomous systems devoid of consciousness and will, as well as the need to reconfigure human criminal liability in the face of technological automation. Based on the principle of nulla poena sine culpa and the theory of guarantees, it argues that culpability remains an exclusive attribute of the human person, falling on programmers, developers, and operators who, through negligence, incompetence, or omission, create legally significant risks. It also discusses the interface between the General Data Protection Law (Law No. 13,709/2018) and Criminal Law, highlighting the importance of digital ethics and accountability as elements that prevent liability. Finally, it is proposed to consolidate the concept of "guilt for creating algorithmic risk" and to formulate objective criteria for criminal prosecution for conduct involving artificial intelligence, in accordance with the principles of legality, human dignity and legal

[1] Mestre em Direito da Sociedade da Informação pelo Centro Universitário das Faculdades Metropolitanas Unidas. Graduado em Direito pela Universidade São Judas Tadeu. Especialista em Direito Civil e Processo Civil pela Escola Paulista de Direito. Especialista em Direito Constitucional e Direito Administrativo pela Escola Paulista de Direito. Especialista em Direito Ambiental Empresarial pelo Centro Universitário das Faculdades Metropolitanas Unidas.
[2] Pós-graduado em Processo Penal pelo Instituto de Direito Penal Econômico Europeu (IDPEE) da Universidade de Coimbra, em parceria com o Instituto Brasileiro de Ciências Criminais (IBCCRIM), e Pós-graduado em Direito Penal Econômico pela Fundação Getúlio Vargas (FGV). Pós-graduando em Data Protection Officer (DPO) – Privacidade e Proteção de Dados pela Escola Superior de Advocacia (ESA), dedicando-se à adequação corporativa à LGPD e ao compliance digital.

certainty.

Keywords: Artificial Intelligence; Criminal Liability; Algorithmic Culpability; Digital Mens Rea.

INTRODUÇÃO

O avanço da inteligência artificial (IA) trouxe novos desafios ao Direito Penal, especialmente quanto à identificação da culpabilidade e da intenção criminosa (mens rea) em ações praticadas por sistemas autônomos. O conceito de mens rea digital busca adaptar a ideia tradicional de dolo e culpa ao contexto tecnológico, no qual algoritmos podem causar danos sem consciência ou vontade própria.

A culpabilidade algorítmica surge como instrumento para atribuir responsabilidade penal aos desenvolvedores, operadores e gestores de IA considerando fatores como previsibilidade do dano, controle humano e diligência técnica. Assim, ainda que a máquina não possa ser punida, o ser humano responsável pela criação ou supervisão do sistema pode responder penalmente, conforme os princípios constitucionais da Culpabilidade e da pessoalidade da pena previsto no sistema jurídico brasileiro.

Nessa conjuntura, coloca-se a seguinte indagação: pode haver culpabilidade sem consciência? em outras palavras, seria possível imputar a responsabilidade penal a uma inteligência artificial desprovido de vontade, emoção ou até mesmo juízo de valor?

A presente pesquisa parte do pressuposto de que não há culpabilidade sem consciência, pois, a culpabilidade é uma categoria estritamente humana, alicerçada na capacidade de autodeterminação ética e na possibilidade de agir de modo diverso. Assim, os sistemas de inteligência artificial, malgrado sua autonomia decisória, não se qualificam como sujeitos de imputação penal direta. Entretanto, o avanço da culpabilidade algorítmica impõe a redefinição dos contornos da responsabilidade penal humana face à IA, especialmente diante de condutas imprudentes, omissivas ou tecnicamente falhas que exponencializam riscos sociais mediante sistemas automatizados.

No futuro, o Direito Penal precisará evoluir para lidar com os crimes praticados por inteligência artificial (IA delitiva), estabelecendo critérios objetivos de imputação sem violar os princípios da legalidade, da culpabilidade e da dignidade da pessoa humana. Ao examinar a IA delitiva, é fundamental classificá-la sob a ótica da participação da tecnologia no ilícito penal. O computador – ou, neste caso, o sistema algorítmico – pode ser o "alvo ou a ferramenta para a prática do crime" ("the computer must be either the target or tool for committing the crime") ou, ainda, "incidental ao crime, significando que o computador está envolvido na prática do crime ou está sendo usado meramente como um dispositivo de armazenamento" ("the computer may also be incidental to a crime, meaning that the

computer is either involved in the commission of a crime, or the computer is being used merely as a storage device") (Holt; Bossler; Seigfried-Spellar, p. 508, tradução nossa). A IA delitiva, ao transcender o papel de mera ferramenta passiva, força o Direito Penal a reavaliar sua resposta.

Em síntese, a mens rea digital e a culpabilidade algorítmica representam a tentativa de equilibrar inovação tecnológica e responsabilidade jurídica. Isso é crucial porque, embora a intenção criminosa seja atemporal, guiada por motivações humana como "a ganância, a luxúria, o poder, a vingança, a aventura e o desejo de pegar o 'fruto proibido'" ("greed, lust, power, revenge, adventure, and the desire to take 'forbidden fruit'"), "o elemento de novidade reside na capacidade sem precedentes da tecnologia de facilitar a ação sobre essas motivações" ("The element of novelty resides in the unprecedented capacity of technology to facilitate acting on these motivations") (Holt; Bossler; Seigfried-Spellar, p. 440, tradução nossa).

Desta forma, a redefinição da responsabilidade humana é a via indispensável para assegurar que a inteligência artificial opere dentro dos limites éticos e legais do Estado Democrático de Direito.

1 A *MENS REA* E SUA TRANSPOSIÇÃO AO CONTEXTO DIGITAL

Tradicionalmente, o Direito Penal brasileiro se estrutura sobre o tripé do crime: fato típico, antijurídico e culpável, conforme o Decreto-Lei nº 2.848/1940 (Código Penal). O dolo e a culpa são elementos centrais da culpabilidade, expressando a consciência e a vontade de realizar a conduta criminosa (art. 18 do Código Penal). No entanto, quando o agente é uma entidade digital — um algoritmo ou um sistema de IA —, a presença de vontade e consciência se torna juridicamente incerta.

O conceito de mens rea digital surge, assim, como tentativa de compreender a intencionalidade subjacente às ações praticadas por sistemas autônomos. A dogmática penal, mesmo ao lidar com o cibercrime, insiste na prova da intenção humana: a legislação estrangeira, ao incriminar o acesso não autorizado a sistemas, por exemplo, exige que o agente atue "intencionalmente (knowingly) ou entregando intencionalmente (willfully) essa informação a outra pessoa não autorizada a recebê-la" ("knowingly accessing a computer [...] and willfully deliver that information to another person not entitled to receive it") (Holt; Bossler; Seigfried-Spellar, 2018, p. 119, tradução nossa). Esta persistência em requerer a vontade consciente reforça que a mens rea deve ser adaptada, mas sempre imputada ao ser humano responsável. Ainda que o algoritmo não possua consciência moral, suas decisões podem refletir padrões de comportamento previsíveis ou

negligentes por parte de seus programadores ou operadores humanos. Nesse cenário, a responsabilidade desloca-se do ente digital para aqueles que conceberam, treinaram ou supervisionaram a IA, aplicando-se o princípio de que não há crime sem culpabilidade (nulla poena sine culpa), previsto implicitamente no art. 5º, inciso XLV, da Constituição Federal, que consagra a responsabilidade pessoal e subjetiva.

2 CULPABILIDADE ALGORÍTMICA E A RESPONSABILIDADE PENAL DOS DESENVOLVEDORES

A culpabilidade algorítmica representa o esforço doutrinário de mensurar a responsabilidade penal humana diante de decisões tomadas por sistemas de inteligência artificial (IA). Trata-se de uma reconstrução teórica da culpabilidade voltada a contextos em que a autonomia técnica das máquinas interfere na previsão e no controle humano dos resultados.

Segundo Frazão e Mulholland (2019, p. 6), a inteligência artificial pode ser compreendida como "*todo sistema computacional que simula a capacidade humana de raciocinar e resolver problemas, por meio de tomada de decisões baseadas em análises probabilísticas*". Essa definição evidencia que, embora a IA execute tarefas cognitivamente complexas, sua atuação permanece restrita a padrões estatísticos e matemáticos, destituídos de consciência moral, o que inviabiliza a atribuição direta de dolo ou culpa à máquina.

A imputação penal, portanto, dependerá da análise de três elementos fundamentais: (i) a previsibilidade do dano, (ii) o grau de controle humano sobre o sistema e (iii) a diligência técnica empregada no desenvolvimento e supervisão do algoritmo. Assim, se um desenvolvedor ou operador deixa de implementar salvaguardas éticas e técnicas capazes de evitar resultados lesivos, sua omissão poderá configurar culpa consciente, nos termos do art. 18, II, do Código Penal.

No plano teórico, Ferrajoli (2002, p. 390-393) explica que o modelo garantista se estrutura sobre o axioma *nulla actio sine culpa*, segundo o qual "*nenhum fato ou comportamento humano é valorado como ação se não é fruto de uma decisão; consequentemente, não pode ser castigado, nem sequer proibido, se não é intencional, isto é, realizado com consciência e vontade por uma pessoa capaz de compreender e de querer*". Esse princípio reafirma que a culpabilidade é condição ética e jurídica da pena, vinculada à consciência e à voluntariedade humanas, excluindo, portanto, qualquer hipótese de punição de entes destituídos de razão.

Por outro lado, a autonomia crescente das máquinas desafia diretamente o

princípio da responsabilidade subjetiva, pilar do Direito Penal contemporâneo. Conforme o art. 5º, II, da Constituição Federal, "ninguém será obrigado a fazer ou deixar de fazer alguma coisa senão em virtude de lei". Nesse sentido, o próprio Ferrajoli (2002, p. 393) adverte que o sistema penal garantista "*veta a responsabilidade impessoal, objetiva ou não culpável*", reafirmando que somente o autor consciente e voluntário de um injusto pode ser legitimamente punido.

Essa leitura é reforçada por Ferré Olivé et al. (2017, p. 431), ao sustentarem que "*a culpabilidade converte-se no principal pressuposto da pena. Nesse contexto, mantém plena vigência o aforismo: nulla poena sine culpa*". Assim, mesmo em face de sistemas inteligentes, o juízo de reprovação penal permanece indissociável da vontade e da racionalidade humanas, o que torna inviável a punição de entes artificiais.

A culpabilidade algorítmica não representa uma nova forma de punição de máquinas, mas sim um avanço teórico voltado à compreensão da responsabilidade penal humana diante da automação. A inteligência artificial, por carecer de consciência, vontade e discernimento moral, não pode ser considerada sujeito ativo de crime, pois lhe faltam os elementos essenciais da culpabilidade.

Contudo, a expansão da autonomia dos sistemas tecnológicos impõe ao Direito Penal a tarefa de redefinir os limites da imputação humana. O desenvolvimento, a programação e a operação de algoritmos passam a exigir um padrão mais elevado de diligência técnica, ética e preventiva. Sempre que uma decisão automatizada causar dano previsível em razão de falha humana, omissão de segurança ou descuido na supervisão, poderá ser configurada a responsabilidade penal do agente envolvido.

Nesse contexto, os programadores, desenvolvedores e operadores assumem posição central: são eles que detêm o domínio técnico e o controle potencial sobre o risco criado. A ausência de medidas adequadas de segurança, de transparência algorítmica ou de mecanismos de correção pode caracterizar negligência ou culpa consciente, permitindo a imputação penal em situações de dano evitável.

Assim, a culpabilidade algorítmica deve ser entendida como uma extensão do dever de cuidado no âmbito tecnológico. O foco da punição permanece sobre o ser humano que cria, treina ou supervisiona a máquina, e não sobre a máquina em si. Essa perspectiva assegura a coerência entre o progresso tecnológico e os princípios fundamentais do Direito Penal, preservando o postulado de que não há pena sem culpa nem responsabilidade sem consciência.

3 A PROTEÇÃO DE DADOS E A ÉTICA PENAL DIGITAL

A consolidação da inteligência artificial como instrumento de decisão e análise de dados trouxe à tona uma nova dimensão da responsabilidade penal: a ética digital. Nesse contexto, a Lei nº 13.709/2018 — Lei Geral de Proteção de Dados (LGPD), inspirada no Regulamento Geral de Proteção de Dados da União Europeia (GDPR), desempenha papel fundamental ao estabelecer o princípio da responsabilidade e da prestação de contas (accountability). Esse princípio impõe aos controladores e operadores de dados o dever de demonstrar a adoção de medidas técnicas e administrativas eficazes para garantir a segurança, a integridade e a privacidade das informações tratadas.

Tal lógica normativa aproxima-se diretamente da concepção de culpabilidade algorítmica, ao reconhecer que o simples manejo de sistemas automatizados envolve riscos socialmente relevantes e que o agente humano deve responder pela criação, supervisão e controle desses riscos. Assim, tanto no âmbito da proteção de dados quanto no Direito Penal, a responsabilidade não decorre apenas do resultado, mas também da conduta preventiva e diligente exigida de quem atua em contextos tecnológicos.

Nos chamados casos de IA delitiva — situações em que sistemas inteligentes produzem efeitos criminosos, como fraudes financeiras automatizadas, deepfakes difamatórios, ataques cibernéticos autônomos ou manipulação ilícita de bases de dados — , o Direito Penal é desafiado a mapear toda a cadeia de imputação humana envolvida. Isso significa considerar desde os desenvolvedores e programadores até as instituições que financiam, supervisionam ou lucram com a aplicação indevida de algoritmos.

Contudo, essa expansão de responsabilidade deve manter-se ancorada nos princípios constitucionais da legalidade e da dignidade da pessoa humana (art. 1º, III, da Constituição Federal).

A dignidade, enquanto fundamento da República, impede a despersonalização da culpa e reafirma que a sanção penal só pode incidir sobre indivíduos dotados de consciência moral e liberdade de escolha. O controle ético sobre a tecnologia, portanto, não é apenas uma exigência de segurança jurídica, mas também uma garantia civilizatória, que preserva o ser humano como centro e limite da atuação penal.

Dessa forma, a convergência entre o Direito Penal e a proteção de dados revela a necessidade de um modelo normativo capaz de equilibrar inovação tecnológica, responsabilidade e direitos fundamentais. A ética penal digital deve servir como filtro para o uso legítimo da inteligência artificial, assegurando que o avanço tecnológico se submeta aos valores de prudência, transparência e respeito à dignidade humana — pilares indispensáveis à justiça penal no século XXI.

4 PERSPECTIVAS FUTURAS E O DESAFIO DA IMPUTAÇÃO PENAL AUTOMATIZADA

A consolidação da culpabilidade algorítmica como parâmetro de imputação humana no contexto da inteligência artificial não encerra a discussão, mas inaugura um novo campo de desafios para o Direito Penal contemporâneo.

Se, no presente, a imputação recai sobre os agentes que concebem, operam ou supervisionam sistemas inteligentes, o futuro tende a exigir uma reformulação normativa mais precisa para lidar com níveis crescentes de autonomia tecnológica.

Nesse cenário, emerge o conceito de culpa pela criação de risco algorítmico, expressão que designa a responsabilidade daquele que introduz ou mantém em funcionamento um sistema cuja operação autônoma pode gerar resultados lesivos previsíveis. Trata-se de uma evolução natural da ideia de culpabilidade subjetiva, adaptada a um contexto em que o controle humano é mediado por estruturas de decisão automatizadas.

Essa nova dimensão da culpa não rompe com os princípios tradicionais do Direito Penal, mas os atualiza à luz das transformações tecnológicas. O dever de cuidado — antes restrito a atos diretos de vigilância, escolha ou execução — passa a abranger também a previsibilidade técnica e o domínio informacional. Em outras palavras, quem cria ou utiliza algoritmos capazes de afetar bens jurídicos relevantes deve antecipar os riscos decorrentes de seu comportamento, sob pena de incorrer em responsabilidade penal por omissão ou negligência.

O legislador penal do futuro será desafiado a positivar critérios objetivos de imputação para condutas relacionadas à inteligência artificial, preservando os pilares constitucionais da legalidade (art. 1º do Código Penal) e da anterioridade penal (art. 5º, XXXIX, da Constituição Federal). A normatização deverá definir parâmetros de culpa, dolo e risco tecnológico de modo a garantir previsibilidade jurídica e evitar punições arbitrárias baseadas em interpretações extensivas.

Enquanto esse marco regulatório não se consolida, os tribunais terão papel fundamental na adaptação da dogmática penal, aplicando, por analogia, os princípios clássicos da imputação subjetiva, da culpa in vigilando e da culpa in eligendo aos agentes humanos envolvidos em sistemas de IA delitiva. Tais fundamentos asseguram que a inovação não se converta em impunidade, nem em responsabilização automática, mantendo o equilíbrio entre desenvolvimento tecnológico e tutela penal.

Portanto, o futuro da imputação penal automatizada exigirá um Direito Penal tecnicamente sensível e eticamente responsável, capaz de distinguir a autonomia funcional

da máquina da responsabilidade consciente do ser humano. O desafio não é apenas normativo, mas filosófico: garantir que a expansão da inteligência artificial se submeta ao princípio fundamental de que a culpa — e, portanto, a pena — permanece atributo exclusivo da pessoa humana.

CONCLUSÃO

A emergência da inteligência artificial delitiva impõe ao Direito Penal contemporâneo um novo horizonte de reflexão sobre os limites da responsabilidade e da imputação. A transposição da mens rea tradicional para o contexto digital revela que, embora os sistemas algorítmicos possam agir de modo autônomo, eles carecem dos elementos essenciais da culpabilidade — consciência, vontade e discernimento moral. Assim, permanece insustentável, sob o ponto de vista jurídico, qualquer pretensão de atribuir responsabilidade penal direta a entes artificiais.

A culpabilidade algorítmica deve ser compreendida como um instrumento teórico voltado à identificação e mensuração das falhas humanas que permitem a ocorrência de resultados lesivos em contextos de automação. Programadores, desenvolvedores, operadores e gestores que detêm o domínio técnico e a capacidade de controle sobre sistemas autônomos assumem posição central no juízo de imputação, especialmente diante da previsibilidade dos riscos e da omissão em adotar medidas de segurança, transparência e supervisão adequada.

Nesse sentido, a responsabilidade penal na era da inteligência artificial representa uma extensão do dever de cuidado, aplicável às novas formas de risco criadas por tecnologias autônomas. O Direito Penal deve evoluir para incorporar parâmetros de diligência técnica e ética compatíveis com a complexidade dos sistemas algorítmicos, sem recorrer à responsabilização objetiva ou à ficção de vontade nas máquinas.

O futuro da imputação penal dependerá da capacidade do ordenamento jurídico de estabelecer critérios objetivos para distinguir o erro humano da falha algorítmica, preservando os princípios da legalidade, da previsibilidade e da proporcionalidade. A mens rea digital e a culpabilidade algorítmica não buscam substituir a consciência humana, mas reafirmar seu papel central como condição indispensável para a existência da culpa e, portanto, da pena.

A consolidação do conceito de “culpa pela criação de risco algorítmico” surge, assim, como caminho possível para responsabilizar aqueles que, por negligência técnica ou omissão consciente, colocam em funcionamento sistemas capazes de causar danos previsíveis. O equilíbrio entre inovação tecnológica e responsabilidade penal será o grande

desafio das próximas décadas, impondo ao Direito o dever de garantir que a inteligência artificial permaneça subordinada aos valores éticos, jurídicos e humanos que sustentam o Estado Democrático de Direito.

REFERÊNCIAS

BRASIL. **Código Penal**. Decreto-Lei nº 2.848, de 7 de dezembro de 1940. Disponível em: https://www.planalto.gov.br/ccivil_03/decreto-lei/del2848compilado.htm. Acesso em: 20 out. 2025.

FERRAJOLI, Luigi. **Direito e razão: teoria do garantismo penal**. 3. ed. rev. São Paulo: Revista dos Tribunais, 2002.

FERRÉ OLIVÉ, Juan Carlos; NÚÑEZ PAZ, Miguel Ángel; OLIVEIRA, William Terra de; BRITO, Alexis Couto de. **Direito penal brasileiro: parte geral – princípios fundamentais e sistema.** 2. ed. São Paulo: Saraiva Jur, 2017.

FRAZÃO, Ana; MULHOLLAND, Caitlin S. **Inteligência artificial e direito: ética, regulação e responsabilidade**. Rio de Janeiro: Forense, 2019.

HOLT, Thomas J.; BOSSLER, Adam M.; SEIGFRIED-SPELLAR, Kathryn C. **Cybercrime and digital forensics**: an introduction. 2. ed. London: Routledge, 2018.

JAKOBS, Günther. **Derecho penal**: parte general**.** 2. ed. Madrid: Marcial Pons, 1997.

SILVA SÁNCHEZ, Jesús-María. **La expansión del Derecho penal**: aspectos de la política criminal en las sociedades postindustriales. 2. ed. Madrid: Civitas, 2011.

ZAFFARONI, Eugenio Raúl; José Henrique Pierangeli. **Manual de Direito Penal Brasileiro**: Parte Geral. 5ª ed. rev. e atual. São Paulo: Editora Revista dos Tribunais. 2004.