

PROVA PENAL E INTELIGÊNCIA ARTIFICIAL: limites epistemológicos e garantias processuais na valoração de interferências algorítmicas

CRIMINAL EVIDENCE AND ARTIFICIAL INTELLIGENCE: epistemological limits and procedural guarantees in the evaluation of algorithmic inferences

Isis de Almeida Gomes¹

Recebido/Received: 21.03.2026/Mar 21th, 2026

Aprovado/Approved: 27.04.2026/Apr 27th, 2026

RESUMO: Este artigo examina a valoração da prova penal produzida por inteligência artificial (IA), sob uma perspectiva crítica que destaca seus limites epistemológicos, as garantias processuais e os riscos da automação decisória. Analisa-se a crescente incorporação de sistemas automatizados na investigação criminal e na formação do convencimento judicial, frequentemente sem transparência, auditabilidade e supervisão humana adequadas, o que pode comprometer o contraditório, a ampla defesa e o devido processo legal. Por meio de abordagem qualitativa, com análise doutrinária, jurisprudencial e comparada, sustenta-se que, embora a IA possa atuar como instrumento auxiliar, seu uso acrítico fragiliza a legitimidade da decisão penal. Conclui-se que sua adoção deve ser subsidiária, preservando a centralidade do juízo humano e a proteção dos direitos fundamentais.

PALAVRAS-CHAVE: inteligência artificial; processo penal; prova penal; garantias processuais; automação decisória.

ABSTRACT: This article examines the evidentiary assessment of criminal evidence produced by artificial intelligence (AI), adopting a critical perspective that highlights its epistemological limits, procedural guarantees, and the risks of automated decision-making. It analyzes the growing incorporation of automated systems into criminal investigations and judicial reasoning, often without adequate transparency, auditability, and human oversight, which may undermine adversarial proceedings, the right to a full defense, and due process of law. Using a qualitative approach based on doctrinal, jurisprudential, and comparative analysis, the study argues that, although AI may function as an auxiliary tool, its uncritical use weakens the legitimacy of criminal decisions. It concludes that its adoption must be strictly subsidiary, preserving the centrality of human judicial reasoning and the protection of fundamental rights.

¹ Pós-graduada em Direito Penal e Criminologia pela Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS). Bacharela em Direito pelo Centro Universitário Jorge Amado. Currículo Lattes: <http://lattes.cnpq.br/9045833955539686>. E-mail: isis.algomes@hotmail.com

KEYWORDS: artificial intelligence; criminal procedure; criminal evidence; procedural guarantees; automated decision-making.

INTRODUÇÃO

O avanço acelerado das tecnologias digitais, em especial da inteligência artificial (IA), tem produzido impactos profundos sobre as estruturas tradicionais do sistema de justiça criminal. No âmbito do processo penal, observa-se a crescente incorporação de ferramentas algorítmicas voltadas à produção, análise e interpretação de dados probatórios, como sistemas de reconhecimento facial utilizados em segurança pública, ferramentas de análise de dados massivos (big data policial), mecanismos de inferência preditiva de risco criminal e perícias digitais assistidas por algoritmos. Tal fenômeno pode ser compreendido como a inauguração de uma nova etapa do poder punitivo estatal, marcada pela transição de um modelo centrado na atividade humana para um cenário de mediação tecnológica da prova penal.

Nesse contexto, emerge uma questão central: a prova penal, historicamente concebida como resultado de atos humanos passíveis de reconstrução racional, passa a ser produzida, total ou parcialmente, por sistemas automatizados cuja lógica interna é, muitas vezes, opaca, probabilística e inacessível às partes.

A problemática da opacidade algorítmica pode ser compreendida como elemento central desse cenário. Conforme desenvolvido por Frank Pasquale, em *The Black Box Society*, sistemas algorítmicos contemporâneos tendem a operar como “caixas-pretas”, caracterizadas por assimetria informacional, limitação de transparência e dificuldades de controle e responsabilização. Tal característica pode comprometer o escrutínio crítico das inferências produzidas, impactando diretamente a possibilidade de controle jurídico dessas tecnologias no processo penal.

A chamada prova produzida por inteligência artificial tende a desafiar categorias clássicas do processo penal, especialmente no que se refere à sua admissibilidade, confiabilidade e, sobretudo, à sua valoração pelo julgador.

A valoração da prova constitui momento decisivo do processo penal, pois é nela que o juiz, à luz do princípio do livre convencimento motivado, atribui peso, credibilidade e significado aos elementos probatórios constantes dos autos.

Contudo, a introdução de resultados algorítmicos no raciocínio judicial tende a tensionar esse modelo, na medida em que pode induzir à deferência excessiva à tecnologia, podendo levar a substituição da fundamentação racional por uma aceitação acrítica de conclusões apresentadas como tecnicamente superiores ou incontestáveis.

O problema se agrava diante do fenômeno da opacidade algorítmica, isto é, da dificuldade, ou impossibilidade, de compreender os critérios, os dados de treinamento e os processos decisórios internos dos sistemas de inteligência artificial. Tal característica pode comprometer diretamente garantias fundamentais do processo penal democrático, como o contraditório, a ampla defesa e o direito à motivação das decisões judiciais, na medida em que impede o efetivo questionamento da prova e a reconstrução racional de seu percurso formativo.

Além disso, a prova produzida por inteligência artificial não deve ser compreendida como um instrumento neutro, pois está sujeita à reprodução de padrões e distorções presentes nos dados que alimentam os sistemas. Nesse sentido, sua utilização no processo penal envolve implicações que ultrapassam o plano técnico, alcançando dimensões relacionadas à seletividade penal e à própria legitimidade do exercício do poder punitivo.

Diante desse cenário, propõe-se o seguinte problema de pesquisa: é juridicamente legítima a valoração da prova penal produzida por inteligência artificial nos moldes tradicionais do livre convencimento motivado, sem a adoção de critérios reforçados de controle epistêmico e constitucional? Parte-se da hipótese de que a prova produzida por IA não se equipara, em regra, às provas tradicionais, nem deve, em regra, servir como fundamento exclusivo de uma condenação penal, sob o risco de violação às garantias estruturantes do devido processo legal.

O objetivo deste trabalho é analisar criticamente a valoração da prova penal produzida por inteligência artificial, à luz do direito processual penal, da epistemologia jurídica e da criminologia crítica, identificando seus limites, riscos e exigências normativas. Busca-se demonstrar que a utilização legítima dessa modalidade probatória pressupõe a necessidade de transparência, auditabilidade, possibilidade de contraditório técnico e corroboração por outros meios de prova, de modo a mitigar o risco de automação indevida do convencimento judicial e a erosão do garantismo penal.

Para tanto, adota-se uma metodologia de abordagem qualitativa, com base em pesquisa bibliográfica e análise crítica de doutrina nacional e estrangeira, bem como em documentos normativos e estudos interdisciplinares sobre inteligência artificial e justiça criminal. O marco teórico articula contribuições da epistemologia da prova, da criminologia crítica e da teoria das garantias processuais, com foco na análise dos impactos da automação na produção e valoração da prova penal.

Por fim, quanto à estrutura do artigo, inicialmente será analisada a relação entre inteligência artificial e produção de provas no processo penal, com ênfase nos aspectos técnicos e epistemológicos das inferências algorítmicas. Em seguida, serão examinados os limites epistemológicos da prova automatizada, especialmente no que se refere à opacidade dos sistemas, aos vieses estruturais e à natureza probabilística das inferências. Na sequência, serão discutidas as garantias processuais penais frente à automação decisória, com destaque para o contraditório, a ampla defesa, a auditabilidade e o papel do juiz como agente de controle crítico da prova. Posteriormente, serão analisados os riscos da automação decisória no processo penal, com foco na possibilidade de deferência acrítica aos sistemas algorítmicos, na reprodução de vieses estruturais, na opacidade decisória e na fragilização da responsabilização e da imparcialidade judicial. Ao final, serão apresentadas as conclusões do trabalho, com a proposição de parâmetros normativos para a valoração legítima da prova penal produzida por inteligência artificial.

1 INTELIGÊNCIA ARTIFICIAL E PRODUÇÃO DE PROVAS: UM ENSAIO APROFUNDADO

A inteligência artificial (IA) tem emergido como uma das tecnologias mais disruptivas de nosso tempo, impactando não apenas áreas empresariais ou de engenharia, mas também se inserindo progressivamente na investigação criminal, no processamento de provas e na própria dinâmica de formação do convencimento judicial. Ao contrário das provas tradicionais, caracterizadas por fenômenos perceptíveis e verificáveis pelo raciocínio humano, como testemunhos, documentos físicos, perícias laboratoriais e registros fidedignos, as evidências produzidas por IA tendem a consistir em inferências automatizadas extraídas de modelos matemáticos complexos que operam sobre vastos conjuntos de dados. Essa diferença não se

limita ao plano técnico, podendo assumir dimensão epistemológica, na medida em que altera a forma pela qual o direito compreende a verdade material e a legitimidade probatória no processo penal. A literatura especializada tem destacado que a produção de prova por IA não constitui apenas uma nova ferramenta, mas pode ser compreendida como um fenômeno que desafia categorias fundamentais do sistema probatório e demanda reflexão crítica acerca de sua confiabilidade, transparência e valor probatório (Irfan; Bennett; Müller, 2025).

As provas produzidas por IA podem assumir diferentes formatos, que, em geral, podem ser agrupados em três grandes categorias. Em primeiro lugar, há relatórios automatizados, que consistem em documentos gerados por sistemas capazes de sintetizar e interpretar grandes volumes de dados, como o cruzamento de registros de localização, padrões de comunicação digital ou análises de bancos de dados de transações financeiras, com o objetivo de apontar possíveis relações ou tendências. Esses relatórios não se limitam à apresentação de dados brutos, mas podem oferecer interpretações estatísticas e inferências preditivas passíveis de utilização em investigações criminais. Em segundo lugar, algoritmos de IA são amplamente empregados na análise de grandes conjuntos de dados, operando sobre metadados, logs de sistemas, registros telefônicos, redes sociais e demais rastros digitais que, isoladamente, podem não revelar informações relevantes, mas que, quando correlacionados, tendem a sugerir padrões e indícios úteis à persecução penal. Em terceiro lugar, destacam-se as previsões de comportamento, realizadas por sistemas de aprendizagem de máquina (*machine learning*) que, a partir de dados históricos, estimam a probabilidade de ocorrência de determinados eventos. Esses modelos podem deslocar o foco do processo penal de fatos passados para projeções estatísticas, o que tende a suscitar preocupações relevantes sob a ótica do devido processo legal (O'Neil, 2016).

Nesse cenário, mostra-se relevante distinguir, ainda que de forma analítica, diferentes modalidades de sistemas de inteligência artificial potencialmente empregados no âmbito da persecução penal, uma vez que sua funcionalidade e impacto probatório não são homogêneos. Identificam-se, de um lado, sistemas de vigilância automatizada, como tecnologias de reconhecimento facial e análise de vídeo em tempo real, utilizados sobretudo na identificação de suspeitos e na reconstrução de trajetórias comportamentais. De outro lado, verificam-se ferramentas privadas de monitoramento e extração de dados, empregadas por

empresas de tecnologia e provedores de serviços digitais para a coleta, organização e análise de grandes volumes de informações, incluindo metadados, registros de comunicação e interações em ambiente virtual, que podem posteriormente ser utilizados como elementos informativos na investigação criminal. Por fim, destacam-se os sistemas de apoio à decisão judicial e investigativa, baseados em modelos preditivos ou classificatórios, os quais produzem inferências probabilísticas sobre risco, reincidência ou padrões de comportamento, podendo influenciar, direta ou indiretamente, decisões de natureza processual penal. Essa diferenciação permite evidenciar que a expressão “inteligência artificial” não designa um fenômeno unitário, mas um conjunto heterogêneo de tecnologias com distintos graus de intervenção na produção e valoração da prova penal.

O funcionamento desses sistemas automatizados baseia-se em princípios matemáticos e estatísticos que envolvem etapas como coleta de dados, pré-processamento, treinamento do modelo e geração de inferências. Embora tais etapas sejam compreensíveis em nível teórico, sua aplicação prática pode resultar em sistemas altamente complexos e de difícil explicação. Esse fenômeno é frequentemente descrito como “caixa-preta” (*black box*), no qual não se consegue compreender, de forma transparente, como o algoritmo chegou a determinada conclusão, nem mesmo por seus desenvolvedores (O’Neil, 2016). Tal opacidade pode se revelar especialmente problemática no processo penal, em que a transparência metodológica e a possibilidade de contraditório figuram como requisitos essenciais de validade da prova.

Além disso, a natureza dos dados utilizados para treinar esses modelos pode introduzir vieses algorítmicos capazes de reproduzir desigualdades estruturais. Como observa O’Neil, modelos são opiniões incorporadas na matemática (O’Neil, 2016), o que sugere que tais sistemas podem não ser neutros. Nesse sentido, estudos indicam que algoritmos podem gerar impactos discriminatórios ao reproduzirem padrões históricos de desigualdade, uma vez que “o Big Data pode reproduzir padrões existentes de discriminação, herdando vieses presentes nos próprios dados” (Barocas; Selbst, 2016, p. 671).

Essa preocupação encontra certa ressonância na jurisprudência brasileira, ainda que em contextos tecnológicos não estritamente algorítmicos. O Superior Tribunal de Justiça tem consolidado o entendimento de que provas digitais obtidas sem observância de metodologia adequada e sem documentação idônea da cadeia

de custódia tendem a ser consideradas inadmissíveis, por comprometerem sua confiabilidade e dificultarem o exercício do contraditório (Brasil, 2024). Nesse sentido, destaca-se o HC 598.886/SC¹, Rel. Min. Ribeiro Dantas, julgado em 2021, no qual se reconheceu a necessidade de observância rigorosa da cadeia de custódia para a validade da prova digital. Tal orientação pode fornecer base analógica relevante para o tratamento da prova produzida por IA, cuja validade também tende a depender da observância de critérios técnicos transparentes, rastreáveis e verificáveis.

A jurisprudência do Superior Tribunal de Justiça também tem destacado que a validade da prova digital está condicionada à preservação de sua integridade e à possibilidade de verificação de sua origem, como se observa no RHC 131.263/DF². Tal entendimento evidencia que não basta a existência do dado digital, sendo indispensável a demonstração de sua confiabilidade por meio de procedimentos técnicos verificáveis.

A doutrina brasileira reforça essa perspectiva ao sustentar que a cadeia de custódia pode demandar reinterpretação à luz das tecnologias digitais e algorítmicas. Nesse sentido, Sampaio (2025) destaca que:

A preservação da prova, no ambiente automatizado, tende a exigir não apenas a integridade dos dados, mas também a transparência dos métodos empregados na sua análise por sistemas de inteligência artificial, sob o risco de esvaziamento do valor probatório do resultado apresentado.

A comparação entre a prova tradicional e a prova algorítmica sugere a existência de diferenças estruturais relevantes. Enquanto provas clássicas são produzidas por sujeitos identificáveis e passíveis de questionamento em juízo, a prova gerada por IA frequentemente carece de um agente humano capaz de explicar e justificar cada etapa da inferência automatizada. Essa lacuna pode comprometer o contraditório pleno e fragilizar o controle racional da decisão judicial, razão pela qual parte da doutrina internacional sustenta que sistemas algorítmicos devem atuar como instrumentos auxiliares, e não substitutivos, da atividade jurisdicional (Rosen; Citron, 2020).

¹ STJ, HC 598.886/SC, Rel. Min. Ribeiro Dantas, 5ª Turma, julgado em 2020.

² STJ. Recurso Ordinário em Habeas Corpus n. 131.263/DF. Rel. Min. Ribeiro Dantas, 5ª Turma, julgado em 09 jun. 2020, DJe 15 jun. 2020.

Em síntese, a produção probatória por inteligência artificial tende a impor ao processo penal desafios epistemológicos relevantes, relacionados à opacidade dos modelos, aos riscos de vieses estruturais e à natureza probabilística das inferências. A superação desses limites não parece decorrer da rejeição da tecnologia, mas de sua integração crítica ao sistema de garantias processuais, com a adoção de mecanismos de auditabilidade, explicabilidade e possibilidade efetiva de contestação. Nessas condições, a IA poderá contribuir para a justiça penal sem comprometer, ao menos em tese, os princípios constitucionais do contraditório, da ampla defesa e do devido processo legal.

2 LIMITES EPISTEMOLÓGICOS DA PROVA AUTOMATIZADA

A incorporação de sistemas de inteligência artificial na produção probatória penal pode representar uma ruptura relevante no paradigma epistemológico tradicional do processo penal. Se, historicamente, a prova jurídica esteve vinculada à reconstrução de fatos pretéritos por meio de elementos empiricamente verificáveis, a utilização de modelos algorítmicos tende a introduzir um novo regime de produção do conhecimento, baseado em correlações estatísticas, inferências probabilísticas e processamento massivo de dados. Nesse contexto, a prova pode deixar de ser apenas um instrumento de reconstrução fática para assumir também a função de projeção inferencial, o que tende a tensionar os fundamentos da epistemologia jurídica. Nesse sentido, Mireille Hildebrandt destaca que a mediação tecnológica pode redefinir as condições de produção do conhecimento jurídico (Hildebrandt, 2019).

A opacidade algorítmica, nesse contexto, pode ser compreendida não apenas como limitação técnica, mas como estrutura de poder, nos termos propostos por Frank Pasquale. Segundo o autor, a dificuldade de acesso aos critérios internos dos sistemas automatizados pode gerar assimetrias informacionais significativas, nas quais determinados atores detêm controle sobre os processos decisórios enquanto outros permanecem impossibilitados de compreendê-los ou contestá-los. Tal assimetria pode ser especialmente sensível no processo penal, em que a paridade de armas e o contraditório dependem da possibilidade de acesso e compreensão dos elementos que fundamentam a prova.

Diante desse cenário, sustenta-se que a prova produzida por inteligência artificial pode apresentar limites epistemológicos estruturais que dificultam sua valoração autônoma no processo penal, sugerindo a necessidade de sua inserção em um modelo de integração probatória sob controle humano rigoroso. Tais limites não parecem decorrer apenas de dificuldades técnicas, mas também de uma possível tensão entre a lógica probabilística dos sistemas automatizados e o modelo garantista de processo penal, que exige fundamentação racional e possibilidade efetiva de contraditório (Hildebrandt, 2019).

O primeiro desses limites manifesta-se na opacidade dos sistemas algorítmicos, frequentemente descritos como “caixas-pretas”. Em muitos modelos contemporâneos, especialmente aqueles baseados em aprendizado profundo, pode não ser possível reconstruir, de maneira clara, o percurso lógico que conduz da entrada de dados ao resultado final. Como observa Frank Pasquale, quanto mais opaco o sistema, maior o desafio para responsabilização e controle (Pasquale, 2015, p. 3, tradução nossa). Em termos mais amplos, o autor sustenta que a opacidade algorítmica pode comprometer a *accountability* institucional e dificultar o controle jurídico sobre decisões automatizadas (Pasquale, 2015).

No âmbito do processo penal, essa opacidade pode inviabilizar ou, ao menos, dificultar o exercício pleno do contraditório, na medida em que pode impedir que a parte compreenda, questione e eventualmente refute os critérios utilizados na formação da prova. A exigência de transparência metodológica, portanto, não se limita a uma dimensão técnica, podendo ser compreendida como condição relevante de validade da própria prova penal.

O segundo limite epistemológico refere-se à dependência estrutural dos dados utilizados para o treinamento dos sistemas de inteligência artificial. Diferentemente da prova tradicional, cuja materialidade pode ser diretamente examinada, a prova algorítmica depende de bases de dados cuja qualidade, representatividade e integridade nem sempre se mostram plenamente transparentes. Nesse sentido, Tarleton Gillespie afirma que algoritmos não são meros instrumentos técnicos, mas incorporam escolhas humanas e institucionais (Gillespie, 2014, p. 167). De forma complementar, o autor ressalta que essas escolhas tendem a influenciar diretamente os resultados produzidos, afastando a ideia de neutralidade tecnológica (Gillespie, 2014).

Assim, a aparente objetividade das inferências automatizadas pode ocultar processos de seleção e ponderação de dados que tendem a reproduzir desigualdades estruturais. A literatura indica que sistemas algorítmicos devem ser analisados não apenas por sua eficiência técnica, mas também por seus efeitos sociais e jurídicos (Gillespie, 2014).

O terceiro limite relaciona-se à natureza probabilística das inferências produzidas por sistemas de inteligência artificial. Diferentemente da prova tradicional, que se orienta pela demonstração de fatos concretos, a IA opera por meio de correlações e estimativas de probabilidade, o que pode implicar uma mudança significativa no objeto do conhecimento produzido. Como argumenta Bernard E. Harcourt, “a lógica atuarial desloca o foco da responsabilidade individual para padrões estatísticos de risco” (Harcourt, 2007, p. 35). Em termos gerais, o autor critica a expansão de modelos preditivos no campo penal, alertando para os riscos de substituição do juízo jurídico por métricas estatísticas (Harcourt, 2007).

Esse deslocamento pode se revelar particularmente problemático no processo penal, pois pode levar à valorização de expectativas probabilísticas em detrimento da comprovação de fatos efetivamente ocorridos, comprometendo, ao menos em tese, o padrão de prova exigido para a responsabilização penal.

No plano jurídico, tais limitações epistemológicas podem impactar garantias fundamentais do processo penal. A exigência de motivação das decisões judiciais pode se mostrar tensionada diante da utilização de resultados cuja lógica interna não pode ser explicitada. Do mesmo modo, o contraditório e a ampla defesa podem restar comprometidos quando a parte não dispõe de meios adequados para compreender e contestar a prova produzida por sistemas automatizados. Como observa Thomas Streinz, a governança da inteligência artificial tende a exigir mecanismos institucionais de transparência e controle jurídico (Streinz, 2021).

A jurisprudência brasileira, embora ainda incipiente no enfrentamento direto da prova algorítmica, oferece parâmetros relevantes ao exigir a observância da cadeia de custódia e da verificabilidade técnica das provas digitais. Tais exigências podem ser estendidas, por analogia, à prova produzida por inteligência artificial, indicando a necessidade de documentação dos processos algorítmicos, validação independente e possibilidade de auditoria pelos sujeitos processuais, entendimento que encontra respaldo em precedentes como o HC 598.886/SC do Superior Tribunal de Justiça (Brasil, 2023).

Dessa forma, os limites epistemológicos da prova automatizada não apenas sugerem fragilidades técnicas, mas também revelam uma possível tensão estrutural entre a racionalidade algorítmica e o modelo garantista de processo penal. A superação dessa tensão não parece decorrer da exclusão da tecnologia, mas de sua submissão a parâmetros jurídicos rigorosos, que busquem assegurar controle humano, transparência metodológica e efetividade do contraditório. Nessas condições, será possível compatibilizar o uso da inteligência artificial com os princípios fundamentais que regem a atividade probatória no processo penal contemporâneo.

2.1 Opacidade algorítmica como vício de origem da prova e a Teoria dos Frutos da Árvore Envenenada

A análise dos limites epistemológicos da prova produzida por inteligência artificial pode ser significativamente aprofundada a partir de sua articulação com as teorias de invalidade da prova no processo penal, especialmente a teoria dos frutos da árvore envenenada.

Tradicionalmente, essa teoria estabelece que provas derivadas de fontes ilícitas ou inválidas tendem a ser igualmente contaminadas, em razão da ruptura originária com parâmetros de legalidade e confiabilidade. No contexto da inteligência artificial, pode-se sustentar que a opacidade algorítmica, quando impede a verificação da metodologia empregada, pode configurar um vício de origem da prova, não necessariamente por ilicitude formal, mas por insuficiência epistemológica.

Isso ocorre porque a ausência de transparência, auditabilidade e explicabilidade compromete a possibilidade de reconstrução racional do percurso inferencial, elemento central da prova no processo penal. Nessa perspectiva, a prova deixa de ser um meio de conhecimento controlável intersubjetivamente e passa a operar como um resultado tecnicamente produzido, porém não plenamente justificável no plano jurídico.

Assim, inferências produzidas por sistemas não auditáveis podem, em determinadas circunstâncias, irradiar efeitos contaminantes sobre a cadeia probatória, especialmente quando utilizadas como fundamento relevante da decisão judicial. A contaminação, nesse caso, não decorre apenas da ilicitude originária, mas

da impossibilidade de validação epistemológica do processo que gerou a informação.

Tal construção sugere que a admissibilidade e a valoração da prova algorítmica podem estar condicionadas à observância de requisitos mínimos de transparência e controle, sob pena de comprometimento da validade do conjunto probatório.

Nesse sentido, a doutrina processual penal contemporânea oferece importantes subsídios. Conforme leciona Aury Lopes Jr.:

A prova penal somente se legitima quando submetida ao contraditório efetivo, o que pressupõe não apenas acesso ao resultado, mas também compreensão crítica de seu processo de formação. Elementos probatórios cuja origem ou metodologia não possam ser devidamente escrutinadas tendem, portanto, a fragilizar o devido processo legal (Lopes Jr., 2023).

De forma convergente, Gustavo Badaró (2021) sustenta que a validade da prova não se esgota em sua admissibilidade formal, exigindo também confiabilidade epistêmica, entendida como a possibilidade de verificação racional dos critérios que sustentam a inferência probatória. Nesse sentido, a opacidade algorítmica pode comprometer não apenas a força persuasiva da prova, mas sua própria legitimidade jurídica.

A reflexão desenvolvida por Michele Taruffo reforça essa perspectiva ao destacar que:

A decisão judicial racional exige a possibilidade de reconstrução argumentativa do caminho probatório, permitindo sua justificação intersubjetiva. A impossibilidade de explicitação dos critérios decisórios, característica de determinados sistemas algorítmicos, tensiona diretamente esse modelo de racionalidade (Taruffo, 2012).

Sob outro enfoque, a crítica de Frank Pasquale acerca da opacidade algorítmica revela que tais sistemas não apenas dificultam a compreensão técnica de seus resultados, mas também estruturam assimetrias de poder informacional. No processo penal, essa assimetria pode comprometer a paridade de armas, na medida em que a acusação pode se valer de ferramentas cujo funcionamento não é acessível à defesa, dificultando o exercício do contraditório técnico.

Além disso, autores como Mirjan Damaška apontam que:

A prova, em sistemas processuais contemporâneos, não pode ser dissociada de seus pressupostos institucionais de controle, sob pena de esvaziamento de sua função garantidora. A introdução de inferências automatizadas não auditáveis pode representar, nesse sentido, uma ruptura

com a lógica de controle que estrutura o processo penal democrático (Damaška, 1997).

Diante desse quadro, a aplicação da teoria dos frutos da árvore envenenada ao contexto da inteligência artificial pode ser compreendida como uma extensão funcional da exigência de confiabilidade da prova. Não se trata de equiparar automaticamente a opacidade à ilicitude, mas de reconhecer que a ausência de transparência pode comprometer estruturalmente a validade do conhecimento produzido.

Assim, sempre que a prova algorítmica não possibilitar controle técnico, contraditório efetivo e reconstrução racional de seu percurso inferencial, sua utilização poderá ser questionada não apenas no plano de sua valoração, mas também de sua admissibilidade. Nessa perspectiva, a teoria dos frutos da árvore envenenada revela-se instrumento útil para problematizar a extensão dos efeitos de vícios epistemológicos na cadeia probatória.

3 GARANTIAS PROCESSUAIS E O PAPEL DO JUIZ FRENTE À AUTOMAÇÃO DECISÓRIA

As garantias processuais podem ser compreendidas como pilares estruturantes do Estado Democrático de Direito e tendem a assumir especial relevância no processo penal, no qual a tutela dos direitos fundamentais do acusado deve, em regra, prevalecer sobre promessas de eficiência tecnológica. No contexto da produção e valoração de provas por inteligência artificial, essas garantias tendem a se tornar ainda mais sensíveis, uma vez que a complexidade técnica e a opacidade dos sistemas automatizados podem comprometer elementos essenciais do devido processo legal, como o contraditório, a ampla defesa, a igualdade de armas e a imparcialidade judicial. A doutrina processual penal brasileira tem sustentado que a admissibilidade e a valoração da prova não devem prescindir de critérios objetivos de confiabilidade, transparência e possibilidade de controle crítico, sob o risco de esvaziamento das garantias constitucionais que legitimam a atuação do Estado na persecução penal. Nesse sentido, a prova que não se submete ao exame crítico do julgador e da defesa tende a perder densidade jurídica e a se aproximar de uma presunção de autoridade, em detrimento de sua função como elemento racional de convencimento (Mirabete, 2022).

Quando se trata de provas produzidas por sistemas de inteligência artificial, a concretização dessas garantias pode enfrentar desafios relevantes. Um dos mais significativos refere-se à transparência metodológica. Muitos sistemas algorítmicos, especialmente aqueles baseados em técnicas de aprendizado profundo, operam por meio de modelos cuja lógica interna pode não ser plenamente acessível ou compreensível, fenômeno frequentemente descrito como “caixa-preta” epistemológica. Essa opacidade pode comprometer o exercício do contraditório e da ampla defesa, na medida em que pode dificultar que a parte conheça, questione e conteste os fundamentos que levaram à inferência apresentada como prova. Nessa perspectiva, Douzinas e Gearey (2019) defendem que:

A tecnologia, para ser compatível com a justiça, deve ser compreensível, auditável e passível de escrutínio humano; do contrário, a prova automatizada pode se converter em uma forma de autoridade técnica, com potencial de substituir indevidamente o juízo crítico no processo penal.

Outro aspecto central refere-se à necessidade de validação e auditabilidade das provas automatizadas. A confiabilidade de uma inferência algorítmica tende a depender não apenas do resultado final apresentado, mas também da qualidade dos dados utilizados, dos critérios de treinamento do modelo, dos parâmetros adotados e das margens de erro envolvidas. A literatura crítica sobre algoritmos decisórios indica que sistemas utilizados para avaliação de risco ou identificação de padrões podem reproduzir e até amplificar discriminações históricas, sobretudo quando treinados a partir de bases de dados enviesadas. Nesse sentido, O’Neil (2016) sustenta que:

A utilização de algoritmos sem mecanismos rigorosos de supervisão e controle pode colocar em risco a equidade do processo penal e potencialmente gerar decisões injustas sob a aparência de neutralidade matemática.

A jurisprudência brasileira, ainda que não trate diretamente da prova penal produzida por inteligência artificial, já fornece parâmetros relevantes para essa discussão. O Superior Tribunal de Justiça tem reiteradamente afirmado que provas digitais obtidas sem metodologia adequada ou sem documentação idônea da cadeia de custódia tendem a ser consideradas inadmissíveis, por comprometerem a integridade, a autenticidade e a confiabilidade do material probatório, além de dificultarem o exercício do contraditório (Brasil, 2023). Essa orientação jurisprudencial pode ser utilizada, por analogia, no tratamento das provas

automatizadas, indicando a necessidade de observância de critérios mínimos de auditabilidade, rastreabilidade e transparência metodológica como condição relevante de validade.

Em reforço a essa compreensão, o Superior Tribunal de Justiça tem ressaltado que a obtenção e utilização de dados digitais devem observar rigorosamente as garantias processuais, sob pena de invalidade da prova, conforme discutido no HC 455.223/SC³¹. Tal orientação sugere que a incorporação de tecnologias no processo penal não dispensa, mas possivelmente intensifica, a necessidade de observância das garantias constitucionais.

A possibilidade de contraprova técnica constitui outra dimensão essencial das garantias processuais no contexto da inteligência artificial. Diferentemente das provas tradicionais, cuja metodologia pode ser analisada e contestada por peritos independentes, os resultados de sistemas algorítmicos podem se mostrar parcialmente inacessíveis à crítica técnica, sobretudo quando protegidos por sigilo comercial ou quando a complexidade do modelo dificulta sua compreensão. Essa assimetria técnica pode colocar a defesa em posição de desvantagem estrutural, afetando a igualdade de armas e o devido processo legal. Por essa razão, sustenta-se que sistemas de inteligência artificial empregados na produção de provas devem, idealmente, ser passíveis de exame por especialistas independentes, de modo que suas inferências estatísticas possam ser avaliadas à luz de critérios científicos, jurídicos e constitucionais (Sampaio, 2025).

Nesse cenário, o papel do juiz tende a assumir relevo decisivo. O magistrado não deve se limitar a receber passivamente o resultado de uma inferência algorítmica, atribuindo-lhe valor probatório automático ou presumido. Ao contrário, pode-se sustentar que lhe cabe exercer uma função ativa de controle, questionando a metodologia empregada, a origem e a qualidade dos dados, a existência de vieses e a compatibilidade da prova com os princípios do contraditório e da ampla defesa. A literatura internacional tem indicado que:

Decisões judiciais fundadas exclusivamente em resultados automatizados, sem supervisão humana crítica, podem representar risco de arbitrariedade e eventual violação de direitos fundamentais (Rosen; Citron, 2020).

³ STJ. Habeas Corpus n. 455.223/SC. Relator: Ministro Reynaldo Soares da Fonseca. Quinta Turma. Julgado em 12 mar. 2019

Nesse contexto, a inteligência artificial tende a ser mais adequadamente compreendida como instrumento auxiliar à atividade jurisdicional, e não como substituto do juízo humano. Nessa perspectiva, o magistrado pode também atuar como instância de contenção da opacidade algorítmica, conforme problematizado por Frank Pasquale, exigindo postura ativa na verificação dos pressupostos, limites e condições de validade das inferências produzidas por sistemas automatizados.

Por fim, a integração das provas produzidas por inteligência artificial com os meios tradicionais de prova pode se apresentar como uma salvaguarda relevante. Inferências automatizadas não devem, em regra, ocupar posição central ou exclusiva na formação do convencimento judicial, mas funcionar como elementos complementares, a serem analisados em conjunto com testemunhos, laudos periciais e demais provas produzidas sob métodos clássicos de controle. Essa abordagem tende a preservar a centralidade do juiz como garantidor do devido processo legal e a assegurar que a tecnologia contribua para o aprimoramento da atividade jurisdicional sem comprometer a legitimidade democrática do processo penal.

Em síntese, a incorporação da inteligência artificial na produção e valoração da prova penal tende a exigir a reafirmação das garantias processuais como limites relevantes à automação decisória. Transparência metodológica, auditabilidade, possibilidade de contraprova técnica, supervisão humana crítica e integração com os meios probatórios tradicionais não devem ser compreendidas apenas como opções de política judiciária, mas como exigências de natureza constitucional. O juiz, nesse contexto, tende a assumir o papel de guardião contra a transformação da tecnologia em fonte autônoma de autoridade decisória, buscando assegurar que a inovação tecnológica fortaleça e não fragilize o devido processo legal, o contraditório, a ampla defesa e a imparcialidade judicial.

4 RISCOS DA AUTOMAÇÃO DECISÓRIA

A automação decisória no processo penal, impulsionada pelo uso crescente de sistemas de inteligência artificial, pode apresentar riscos relevantes à preservação do devido processo legal, da imparcialidade judicial e da centralidade das garantias fundamentais do acusado. Nesse sentido, Grimm, Grossman e Cormack (2021) defendem que:

Diferentemente da decisão judicial humana, fundada na interpretação crítica das provas, na ponderação de valores jurídicos e na fundamentação racional, os mecanismos automatizados operam a partir de inferências estatísticas e modelos preditivos, que podem ser incorporados ao convencimento judicial sem o necessário escrutínio epistemológico.

Ainda nesta mesma linha, os referidos autores defendem que:

Um dos principais riscos associados a esse fenômeno reside na possível transformação de probabilidades estatísticas em critérios de autoridade, o que pode deslocar o eixo da decisão de fatos juridicamente verificáveis para resultados algorítmicos cuja racionalidade subjacente nem sempre é plenamente acessível ao julgador ou às partes (Grimm; Grossman; Cormack, 2021).

A literatura especializada indica que sistemas de automação decisória tendem a incorporar e reproduzir vieses históricos, sociais e institucionais presentes nos bancos de dados utilizados para seu treinamento. Cathy O'Neil demonstra que algoritmos de avaliação de risco e predição de comportamento podem penalizar, de forma desproporcional, indivíduos pertencentes a grupos socialmente vulneráveis, produzindo um efeito de amplificação das desigualdades estruturais já existentes (O'Neil, 2016). Tal constatação sugere que a automação decisória pode não ser neutra, implicando uma possível transferência de responsabilidade moral e jurídica para sistemas técnicos, o que pode tensionar os princípios constitucionais do devido processo legal, do contraditório e da ampla defesa (Brasil, 1988).

Outro risco relevante consiste na falsa percepção de objetividade e neutralidade associada às decisões automatizadas. A sofisticação técnica dos algoritmos pode induzir operadores do direito a atribuir aos seus resultados um grau excessivo de confiabilidade, como se expressassem verdades necessárias e incontestáveis. No entanto, tais resultados decorrem de modelos probabilísticos condicionados por escolhas humanas, desde a seleção de dados até a definição de parâmetros e pesos decisórios. Conforme alertam Rosen e Citron:

A automatização decisória pode gerar uma ilusão de certeza, reduzindo a disposição crítica do julgador e favorecendo decisões potencialmente acríicas ou arbitrárias (Rosen; Citron, 2020).

Embora o ordenamento jurídico brasileiro ainda não enfrente diretamente decisões judiciais plenamente automatizadas, a jurisprudência oferece parâmetros analógicos relevantes no tratamento das provas digitais e dos métodos periciais tecnológicos. O Superior Tribunal de Justiça tem afirmado que provas obtidas sem metodologia verificável ou sem documentação idônea tendem a ser consideradas

inadmissíveis, por comprometerem a confiabilidade do processo e dificultarem o exercício do contraditório. Tal orientação também se verifica no HC 598.886/SC, no qual se reforça a exigência de integridade e rastreabilidade da prova digital. Por analogia, a utilização de sistemas de automação decisória pode demandar a observância de critérios rigorosos de transparência, auditabilidade e possibilidade efetiva de contestação, de modo a assegurar que a decisão final permaneça sob controle humano e que o juiz atue como garante ativo das liberdades fundamentais.

Deve-se considerar, ainda, o risco de desumanização do processo penal. Inferências automatizadas, especialmente aquelas voltadas à previsão de reincidência ou periculosidade, tendem a desconsiderar fatores contextuais, subjetivos e sociais relevantes à apreciação do caso concreto. A redução da complexidade humana a padrões estatísticos pode conduzir à adoção de medidas potencialmente desproporcionais, dissociadas de uma análise individualizada e eticamente orientada. Douzinas e Gearey ressaltam que:

A justiça penal não pode ser reduzida a uma equação matemática, pois o direito exige interpretação crítica, ponderação de valores e responsabilidade moral, elementos que permanecem intrinsecamente humanos (Douzinas; Gearey, 2019).

Por fim, a automação decisória também pode suscitar o problema da opacidade institucional e da responsabilização difusa. Quando decisões judiciais são influenciadas por sistemas algorítmicos complexos, pode se tornar difícil atribuir responsabilidade por eventuais erros ou injustiças, deslocando-se a culpa do decisor humano para o sistema ou para seus desenvolvedores. Essa fragmentação da responsabilidade pode comprometer o princípio da *accountability*, essencial ao Estado Democrático de Direito, e eventualmente enfraquecer a confiança pública no sistema de justiça penal (Sampaio, 2025).

Diante desses riscos, a doutrina tende a convergir no sentido de que a automação decisória deve possuir caráter predominantemente subsidiário, funcionando como instrumento auxiliar à atividade jurisdicional, e não como substituto do juízo humano. Nesse contexto, sustenta-se que toda inferência automatizada deve estar sujeita à supervisão crítica do magistrado, à auditabilidade técnica e à possibilidade de contraprova, com vistas à preservação do contraditório, da ampla defesa e do devido processo legal. Nessas condições, torna-se possível

compatibilizar o uso da inteligência artificial com a racionalidade garantista que orienta o processo penal contemporâneo.

CONSIDERAÇÕES FINAIS

A produção de provas por meio de sistemas de inteligência artificial pode ser compreendida como uma das transformações mais relevantes do processo penal contemporâneo, ao mesmo tempo em que tende a tensionar categorias dogmáticas clássicas relacionadas à verdade, à prova e ao convencimento judicial. Embora tais tecnologias apresentem potencial para ampliar a capacidade investigativa do Estado, sua utilização no âmbito penal pode revelar limites epistemológicos e riscos processuais que não devem ser ignorados.

Ao longo deste trabalho, buscou-se demonstrar que a prova produzida por IA não se confunde, em regra, com os meios probatórios tradicionais, na medida em que resulta de inferências automatizadas baseadas em modelos estatísticos complexos, frequentemente marcados por opacidade metodológica e suscetíveis à reprodução de vieses históricos. A incorporação acrítica desses outputs ao processo penal pode converter probabilidades em critérios de autoridade, o que tende a enfraquecer o contraditório, a ampla defesa e o próprio devido processo legal.

Indicou-se, ainda, que a automação decisória não deve substituir o juízo humano sem que haja risco de comprometimento da racionalidade da decisão judicial e da responsabilização institucional. Nesse sentido, a centralidade do juiz, enquanto garantidor dos direitos fundamentais, sugere que toda inferência algorítmica deva ser submetida a controle crítico, auditabilidade e possibilidade efetiva de contraprova, preservando-se a lógica do convencimento motivado e a exigência de fundamentação.

Diante dessas premissas, sugere-se que a inteligência artificial deva ocupar papel predominantemente subsidiário na produção e valoração da prova penal, funcionando como instrumento auxiliar, e não como fonte autônoma de verdade processual. A legitimidade de seu uso tende a depender da observância de critérios rigorosos de transparência, explicabilidade e controle, sob o risco de comprometimento da justiça do caso concreto e da confiança pública no sistema penal.

Por fim, a compatibilização entre inovação tecnológica e garantias processuais não parece ser apenas desejável, mas tende a se apresentar como condição relevante para que a inteligência artificial possa ser incorporada de forma constitucionalmente adequada ao processo penal.

REFERÊNCIAS

BADARÓ, Gustavo Henrique. **Processo penal**. 9. ed. São Paulo: Revista dos Tribunais, 2021.

BAROCAS, Solon; SELBST, Andrew D. Big Data's Disparate Impact. **California Law Review**, v. 104, n. 3, 2016.

BRASIL. Superior Tribunal de Justiça. **Habeas Corpus n. 598.886/SC**. Relator: Ministro Ribeiro Dantas. 5ª Turma. Julgado em 27 out. 2020. Diário da Justiça Eletrônico, Brasília, 03 nov. 2020.

BRASIL. Superior Tribunal de Justiça. **Recurso Ordinário em Habeas Corpus n. 131.263/DF**. Relator: Ministro Ribeiro Dantas. 5ª Turma. Julgado em 09 jun. 2020. Diário da Justiça Eletrônico, Brasília, 15 jun. 2020.

BRASIL. Superior Tribunal de Justiça. **Habeas Corpus n. 455.223/SC**. Relator: Ministro Reynaldo Soares da Fonseca. Quinta Turma. Julgado em 12 mar. 2019. Disponível em: <https://www.stj.jus.br>. Acesso em: 26 abr. 2026.

DAMAŠKA, Mirjan R. **Evidence Law Adrift**. New Haven: Yale University Press, 1997.

DOUZINAS, Costas; GEAREY, Adam. **Critical Jurisprudence and the Role of Technology in Law**. Oxford: Hart Publishing, 2019.

GILLESPIE, Tarleton. The relevance of algorithms. In: GILLESPIE, Tarleton; BOCZKOWSKI, Pablo; FOOT, Kirsten (org.). **Media Technologies: Essays on Communication, Materiality, and Society**. Cambridge: MIT Press, 2014. p. 167–194

GRIMM, Robert; GROSSMAN, Richard; CORMACK, Gordon. **Artificial Intelligence in Criminal Justice: Risks and Opportunities**. Cambridge: Cambridge University Press, 2021.

HARCOURT, Bernard E. **Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age**. Chicago: University of Chicago Press, 2007.

HILDEBRANDT, Mireille. **Law for Computer Scientists and Other Folk**. Oxford: Oxford University Press, 2019.

IRFAN, M.; Bennett, C.; Müller, V. **AI and Criminal Evidence**: Emerging Challenges. Oxford: Oxford University Press, 2025.

LOPES JR., Aury. **Direito processual penal**. 20. ed. São Paulo: Saraiva, 2023.

MIRABETE, Julio Fabbrini. **Processo Penal**. 33. ed. São Paulo: Atlas, 2022.

O'NEIL, Cathy. **Weapons of Math Destruction**: How Big Data Increases Inequality and Threatens Democracy. New York: Crown, 2016.

PA Leari, Luciano. **Epistemologia e Prova no Processo Penal**. Porto Alegre: Livraria do Advogado, 2021.

PASQUALE, Frank. **The Black Box Society**: The Secret Algorithms That Control Money and Information. Cambridge: Harvard University Press, 2015.

ROSEN, Jeffrey; CITRON, Danielle. The Challenges of AI Evidence in Legal Contexts. **Yale Law Journal**, v. 130, n. 5, 2020.

SAMPAIO, Raiane Xavier Moura. **Provas Digitais e Inteligência Artificial no Processo Penal**. Salvador: Juspodivm, 2025.

STREINZ, Thomas. Artificial Intelligence and the Law: An Overview. In: DE GREGORIO, Giovanni; POLLICINO, Oreste (eds.). **The Oxford Handbook of Digital Constitutionalism**. Oxford: Oxford University Press, 2021.

TARUFFO, Michele. **La prova dei fatti giuridici**. Milano: Giuffrè, 2012.