

# **A TRANSPARÊNCIA E A EXPLICABILIDADE COMO PRINCÍPIOS NORTEADORES DA APLICAÇÃO DA INTELIGÊNCIA ARTIFICIAL NO DIREITO**

Felipe César Pires de Oliveira<sup>1</sup>

Orientador: Prof. Dr. Almir Garcia Fernandes<sup>2</sup>

## **RESUMO:**

O presente trabalho analisa a crescente integração da Inteligência Artificial (IA) ao Poder Judiciário brasileiro, com foco na tensão existente entre a busca por eficiência processual e a preservação das garantias fundamentais. Adota-se metodologia de pesquisa descritiva e exploratória para investigar os riscos associados à opacidade algorítmica e aos vieses automatizados, exemplificados por casos concretos de discriminação tecnológica. O estudo demonstra como o fenômeno da "caixa-preta" ameaça o devido processo legal e o dever constitucional de fundamentação das decisões. Examina-se a essencialidade dos princípios da transparência e da explicabilidade como requisitos indispensáveis de validade ética e jurídica, à luz da legislação pátria, da Resolução CNJ nº 615/2025 e da Recomendação sobre a Ética da IA da UNESCO. Discute-se a insuficiência da mera abertura de código-fonte (transparência formal) frente à necessidade de tradução da lógica algorítmica (explicabilidade material). Conclui-se que a modernização da justiça depende da implementação de mecanismos que garantam a inteligibilidade das decisões automatizadas, assegurando que a tecnologia atue como instrumento de acesso à justiça sob rigoroso controle humano e democrático, viabilizando o efetivo exercício do contraditório e da ampla defesa.

**Palavras-chave:** Inteligência Artificial. Princípios. Explicabilidade. Transparência. Vieses algorítmicos.

---

<sup>1</sup> Discente do curso de graduação de Direito da Faculdade Jacy de Assis, da Universidade Federal de Uberlândia (UFU).

<sup>2</sup> Doutor em Direito Comercial pela Pontifícia Universidade Católica de São Paulo (PUC-SP). Mestre em Direito pela Universidade de Franca. Especialista em Direito Processual Civil e graduado em Direito pela Universidade Federal de Uberlândia. Professor da Faculdade de Direito da Universidade Federal de Uberlândia (UFU).

## 1. INTRODUÇÃO

Frente aos avanços tecnológicos atuais, impossível pensar no cotidiano de boa parte da população sem considerar o uso de tais ferramentas. Da realização de tarefas diárias até o usufruto de momentos de lazer, os sistemas tecnológicos se mostram indispensáveis, tidos como parte essencial do modo de vida moderno; e, dessa realidade, não escapa o Direito. Dentre tais tecnologias, destaca-se a Inteligência Artificial, que vem revolucionando cada vez mais o cenário jurídico contemporâneo, nas suas mais diversas utilidades.

Sob tal perspectiva, vê-se necessário um estudo cada vez mais aprofundado de tais tecnologias, de seus riscos, benefícios, funções e demais questões que circundam o tema e que se mostram de suma importância frente a análise de diversos temas na atualidade, como será feito neste trabalho em relação ao âmbito jurídico - com um foco na no sistemas inteligentes, que vêm se destacando na seara jurídica. O uso cada vez maior destas tecnologias nos dias de hoje, tanto pelo Poder Público quanto pelos operadores do Direito, exige uma reflexão crítica de sua aplicação e de seus impactos éticos e jurídicos, de modo a evitar a lesão aos direitos fundamentais dos jurisdicionados.

Como exemplo, imagine, pois, que o Poder Judiciário adote uma ferramenta de inteligência artificial que automatize a aplicação de precedentes e temas repetitivos, vinculando processos a teses jurídicas já fixadas pelos tribunais, mas que, contudo, não permite aos jurisdicionados entender a motivação da referida decisão ou mesmo questioná-la, devido à opacidade do sistema; infligindo, por consequência, o direito à decisão fundamentada, consoante disposto no art. 93, inciso IX, da Constituição Federal.

Neste sentido, não obstante os claros avanços operacionais que a Inteligência Artificial é capaz de gerar no trabalho jurídico (principalmente àqueles relacionados à celeridade processual), os princípios da abertura e da explicabilidade são essenciais para garantir que a inteligência artificial não se torne um obstáculo ao exercício da cidadania e garantam o uso responsável e transparente destas tecnologias. A transparência refere-se à abertura quanto ao funcionamento do sistema, permitindo que se saiba quais dados foram utilizados e quais objetivos a ferramenta persegue; ou seja, trata-se de disponibilizar ao público em geral, a estruturação do sistema. Já a explicabilidade - enquanto complemento indispensável da abertura, vai além, exigindo que o sistema seja capaz de traduzir processos matemáticos complexos em justificativas compreensíveis para seres humanos.

Vale ressaltar, ainda, que a discussão aqui proposta não se limita ao campo teórico, mas reflete uma urgência normativa global e nacional. Diante de iniciativas regulatórias recentes, como as resoluções do Conselho Nacional de Justiça (CNJ) e os debates internacionais acerca do assunto, a compreensão profunda da transparência e da explicabilidade torna-se um pressuposto de validade para a modernização do Judiciário. Com a efetiva implementação desses princípios será possível harmonizar a inevitável eficiência tecnológica com as garantias inegociáveis do devido processo legal, impedindo que a "caixa-preta" dos algoritmos substitua a fundamentação racional das decisões judiciais.

Neste contexto, o presente artigo tem por objetivo analisar, mediante pesquisa descritiva e exploratória sobre o tema, a essencialidade dos princípios da transparência e explicabilidade para a aplicação justa e ética das tecnologias de Inteligência Artificial na seara judicial. Ademais, tem como objetivos específicos introduzir, ainda que brevemente, à história da Inteligência Artificial e sua relação com o Direito, o grande avanço de tais sistemas inteligentes no Direito na atualidade brasileira (demonstrando sua atual aplicação, principalmente, no Poder Judiciário), demonstrar os principais entraves éticos e judiciais à aplicação de tais tecnologias no ramo jurídico e como os princípios acima expostos são, em realidade, complementares e necessários um ao outro.

## **2. Breve introdução à história da Inteligência Artificial**

De modo a melhor entender quais princípios devem nortear a aplicação de sistemas inteligentes no Direito, entende-se necessário uma introdução, ainda que breve e superficial, sobre o que seria a “Inteligência Artificial” em si; para, então, analisar suas aplicabilidades no âmbito judicial.

A “Inteligência Artificial”, enquanto campo de estudo, surge em 1956, na “*Dartmouth Conference's*”, pelo seu organizador e também cientista da computação John McCarthy (Copeland, 1993, p. 28). Contudo, as discussões acerca da possibilidade de máquinas inteligentes têm seu principal impulso com o cientista da computação e matemático britânico Alan Turing, com a publicação de seu artigo “*Computing Machinery and Intelligence*”, em 1950 (Copeland, 1993, p. 29).

Logo no início de seu trabalho, Turing estabelece uma das principais questões sobre a questão da inteligência artificial: “*Can machines think?*”<sup>3</sup> (Turing, 1950, p. 433). Neste artigo, o cientista britânico estabelece importantes ideais que são utilizadas até hoje no desenvolvimento de tais tecnologias, como o “jogo da imitação”: teste que mediria a capacidade de uma máquina se passar por um ser pensante e, deste modo, ser indistinguível de um ser humano para aquele que analisa a situação (Zilio, 2008, p. 209).

Dito isso, como poderíamos conceituar “inteligência artificial”? De modo geral, a literatura não estabelece uma conceituação definitiva sobre o termo. Por um lado, a inteligência artificial pode ser entendida como o conjunto de técnicas e métodos, a ser utilizado em uma máquina, que visa “imitar” a capacidade cognitiva humana (Sanctis, 2020, p.119). De modo semelhante, Haugeland (1985, *apud* Russel, Norvig, 2013, p. 23) estabelece a inteligência artificial como um empenho humano em fazer com que máquinas, assim como os seres humanos, pensem. Outros, como Rich e Knight (1991, *apud* Russel, Norvig, 2013, p. 23), entendem a inteligência artificial como uma área de estudo que visa analisar como os computadores podem desempenhar atividades que são, a priori, humanas - como, por exemplo, aquelas que são realizadas no âmbito jurídico, de modo geral.

Inobstante a diversidade de definições, certo ponto é comum à maioria: a Inteligência Artificial busca, pela aplicação de determinados métodos algorítmicos, fazer com que sistemas tecnológicos reproduzam o comportamento humano diante de certas tarefas. Contudo, com a atual evolução da disciplina, especialmente com o surgimento de novos modelos de linguagem e redes neurais profundas, o uso de tais sistemas inteligentes tem ido além da mera automação de processos repetitivos, buscando a mimetização, cada vez maior, de faculdades cognitivas complexas, como a de aprendizado e tomada de decisões, extremamente importantes ao campo jurídico.

Ante o exposto, a partir de agora, buscar-se-á entender, de forma mais aprofundada, quais são as principais aplicações de tais sistemas inteligentes no âmbito judicial e exemplos de sua utilização para, então, passarmos à análise da transparência e da interpretabilidade como princípios basilares no que diz respeito à utilização da IA no direito.

### **3. A Inteligência Artificial e o Direito**

Como falado anteriormente, é intensa a difusão das tecnologias na atualidade. As ferramentas tecnológicas vêm se tornando cada vez mais utilizadas na realização de tarefas

---

<sup>3</sup> Em tradução livre: “máquinas podem pensar?”

que antes não dependiam destas, realidade que também alcança o direito de forma cada vez mais clara - visto que a evolução do Direito não é estranha à evolução das próprias ciências naturais (Magalhães, 2005, p. 336); fato que, independentemente de uma posição tecnófila ou não, trás a necessidade de uma análise apurada da situação.

No âmbito jurídico, a busca por tempestividade e eficiência vem criando um ambiente cada vez mais atrativo para o avanço de tecnologias de inteligência artificial; apresentando-se como uma solução vital para o enfrentamento do excessivo volume de processos e para a redução do tempo de tramitação das demandas nos tribunais. Essa expansão, no entanto, não se limita apenas à otimização de fluxos administrativos, mas alcança o próprio cerne da atividade intelectual jurídica, oferecendo ferramentas que vão desde a predição de resultados baseada em jurimetria até a automação de peças processuais complexas.

Neste sentido, uma das principais aplicabilidades de tais tecnologias no Direito, na atualidade, se dá na chamada “análise preditiva”, que consiste no uso de ferramentas de inteligência artificial na análise de dados passados para a criação de modelos preditivos (Barddal, Freitas, 2019, p. 110); ou seja, uma tentativa de antecipar decisões futuras, baseando-se, por exemplo, em decisões passadas. Trata-se, portanto, de uma tentativa de antecipar o desfecho de litígios futuros ao correlacionar diferentes variáveis específicas. Portanto, a análise preditiva oferece ao operador do Direito a capacidade de mitigar riscos e otimizar a tomada de decisão, o que contribui diretamente para a redução do congestionamento do Poder Judiciário. Ao fornecer prognósticos precisos, essas ferramentas desestimulam litígios de sucesso improvável, promovendo uma justiça mais célere e alinhada à realidade da era dos dados..

Os sistemas inteligentes são também comumente usados para a análise técnica e interpretação de informações que lhes são disponibilizadas (como, por exemplo, a admissibilidade de recursos em certos processos), bem como para a elaboração e interpretação de contratos, peças jurídicas, e outros semelhantes (Soares, Kauffman, Chao, 2020,p. 119).

Outrossim, demais funcionalidades da Inteligência Artificial no Direito são temas polêmicos, como, por exemplo, a possibilidade de tais tecnologias proferirem decisões no bojo de processos - principalmente criminais. Exemplo clássico de tal questão ocorreu no Estado norte-americano de Nova York, que utilizava um sistema de inteligência artificial - denominado COMPAS (*Correctional Offender Management Profiling for Alternative Sanctions*), para realizar decisões sobre a liberdade condicional dos presos (Cozman, Kaufman, 2022, p. 203).

Ocorre, porém, que estudos sobre o sistema mostraram que este apresentava certas distorções em sua execução; ocasionando, por exemplo, na classificação de pessoas de pele escura como pessoas de “alto risco” duas vezes maior do que em comparação aos réus de pele clara (Cozman, Kaufman, 2022, p. 203). Trata-se, por óbvio, de um nítido viés algorítmico que evidencia os perigos de se automatizar o juízo de valor jurídico sem o devido filtro ético e constitucional.

Além da análise preditiva e das funções acima citadas, outra aplicabilidade dos sistemas inteligentes no direito se refere à identificação e aplicação de precedentes em processos instaurados - função essa amplamente explorada pelas instituições brasileiras. Um exemplo de tal tecnologia é a ferramenta “VICTOR”. Originado de uma parceria entre a Universidade de Brasília (UNB) e o Supremo Tribunal Federal (STF), o programa tem por objetivo a leitura de recursos extraordinários e a indicação de temas de repercussão geral que estão vinculados com estes, de modo a dar maior celeridade ao trâmite processual (Gaio Júnior, Silva, 2022, p. 76).

No Judiciário brasileiro, destaca-se também os sistemas “Sócrates” e “Athos”, utilizados pelo Superior Tribunal de Justiça (Ernesto, Pinto, 2022, p. 928). A primeira ferramenta, “Sócrates”, tem sua atuação pautada na análise de recursos e acórdãos recorridos, de modo a identificar demandas repetitivas e fornecer informações essenciais dos processos aos Ministros, facilitando-lhes o trabalho (Ernesto, Pinto, 2022, p. 928). Já o sistema “Athos” atua, principalmente, com foco na identificação e agrupamento de temas repetitivos - com o intuito de fortalecer a formação de precedentes qualificados no Tribunal (Ernesto, Pinto, 2022, p. 929).

Entretanto, a proliferação dessas ferramentas no ecossistema de justiça impõe um paradoxo que precisa ser enfrentado: o risco de que a busca pela eficiência operacional acabe por sacrificar a inteligibilidade do processo. À medida que algoritmos complexos assumem funções de triagem e auxílio à decisão, ergue-se o perigo da “opacidade tecnológica”, onde a racionalidade jurídica é substituída por cálculos probabilísticos inacessíveis às partes. Nesse cenário, a modernização do Judiciário não pode significar o abandono das garantias constitucionais; pelo contrário, exige que a tecnologia seja submetida aos mesmos rigores de publicidade e fundamentação que regem os atos humanos, sob pena de tornar o Direito refém de uma lógica matemática inaudível.

Assim sendo, diante da visível difusão de tais sistemas na esfera judicial, viu-se necessária a determinação de princípios que visassem regular a aplicação destas tecnologias, de modo a proteger os direitos individuais daqueles que de alguma forma estão envolvidos

com as atividades jurídicas. Dentre as regras que visam orientar a introdução de sistemas de inteligência artificial no direito, duas se destacam: a transparência e a interpretabilidade, a serem agora explicadas.

#### **4. A Transparência e a Explicabilidade como princípios norteadores da aplicação da Inteligência Artificial no Direito.**

No que se refere ao uso de sistemas de inteligência artificial no Direito, uma preocupação vem à tona: a opacidade algorítmica. Trata-se, de modo geral, da impossibilidade de se entender como determinada ferramenta de inteligência artificial gerou seus resultados; ou seja, desconhece-se os caminhos pelos quais o sistema chegou a suas conclusões (Pecego, Teixeira, 2024, p. 6).

Dentre as causas de tal opacidade jurídica, destacam-se a existência de sistemas proprietários - ou seja, que apresentam o código fonte protegido, não disponível à população de modo geral; como também a complexidade inerente a tais sistemas que, mesmo que abertos à consulta pública, necessitam de um profundo conhecimento técnico sobre o tema para sua compreensão (Pecego, Teixeira, 2024, p. 6) - fato que destoia da publicidade exigida pelo ordenamento no que se refere à atividade jurídica.

Tal conflito, inclusive, não passou despercebido pela comunidade internacional. A Recomendação sobre a Ética da Inteligência Artificial da UNESCO (UNESCO, 2021), ratificado pelo Brasil e outros 92 países, neste ponto, erigiu expressamente a transparência e a explicabilidade como princípios éticos indispensáveis aos sistemas inteligentes, *in verbis*:

37. Com frequência, a transparência e a explicabilidade dos sistemas de IA são pré-requisitos essenciais para garantir o respeito, a proteção e a promoção dos direitos humanos, das liberdades fundamentais e dos princípios éticos. A transparência é necessária para que os regimes de responsabilidade nacionais e internacionais pertinentes funcionem de forma efetiva. A falta de transparência também pode prejudicar a possibilidade de se contestar efetivamente as decisões com base nos resultados produzidos pelos sistemas de IA e, com isso, infringir o direito a um julgamento justo e a um recurso eficaz, além de limitar as áreas em que esses sistemas podem ser utilizados de maneira legal (UNESCO, 2021, p. 22).

Sobre a explicabilidade, o referido documento internacional explicita:

40. A explicabilidade significa tornar inteligível e fornecer informações sobre o resultado dos sistemas de IA. A explicabilidade desses sistemas também se refere à compreensibilidade sobre a entrada, a saída e o funcionamento de cada bloco de construção dos algoritmos e como ele contribui para o resultado dos sistemas.

Assim, a explicabilidade está estreitamente relacionada com a transparência, pois os resultados e os subprocessos que conduzem a resultados devem almejar serem compreensíveis e rastreáveis, conforme o contexto. Os atores de IA devem se comprometer a garantir que os algoritmos desenvolvidos sejam explicáveis. No caso de aplicativos de IA que afetam o usuário final de uma forma que não seja temporária, facilmente reversível ou de baixo risco, deve-se garantir que seja fornecida uma explicação significativa para qualquer decisão que resultou na ação tomada, para que o resultado seja considerado transparente (UNESCO, 2021, fls. 23).

Sobre a transparência e a explicabilidade, inclusive, a Recomendação sobre a Ética da Inteligência Artificial da UNESCO (UNESCO, 2021) determina a necessidade de um papel ativos dos Estados-membros na sua consecução, como se vê:

70. Os Estados-membros devem definir claramente os requisitos de transparência e explicabilidade dos sistemas de IA, de modo a ajudar a assegurar a confiabilidade de todo o ciclo de vida desses sistemas. Tais requisitos devem envolver o planejamento e a implementação de mecanismos de impacto que considerem a natureza do domínio do aplicativo, o uso pretendido, o público-alvo e a viabilidade de cada sistema de IA específico (UNESCO, 2021, fls. 29).

Neste sentido, buscar-se-á expor, ainda que em linhas gerais, como o princípio da transparência e da interpretabilidade são essenciais na mitigação de tais preocupações e indispensáveis na manutenção da segurança jurídica frente à aplicação de tais tecnologias no âmbito judicial.

#### **4.1 Princípio da Transparência**

O direito de acesso à informação pode ser entendido como um dos pilares do Estado Democrático de Direito, sendo, inclusive, assegurado por tratados internacionais dos quais o Brasil é signatário, bem como pela própria legislação pátria.

No plano infraconstitucional, essa diretriz é materializada precipuamente pelo Código de Processo Civil (Brasil, 2015), que, em seu art. 11, erigiu a publicidade à categoria de norma fundamental do processo, determinando que todos os julgamentos e atos processuais sejam públicos, sob pena de nulidade. Indissociável dessa exigência, o mesmo dispositivo legal impõe o dever de motivação (ou fundamentação) das decisões judiciais, de forma a tornar o mais transparente possível a decisão exarada no processo, possibilitando ao jurisdicionado entender as razões de tal decisão ter sido tomada de certo modo.

Na esfera criminal, a lógica é a mesma. O Código de Processo Penal (Brasil, 1941), em seu art. 792, estabelece a publicidade das audiências, sessões e dos atos processuais como

regra. Ademais, o dever de motivação também se impõe com rigor máximo na seara penal como estabelecido no art. 381, inciso III, do CPP (Brasil, 1941), onde o magistrado é obrigado a externar as razões de fato e de direito que formaram o seu convencimento, no que diz respeito às sentenças criminais.

Dessa simbiose entre publicidade e motivação emerge, portanto, a verdadeira transparência dos atos processuais. Não se trata apenas de permitir que o cidadão assista a um julgamento, mas de garantir que ele compreenda os motivos da decisão tomada. Enquanto elemento essencial de um regime democrático, tal direito apresenta uma importante função na manutenção da crença e confiabilidade nas instituições públicas por parte da população, já que dá a esta o poder de fiscalização da máquina pública e seu funcionamento.

Na didática de Ingo Wolfgang Sarlet (Sarlet, Marinoni, Mitidiero, 2025, p. 493), o direito de acesso à informação - diverso do direito à informação (o direito, em si, de ser informado) e da liberdade de informação (que seria o direito de informar), trata-se da capacidade do indivíduo de acessar determinadas informações consideradas essenciais a um Estado Democrático de Direito e ao resguardo de outros direitos fundamentais. É, portanto, um pressuposto indispensável à própria transparência da administração pública e da jurisdição, instrumentalizando o controle social do poder e permitindo que o cidadão fiscalize a legitimidade dos atos estatais. Sob este norte, argumenta o citado autor:

Ademais disso, calha sublinhar que o dever constitucional de transparência, publicidade e informação, além de permitir o controle social, imprescindível a um Estado Democrático de Direito, assegura (ou facilita) a fruição (e proteção) de outros direitos fundamentais (não apenas, mas em especial os direitos de participação política e de crítica, e também os direitos sociais), pois o acesso a informações atuais, corretas e completas, por exemplo, em matéria orçamentária, permite fiscalizar a destinação de recursos para os fins constitucionalmente previstos, como é o caso dos direitos à saúde e à educação, os quais inclusive têm assegurado investimento mínimo de recursos públicos (Sarlet, Marinoni, Mitidiero, 2025, p. 496).

No que se refere às tecnologias de inteligência artificial, a questão da falta de transparência vem se tornando um tema cada vez mais debatido, principalmente nas discussões acerca de seus riscos relacionados à questão da reprodução, por tais tecnologias, preconceitos sociais (chamados, no estudo das inteligências artificiais, de “vieses algorítmicos”). É neste sentido que o acesso às informações do sistema inteligente se mostra indispensável”.

Tal acesso à configuração dos sistemas inteligentes, principalmente aqueles voltados ao Direito e utilizados pelo Poder Público, mostra-se essencial para que terceiros possam

avaliar tais tecnologias em sua dimensão ética, bem como para que seja possível a mensuração de seus impactos sociais e jurídicos e a identificação de possíveis vieses (Maranhão, Junquilho, Tasso, 2023, p. 150). Entende-se, desse modo, que a transparência de tais algoritmos é parte do direito de acesso à informação (Vestri, 2021, p. 381).

Quando aplicados ao direito, a abertura destes sistemas inteligentes se mostra ainda mais importante, tendo em vista que estas tecnologias serão aproveitadas em um terreno extremamente sensível e de alto impacto social, devendo-se garantir, portanto, a acessibilidade aos cidadãos para conhecerem o modo como o algoritmo toma as decisões (Vestri, 2021, p. 382). A transparência quanto à operação do sistema inteligente aplicado ao direito se dirige aos próprios usuários do sistema, para que estes possam entender quais os critérios que levaram a IA a tomar determinada decisão; mas, por ser aplicada no âmbito do Direito Público, deve ir além dos jurisdicionados. A abertura do sistema se dirige, também, à sociedade como um todo, submetendo tais sistemas à opinião pública; aos magistrados, para que estes aproveitem da melhor forma estes sistemas e, por fim, aos próprios órgãos que possam auxiliar na regulação e acompanhamento destas tecnologias, potencializando suas utilidades e evitando possíveis riscos (Maranhão, Junquilho, Tasso, 2023, p. 150).

Trata-se, portanto, de um ir além da fundamentação dos atos jurisdicionais. É possibilitar ao jurisdicionado acessar a "gramática" interna do algoritmo, permitindo que a motivação das decisões judiciais não seja apenas lida, mas efetivamente auditada. É necessário que o jurisdicionado saiba, por exemplo, quais foram os critérios adotados pelo sistema para tomar determinada decisão, se os dados utilizados pela ferramenta não estão enviesados ou se, a partir dos critérios adotados, a Inteligência Artificial em uso é capaz de tomar decisões livres de preconceitos sociais (Ernesto, Pinto, 2022, p. 931).

Nesse sentido, a opacidade algorítmica não representa apenas uma limitação técnica, mas um verdadeiro obstáculo ao exercício do contraditório e da ampla defesa. Conhecer o arcabouço estrutural e informacional do sistema responsável por determinada decisão é condição indispensável para que a parte interessada possa exercer seu direito de resistência à decisão prolatada de maneira efetiva - uma vez que não é possível refutar decisões em que suas premissas são invisíveis ou inacessíveis.

Um clássico exemplo acerca da falta de tal transparência é o já citado caso COMPAS - sistema de inteligência artificial, utilizado no Estado de Nova York, para decisões relativas à concessão, ou não, de liberdade condicional aos réus. O sistema referido é denominado de "sistema proprietário": o algoritmo utilizado para se determinar se o réu será posto em liberdade condicional não é público; tratando-se de uma propriedade da empresa criadora do

sistema - de modo a não permitir que o réu saiba quais foram os motivos, por exemplo, da recusa de sua liberdade condicional (Angwin *et al.*, 2016 *apud* Alves, Andrade, 2022, p. 356). Tal problemática alcançou a Suprema Corte do Estado de Wisconsin (caso *State vs. Loomis*), que tornou clara a preocupação com a utilização de algoritmos de sentenciamento dotados de opacidade (Pecego, Teixeira, 2024, p. 9).

#### **4.1.1 Da insuficiência da Transparência para a compreensão da estrutura algorítmica**

Em que pese a necessidade da disponibilização de informações acerca do funcionamento dos sistemas inteligentes aplicados ao Direito, de modo a assegurar a segurança jurídica e o controle democrático de tais tecnologias pela sociedade civil; tem-se, ainda, a necessidade de exposição de tais métodos de funcionamento de forma clara e compreensível à população em geral - visto que a mera abertura do código fonte de tais ferramentas se mostra insuficiente para garantir a transparência efetiva (Ferrari, 2018, p. 11).

A simples exibição de milhares de linhas de código ou a exposição de pesos matemáticos e parâmetros estatísticos não revela, por si só, o "porquê" de um resultado específico. Para um cidadão comum, ou mesmo para um magistrado, a visualização de uma arquitetura algorítmica complexa é, de fato, ininteligível. O dado técnico está presente, mas o sentido lógico da inferência permanece oculto. Assim, a transparência meramente formal acaba por gerar uma "ilusão de publicidade", onde o acesso à estrutura não se traduz em compreensão sobre os critérios de justiça aplicados ao caso concreto.

Não se trata, no entanto, de ignorar o valor da publicidade dos algoritmos, mas de reconhecer que a transparência real exige uma ponte entre a complexidade computacional e o entendimento jurídico. A eficácia do controle social depende de uma "explicabilidade" que permita ao jurisdicionado compreender os motivos e a lógica por trás de uma recomendação ou decisão automatizada. Sem esse esforço de tradução e clareza, corre-se o risco de substituir o antigo hermetismo do "juridiquês" por um nova e impenetrável linguagem técnica, mantendo o cidadão à margem dos processos que definem seus direitos e deveres.

#### **4.2 Princípio da explicabilidade**

Os algoritmos podem ser entendidos como uma ordenação de regras que definem como se deve resolver determinado problema (Teixeira, 1998, *apud* Gaio Júnior, Silva, 2022, p. 66). No que se refere à inteligência artificial, seus algoritmos, por combinarem matérias

como cálculo e raciocínio, se mostram extremamente complexos, inclusive pela grande quantidade de variáveis codificadas e dados analisados (Caplan *et al*, 2018, apud Maranhão, Florêncio, Almada, 2021, p. 158).

Nesta senda, tem-se que a mera disponibilização do código fonte de tais sistemas inteligentes se mostra ineficiente frente à complexidade matemática que envolve seu funcionamento; transformando tais algoritmos em verdadeiras “caixas-pretas”, em que a observação do *output* dado pelo sistema mostra-se incapaz de permitir a um ser humano a compreensão dos processos internos que conduziram os *inputs* (ou seja, aquilo que foi fornecido ao sistema pelo usuário) até a conclusão fornecida (Ferrari, 2018, p. 11).

Por esta razão, a disponibilização de informações compreensíveis acerca do funcionamento da tecnologia de inteligência artificial aplicada ao direito se mostra de suma valia em um Estado Democrático de Direito que venha a optar por sua utilização, de forma a evitar, por exemplo, possíveis vieses algorítmicos na execução das funções devidas pelo sistema.

Tais vieses algorítmicos apresentam, de modo geral, três origens principais: na coleta dos dados a serem utilizados pelo sistema como base de funcionamento (que deve ser representativa do todo e livre de preconceitos sociais), na preparação destes mesmos dados (quando, já coletados, devem ser selecionados, pelo desenvolvedor, os atributos a serem considerados em tais dados pelo algoritmo) e, por fim, na definição do próprio objetivo a ser alcançado pelos algoritmos (Hao, 2019, *apud* Cozman, Kaufman, 2022, p. 200).

Quanto à estruturação da própria ferramenta inteligente, é imperativo considerar a influência subjetiva do fator humano no desenvolvimento do código. A suposta neutralidade da tecnologia é mitigada pelas escolhas conscientes ou inconscientes dos programadores, que podem transferir seus próprios preconceitos e visões de mundo para a arquitetura do sistema (Cozman, Kaufman, 2022, p. 201).

Um exemplo nítido dessa influência pode ocorrer no desenvolvimento de algoritmos de reconhecimento facial para monitoramento de segurança pública. Imagine que um programador responsável pela arquitetura do sistema, influenciado por um preconceito estrutural inconsciente, decida que o algoritmo deve ser "mais sensível" ou "rigoroso" ao identificar padrões de vestimenta ou comportamentos que ele associa subjetivamente à criminalidade (como o uso de certas peças de roupa comuns em periferias). O sistema, então, passa a gerar alertas de "atitude suspeita" com muito mais frequência para indivíduos com essas características, não por uma correlação estatística real com o crime, mas porque a lógica interna da ferramenta foi calibrada sob a ótica preconceituosa do desenvolvedor. Dessa forma,

a tecnologia automatiza e escala a discriminação, escondendo sob um manto de objetividade matemática um preconceito humano que foi injetado diretamente na estrutura do código.

Tal cenário não é meramente ilustrativo, visto que a materialização desses vieses algorítmicos já foi documentada em casos concretos. Um exemplo paradigmático reside no sistema de triagem de recrutamento automatizado da Amazon, implementado em 2014 e descontinuado no início de 2017, visto que a ferramenta demonstrou não ser neutra em termos de gênero, passando a privilegiar sistematicamente candidatos homens (Cozman, Kaufman, 2022, p. 198). Essa distorção ocorreu porque os algoritmos de IA foram treinados para identificar padrões em uma base de dados de currículos enviados à empresa ao longo de uma década, consideravelmente marcada pelo domínio masculino na indústria de tecnologia (Cozman, Kaufman, 2022, p. 198); fato que se reproduziu no funcionamento do sistema inteligente.

O grande impasse, portanto, reside no fato de que essas distorções, sejam elas oriundas de uma coleta de dados falha ou de escolhas subjetivas de estruturação da tecnologia, permanecem ocultas nas camadas profundas da "caixa-preta", tornando o veredito da máquina inaudível sob a ótica do devido processo legal.

Nesse contexto, a superação do fenômeno da "caixa-preta" exige a transição da mera transparência formal para uma explicabilidade efetiva. Não basta que o Estado ou as empresas forneçam o emaranhado de cálculos matemáticos que compõem o modelo; é indispensável que se apresente uma narrativa compreensível sobre os critérios de relevância adotados pela máquina. Sem que se conheça o "peso" atribuído a cada variável no processo de tomada de decisão, o jurisdicionado encontra-se em uma posição de vulnerabilidade informacional, na qual o contraditório se torna impossível; afinal, não se pode contestar uma conclusão cujas premissas permanecem ocultas sob o argumento do segredo comercial ou da complexidade tecnológica.

Ademais, a utilização de sistemas inteligentes no âmbito do Direito provoca um deslocamento crítico no ônus da prova e na presunção de veracidade. Quando um magistrado fundamenta suas ações em um *output* algorítmico, a natureza pretensamente objetiva do sistema pode conferir à decisão um aspecto de infalibilidade, dificultando a revisão humana. Se a estrutura do código contém vieses implícitos, como no exemplo do reconhecimento facial, a inversão da lógica democrática é completa: o indivíduo passa a ter o ônus quase impossível de provar que a "lógica matemática" do sistema está enviesada. Portanto, a implementação da inteligência artificial deve vir acompanhada de mecanismos rígidos de auditoria externa e supervisão humana; garantindo que a tecnologia sirva como suporte à

justiça, e não como um novo instrumento de perpetuação de desigualdades históricas - e, neste sentido, os princípios da transparência e explicabilidade se mostram indispensáveis.

Tal exigência da explicabilidade foi recentemente externada pelo Conselho Nacional de Justiça (CNJ) através da Resolução de n.º 615/2025 (Brasil, 2025), que visa possibilitar a inserção de tecnologias de Inteligência Artificial na atuação jurisdicional, sempre que em benefício dos jurisdicionados (Tamer, 2025, p. 22).

A referida Resolução, em seu art. 3º, inciso II, elevou a explicabilidade à categoria de princípio regente, definindo-a expressamente, em seu art. 4º, inciso XVIII, como a garantia de "compreensão clara, sempre que tecnicamente possível, de como as decisões são tomadas pela IA" (Brasil, 2025). Dessa forma, impôs-se, com esta Resolução, uma barreira intransponível ao uso de sistemas inteligentes "ininteligíveis" no Judiciário, que impedem o jurisdicionado de entender o caminho, tomado pela tecnologia, para a tomada de certa decisão no bojo de seu processo, convertendo o processo em mera adesão à resultados obscuros, dificultando, por isso mesmo, o exercício do contraditório.

Outrossim, tal diretriz de inteligibilidade não é exclusividade do debate brasileiro, encontrando-se também prevista em instrumentos normativos internacionais. O Regulamento de Inteligência Artificial da União Europeia (União Europeia, 2024) institui, em seu art. 86, que o jurisdicionado, quando afetado por uma decisão tomada com base em sistemas inteligentes de alto risco, pode exigir uma explicação clara e significativa sobre o papel da IA e os parâmetros lógicos adotados. Tal previsão ratifica a tese de que, no Estado Democrático de Direito, a eficiência tecnológica não pode sobrepor-se à garantia fundamental de compreender a racionalidade das decisões estatais.

Por fim, no âmbito jurídico, outros autores apontam também a importância da chamada "*Explainable Artificial Intelligence (XAI)*" (em português: "inteligências artificiais explicáveis"); que, basicamente, são tecnologias de inteligência artificial capazes de explicarem seus resultados (Alves, Andrade, 2022, p. 361). Ou seja, além da exposição inteligível de sua estrutura ao público em geral, tais sistemas estariam habilitados, portanto, a fornecer explicações razoáveis de suas ações para seus usuários (de acordo com seu público-alvo, que nas aplicações jurídicas é bem amplo), o que, sem dúvida, tornaria suas execuções mais compreensíveis àqueles que o utiliza (Alves, Andrade, 2022, p. 361).

## 5. CONCLUSÃO

O presente trabalho buscou evidenciar a importância da transparência e da

explicabilidade como pilares inegociáveis para a integração da inteligência artificial no ordenamento jurídico brasileiro. Como observado, a transição para um Direito cada vez mais digitalizado - enquanto consequência de um mundo também cada vez mais digital; não é apenas uma questão de celeridade processual ou eficiência administrativa; é, primordialmente, um desafio de preservação das garantias fundamentais e do Estado Democrático de Direito.

A análise do fenômeno da "caixa-preta" e dos vieses algorítmicos - exemplificados por casos como o sistema COMPAS e do sistema de triagem de currículos da Amazon - demonstrou que a neutralidade tecnológica é um mito. Por trás da precisão matemática, residem escolhas humanas, bases de dados historicamente viciadas e códigos protegidos por segredos comerciais que, se não submetidos ao crivo da publicidade e da compreensão, podem institucionalizar injustiças e inviabilizar o exercício do contraditório e da ampla defesa.

Conclui-se, portanto, que a transparência formal (acesso ao código) é insuficiente se não for acompanhada pela explicabilidade material (capacidade de traduzir a lógica algorítmica em termos jurídicos e humanos). Para que sistemas como o VICTOR e outras ferramentas de análise preditiva cumpram seu papel de auxiliar o Poder Judiciário sem comprometer a segurança jurídica, é imperativo que o jurisdicionado não seja um mero objeto de processamento de dados, mas um sujeito de direitos capaz de compreender e, se necessário, contestar as premissas de uma decisão automatizada.

Em última análise, a inteligência artificial deve servir como um instrumento de ampliação do acesso à justiça, e não como uma nova barreira tecnocrática. O futuro do Direito na era digital depende da nossa capacidade de garantir que, embora as máquinas possam auxiliar no processamento da norma, a palavra final e a fundamentação ética permaneçam sob o domínio da inteligibilidade humana e do rigor constitucional.

### **Referências bibliográficas:**

ALVES, Marco Antônio Sousa; ANDRADE, Otávio Morato. Da “Caixa-Preta” à “Caixa de Vidro”: o Uso da Explainable Artificial Intelligence (XAI) para Reduzir a Opacidade e Enfrentar o Enviesamento em Modelos Algorítmicos. **Revista de Direito Público**, Brasília, v. 18, n. 100, p. 349-373, out./dez. 2021. Disponível em: <https://doi.org/10.11117/rdp.v18i100.5973>. Acesso em: 08 fev. 2026.

BRASIL. Conselho Nacional de Justiça. **Resolução nº 615**, de 11 de março de 2025.

Estabelece diretrizes para governança de soluções desenvolvidas com recursos de inteligência artificial no Poder Judiciário. Disponível em: <https://atos.cnj.jus.br/files/original1555302025031467d4517244566.pdf>. Acesso em: 10 fev. 2026.

BRASIL. Decreto-Lei nº 3.689, de 3 de outubro de 1941. **Código de Processo Penal**. Diário Oficial da União, Rio de Janeiro, 13 out. 1941. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/decreto-lei/del3689.htm](http://www.planalto.gov.br/ccivil_03/decreto-lei/del3689.htm). Acesso em: 08 fev. 2026.

BRASIL. Lei nº 13.105, de 16 de março de 2015. **Código de Processo Civil**. Diário Oficial da União, Brasília, DF, 17 mar. 2015. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2015/lei/l13105.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13105.htm). Acesso em: 08 fev. 2026.

COPELAND, Jack. **Inteligencia artificial: Una introducción filosófica**. Madrid: Alianza Editorial, 1993.

COZMAN, Fabio Gagliardi; KAUFMAN, Dora. Viés no aprendizado de máquina em sistemas de inteligência artificial: a diversidade de origens e os caminhos de mitigação. **Revista USP**, São Paulo, n. 135, p. 195–210, out/nov/dez. 2022. Disponível em: <https://doi.org/10.11606/issn.2316-9036.i135p195-210>. Acesso em: 08 fev. 2026.

DE SANCTIS, Fausto Martin. **Inteligência Artificial e Direito**. 1 ed. São Paulo: Almedina, 2020.

ERNESTO, Leandro Miranda; PINTO, Henrique Alves. INTELIGÊNCIA ARTIFICIAL APLICADA AO DIREITO: POR UMA QUESTÃO DE ÉTICA. **Revista Jurídica Luso-Brasileira**, Lisboa, n. 6, ano 08 (2022), 2022. Disponível em: <https://www.cidp.pt/publicacao/revista-juridica-lusobrasileira-ano-8-2022-n-6/281>. Acesso em: 08 fev. 2026.

FERRARI, Isabela. Accountability de algoritmos: a falácia do acesso ao código e caminhos para uma explicabilidade efetiva. In: **INTELIGÊNCIA ARTIFICIAL: 3º Grupo de**

**Pesquisa do ITS.** Rio de Janeiro: ITS Rio, 2018. Disponível em: <https://itsrio.org/pt/publicacoes/inteligencia-artificial-gp3/>. Acesso em: 12 fev. 2026.

FREITAS, Cinthia Obladen de Almendra; BARDDAL, Jean Paul. Análise preditiva e decisões judiciais: controvérsia ou realidade? **Revista Democracia Digital e Governo Eletrônico**, Florianópolis, v. 1, n. 18, p. 107-126, 2019. Disponível em: <https://buscalegis.ufsc.br/revistas/index.php/observatoriodoegov/article/view/314>. Acesso em: 08 fev. 2026.

GAIO JUNIOR, Antônio Pereira; SILVA, Fábila Antonio. DIREITO, PROCESSO E INTELIGÊNCIA ARTIFICIAL. DIÁLOGOS NECESSÁRIOS AO EXERCÍCIO DA JURISDIÇÃO. **Revista Eletrônica de Direito Processual**, Rio de Janeiro, v. 24, n. 1, jan/abr 2023. Disponível em: <https://doi.org/10.12957/redp.2023.72240>. Acesso em: 8 fev. 2026.

MAGALHÃES, Renato Vasconcelos. Inteligência Artificial e Direito - Uma breve introdução história. **Revista Direito e Liberdade**, Mossoró, v. 1, n. 1, p. 355-370, jul/dez. 2005. Disponível em: <https://doi.org/10.53798/suprema.2021.v1.n1.a20>. Acesso em: 08 fev. 2026.

MARANHÃO, Juliano Souza de Albuquerque; FLORÊNCIO, Juliana Abrusio; ALMADA, Marco. Inteligência artificial aplicada ao direito e o direito da inteligência artificial. **Suprema: Revista de Estudos Constitucionais**, Distrito Federal, v. 1, n. 1, p. 154-180, 2021. Disponível em: <https://doi.org/10.53798/suprema.2023.v3.n2.a231>. Acesso em: 10 fev. 2026.

MARANHÃO, Juliano Souza de Albuquerque; JUNQUILHO, Tainá Aguiar; TASSO, Fernando Antônio. Transparência sobre o emprego de inteligência artificial no Judiciário: um modelo de governança. **Suprema: Revista de Estudos Constitucionais**, Brasília, v. 3, n. 2, p. 145-187, jul./dez. 2023. Disponível em: <https://suprema.stf.jus.br/index.php/suprema/article/view/231>. Acesso em: 08 fev. 2026.

ORGANIZAÇÃO DAS NAÇÕES UNIDAS PARA A EDUCAÇÃO, A CIÊNCIA E A CULTURA (UNESCO). **Recomendação sobre a Ética da Inteligência Artificial**. Paris: UNESCO, 2021. Disponível em: [https://unesdoc.unesco.org/ark:/48223/pf0000381137\\_por](https://unesdoc.unesco.org/ark:/48223/pf0000381137_por).

Acesso em: 09 fev. 2026.

PECEGO, Daniel Nunes; TEIXEIRA, Raphael Lobato Collet Janny. INTELIGÊNCIA ARTIFICIAL NO JUDICIÁRIO: DA OPACIDADE À EXPLICABILIDADE DAS DECISÕES JUDICIAIS. **Revista da Faculdade de Direito da UERJ**, Rio de Janeiro, n. 43, p. 1–22, 2024. Disponível em: <https://doi.org/10.12957/rfd.2024.87850>. Acesso em: 18 jan. 2026.

RUSSEL, Stuart J.; NORVIG, Peter. **Inteligência Artificial**. 3 ed. Rio de Janeiro: Elsevier, 2013.

SARLET, Ingo W.; MARINONI, Luiz G.; MITIDIERO, Daniel. **Curso de Direito Constitucional**. 14. ed. Rio de Janeiro: Saraiva Jur, 2025. E-book. ISBN 9788553626885. Disponível em: <https://integrada.minhabiblioteca.com.br/reader/books/9788553626885/>. Acesso em: 08 fev. 2026.

SOARES, Marcelo Negri; KAUFFMAN, Marcos Eduardo; CHAO, Kuo-Ming. INTELIGÊNCIA ARTIFICIAL: IMPACTOS NO DIREITO E NA ADVOCACIA. **Revista Direito Público**, Brasília, v. 17, n. 93, p.104-133, maio/jun 2020. Disponível em: <https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3555>. Acesso em: 8 fev. 2026.

TAMER, M. A. Regulação da inteligência artificial no Judiciário Brasileiro: análise do objeto e definições da Resolução CNJ nº 615/2025. **Revista do Tribunal Regional Federal da 3ª Região**, São Paulo, v. 36, n. 162, p. 17–36, 2025. Disponível em: <https://doi.org/10.65674/rev-trf3.v36i162.742>. Acesso em: 10 fev. 2026.

TURING. Alan Mathison. Computing machinery and intelligence. **Mind**, v. 59, n. 236, p. 433-460, out. 1950.

UNIÃO EUROPEIA. Parlamento Europeu e Conselho. **Regulamento (UE) 2024/1689, de 13 de junho de 2024**. Estabelece regras harmonizadas em matéria de inteligência artificial e altera os Regulamentos (CE) n.º 300/2008, (UE) n.º 167/2013, (UE) n.º 168/2013, (UE)

2018/858, (UE) 2018/1139 e (UE) 2019/2144 e as Diretivas 2014/90/UE, (UE) 2016/797 e (UE) 2020/1828 (Regulamento da Inteligência Artificial). Jornal Oficial da União Europeia, Bruxelas, L, 2024/1689, 12 jul. 2024. Disponível em: <https://artificialintelligenceact.eu/article/86/>. Acesso em: 10 fev. 2026.

VESTRI, Gabriele. La inteligencia artificial ante el desafío de la transparencia algorítmica: Una aproximación desde la perspectiva jurídico-administrativa. **Revista Aragonesa de Administración Pública**, Zaragoza, n. 56, p. 368-398, 2021. Disponível em: <https://doi.org/10.71296/raap.62>. Acesso em: 08 fev. 2026.

ZILIO, Diego. Inteligência artificial e pensamento: redefinindo os parâmetros da questão primordial de Turing. **Ciências & Cognição**, Rio de Janeiro, v. 14, n. 1, p. 208-218, mar. 2009. Disponível em: [https://pepsic.bvsalud.org/scielo.php?script=sci\\_arttext&pid=S1806-58212009000100013](https://pepsic.bvsalud.org/scielo.php?script=sci_arttext&pid=S1806-58212009000100013). Acesso em: 08 fev. 2026.