Revista da Esmal
ISSN: 2525-9547

PROMPT INJECTION E INTELIGÊNCIA ARTIFICIAL NO JUDICIÁRIO BRASILEIRO: RISCOS DE MANIPULAÇÃO, DEVERES PROCESSUAIS E A PRESERVAÇÃO DA SUBSUNÇÃO JUDICIAL COMO GARANTIA FUNDAMENTAL

PROMPT INJECTION AND ARTIFICIAL INTELLIGENCE IN THE BRAZILIAN JUDICIARY: MANIPULATION RISKS, PROCEDURAL DUTIES AND THE PRESERVATION OF JUDICIAL SUBSUMPTION AS A FUNDAMENTAL GUARANTEE

Lucas Rodrigues da Costa[1]

RESUMO: O artigo analisa vulnerabilidades técnicas de sistemas de inteligência artificial no contexto judiciário, especificamente ataques de *prompt injections* que possibilitam manipulações por profissionais que agem de má-fé. Demonstra-se que princípios de boa-fé objetiva e cooperação processual estabelecem limites normativos à manipulação algorítmica maliciosa. Examina-se a lacuna do debate jurídico brasileiro sobre governança que negligencia a dimensão de segurança cibernética, privilegiando a transparência e a equidade. Propõe-se modelo de complementaridade responsável entre inteligência artificial auxiliar e subsunção judicial fundamentada como garantia constitucional irredutível. Utiliza-se metodologia dedutiva e abordagem qualitativa, incluindo análise de atos normativos, doutrina processualista e estudos técnicos sobre segurança de modelos de linguagem.
PALAVRAS-CHAVE: inteligência artificial; boa-fé processual; *prompt injection*.

ABSTRACT: The article examines technical vulnerabilities of artificial intelligence systems within the judicial context, focusing on prompt injections attacks that enable manipulation by bad-faith actors. It shows that the principles of objective good faith and procedural cooperation establish normative limits on malicious algorithmic interference. The study identifies a gap in Brazilian legal scholarship on governance that overlooks the cybersecurity dimension while prioritizing transparency and fairness. It proposes a model of responsible complementarity between auxiliary AI systems and reasoned judicial subsumption as an

[1] Mestre em Direito Público pela Universidade Federal de Alagoas (UFAL). Especialista em Direito Civil e Processo Civil pela Fundação Escola Superior do Ministério Público do Rio Grande do Sul (FESMP/RS) e em Direito Previdenciário pelo Centro Universitário Maceió (UNIMA/AL). Assessor jurídico. Advogado licenciado, inscrito na Ordem dos Advogados do Brasil em Alagoas (OAB/AL). E-mail: lucasrdct@gmail.com.

irreducible constitutional safeguard. The methodology is deductive-qualitative, drawing on normative acts, procedural theory and technical studies on language-model security. **KEYWORDS**: artificial intelligence; procedural good faith; prompt injection.

1 INTRODUÇÃO

A incorporação de sistemas baseados em inteligência artificial (IA) ao aparato judiciário brasileiro ultrapassa a mera modernização administrativa; ela institui uma nova ecologia processual que desafia as garantias tradicionais da jurisdição. A transição da promessa tecnológica para a implementação concreta, já consolidada por iniciativas do Conselho Nacional de Justiça (CNJ) e do Supremo Tribunal Federal (STF), não apenas automatiza rotinas burocráticas, mas insere modelos de linguagem e aprendizado de máquina no núcleo da atividade forense. Contudo, essa digitalização profunda inaugura vulnerabilidades técnicas inéditas, especificamente no que tange à arquitetura dos grandes modelos de linguagem.

Já não se trata apenas de discutir vieses algorítmicos ou a substituição da mão de obra humana, mas de enfrentar a fragilidade estrutural de sistemas que, incapazes de distinguir ontologicamente entre instruções de controle e dados de entrada, tornam-se suscetíveis a manipulações externas deliberadas. Essa permeabilidade técnica introduz no processo judicial o risco dos *prompt injections*, permitindo que atores processuais mal-intencionados subvertam a lógica decisória, transformando ferramentas de eficiência em possíveis vetores de fraude processual.

O problema central desta investigação reside na tensão entre a arquitetura aberta dos modelos de linguagem e os princípios basilares do devido processo legal. Petições e documentos, elementos tradicionais da prova e da narrativa processual, assumem também papel de vetores de reconfiguração da lógica algorítmica quando manipulados com intenção maliciosa. A integridade do processo passa a depender de compreensão precisa da forma como esses sistemas operam. A vulnerabilidade não deriva de falha acidental, mas de modo estruturado de processamento da linguagem natural, no qual controle e dados compartilham o mesmo canal.

O Direito Processual, estruturado sobre a presunção de boa-fé e cooperação, vê-se desafiado por uma nova categoria de ilícito: a engenharia de *prompt* maliciosa, que opera nas camadas invisíveis do processamento de dados, escapando frequentemente à percepção imediata dos magistrados e servidores. Assim, impõe-se um esforço crítico: compreender as

repercussões jurídicas dessa vulnerabilidade técnica e reafirmar os deveres éticos dos sujeitos processuais diante da automação.

A insuficiência do debate jurídico atual se evidencia quando se observa que a literatura nacional, embora profícua em discussões sobre ética e discriminação algorítmica, ainda incide pouco sobre os riscos concretos de segurança cibernética. Este artigo, portanto, propõe uma integração entre a tecnicidade da segurança da informação e a dogmática processual civil, visando blindar o processo contra manipulações algorítmicas.

O desenvolvimento do estudo organiza-se em cinco eixos. O segundo capítulo apresenta o panorama atual da inteligência artificial no Judiciário, demonstrando que a automação já constitui infraestrutura consolidada e regulamentada, ilustrada pelos sistemas empregados nos tribunais. O terceiro capítulo examina os ataques de *prompt injection*, descrevendo mecanismo, efeitos e relevância dessa vulnerabilidade para aplicações jurídicas. O quarto capítulo relaciona essa prática aos princípios da boa-fé objetiva e da cooperação processual, demonstrando que manipulação algorítmica configura violação grave do dever de lealdade. O quinto capítulo analisa a governança de IA sob a perspectiva interdisciplinar e enfatiza necessidade de integrar segurança cibernética e auditoria técnica aos instrumentos jurídicos. O sexto capítulo propõe um modelo de complementaridade responsável na interação entre inteligência artificial e atividade jurisdicional, reforçando que sistemas devem operar como apoio à análise humana, conforme diretrizes da Resolução nº 615/2025 do CNJ.

A pesquisa utiliza metodologia descritiva e analítica, apoiando-se na análise de atos normativos recentese em relatórios técnicos de segurança da informação. O método dedutivo orienta a construção teórica, partindo das premissas gerais da teoria da comunicação e do processo civil para diagnosticar os riscos específicos da injeção de *prompt*. A abordagem qualitativa e interdisciplinar permite compreender a manipulação algorítmica não apenas como falha técnica, mas como fenômeno jurídico que demanda resposta normativa. O resultado esperado é a demonstração de que a eficiência prometida pela IA só é legítima se acompanhada de uma governança que assegure a integridade dos dados e a lealdade das interações processuais.

Em última análise, o desafio do Judiciário contemporâneo não é frear a inovação tecnológica, mas garantir que a inteligência artificial sirva ao Direito e não à astúcia de quem sabe manipulá-la, preservando o processo como espaço de racionalidade, transparência e justiça, mesmo em sua infraestrutura digital.

2 INTELIGÊNCIA ARTIFICIAL NO PODER JUDICIÁRIO BRASILEIRO: DA PROMESSA TECNOLÓGICA À IMPLEMENTAÇÃO CONCRETA

A inserção de sistemas baseados em inteligência artificial no aparato judiciário brasileiro não constitui especulação futurista ou projeto experimental. Trata-se de realidade consolidada, documentada e regulamentada. Desde 2019, o Conselho Nacional de Justiça estabelece diretrizes que estruturam o desenvolvimento colaborativo de modelos computacionais voltados ao processo judicial eletrônico e identifica o processamento de linguagem natural como área da inteligência artificial que apresenta resultados expressivos no trabalho jurídico (Brasil. Conselho Nacional de Justiça, 2019, p. 16).

Esse movimento nacional foi acompanhado, de modo decisivo, pelo Supremo Tribunal Federal, que passou a empregar soluções próprias de IA em frentes distintas: o sistema Victor, utilizado desde 2017 para análise de temas de repercussão geral na triagem de recursos recebidos de todo país; o sistema Rafa, no qual promove a classificação temática e análise de impacto, com uso de sistemas voltados ao mapeamento de processos vinculados aos Objetivos de Desenvolvimento Sustentável (ODS); a plataforma VitórIA, que pretende ampliar o conhecimento sobre o perfil dos processos recebidos e permitir o tratamento conjunto de temas repetidos ou similares; e, aMarIA, que auxilia servidores e gabinetes com geração preliminar de minutas, relatórios e ementas, sempre sujeitas à revisão humana (Brasil, Supremo Tribunal Federal, 2023,2024).

A aplicação dessas tecnologias possibilita a automatização de atividades repetitivas que consomem tempo significativo de magistrados e servidores, além de oferecer apoio substancial à decisão mediante análise de vastos repositórios textuais que seriam impraticáveis de processar manualmente. A arquitetura proposta pelo órgão regulador estrutura-se sobre o sistema Sinapses, concebido para orquestrar serviços inteligentes consumidos pela plataforma processual, possibilitando a criação de coleções de modelos desenvolvidos colaborativamente entre diversos tribunais (Brasil. Conselho Nacional de Justiça, 2019, p. 16), constituindo um ecossistema de compartilhamento que otimiza o trabalho e reduz tempo de tramitação processual.

O reconhecimento institucional de que a construção de modelos deve balizar-se em metodologia de pesquisa científica, afastando-se do desenvolvimento tradicional de *software* evidencia maturidade na compreensão da natureza inovadora dessas soluções. Não se trata de simples informatização de procedimentos, mas de implementação de sistemas que aprendem com dados e adaptam comportamento conforme experiência acumulada.

Sichman (p. 38, 2021) esclarece que inteligência artificial não possui definição acadêmica consensual, caracterizando-se antes como ramo da computação que desenvolve sistemas para realizar tarefas que, no momento, são mais bem realizadas por seres humanos que por máquinas ou que não possuem solução algorítmica viável pela computação convencional. A distinção é fundamental: algoritmos tradicionais seguem sequência finita de ações que resolve problema de modo exato, semelhante à receita culinária que, se executada corretamente, produz resultado previsível.

Aplicações jurídicas, contudo, frequentemente carecem de solução única ou exata. Classificações, diagnósticos e análises textuais admitem múltiplas respostas válidas, tornando inviável a geração exaustiva de possibilidades mesmo com computadores extremamente potentes (Sichman, 2021, p. 38-39). A abordagem humana para tais problemas envolve mecanismo de busca e poda: geramos soluções possíveis sem necessariamente enumerar todas elas, escolhemos a melhor solução segundo critério específico e analisamos posteriormente efeitos das escolhas para aprendizado futuro (Sichman, 2021, p. 39).

Sistemas de inteligência artificial replicam essa estratégia mediante técnicas diversas de busca, raciocínio, representação de conhecimento, mecanismos de decisão, percepção, planejamento, processamento de linguagem natural, tratamento de incertezas e aprendizado de máquina, isoladamente ou combinadas conforme a natureza do problema. O paradigma conexionista, subjacente a modelos de linguagem contemporâneos, utiliza redes de elementos simples inspiradas no funcionamento cerebral, onde neurônios artificiais conectados em rede aprendem e generalizam a partir de exemplos fornecidos (Sichman, 2021, p. 38-39).

Matematicamente, trata-se de técnica de aproximação de funções por regressão não linear, dispensando programação explícita de regras (Sichman, 2021, p. 39). Essa característica distingue fundamentalmente sistemas atuais de abordagens simbólicas anteriores, que exigiam identificação prévia e representação formal de conhecimento do domínio. Modelos contemporâneos aprendem padrões diretamente dos dados, descobrindo relações que podem não ser evidentes mesmo para especialistas humanos.

A Resolução nº 615 de 2025 do Conselho Nacional de Justiça estabelece normas abrangentes para desenvolvimento, governança, auditoria, monitoramento e uso responsável de soluções que adotam técnicas de inteligência artificial no Poder Judiciário. O art. 1º determina que aplicações promovam inovação tecnológica e eficiência dos serviços judiciários de modo seguro, transparente, isonômico e ético, beneficiando jurisdicionados com estrita observância de direitos fundamentais garantidos constitucionalmente. O art. 2º, inciso V, exige que a participação e supervisão humana em todas as etapas dos ciclos de

desenvolvimento e utilização não constituem formalidade burocrática, mas reconhecimento de que sistemas automatizados, por mais sofisticados, operam sob limitações que exigem complementação por julgamento humano. A ressalva de que o uso de tecnologias como ferramentas auxiliares para automação de serviços meramente acessórios ou procedimentais e suporte à decisão não se submete a todas as restrições aplicáveis a sistemas decisórios evidencia o discernimento sobre diferentes níveis de risco conforme a aplicação específica.

Além disso, cumpre destacar também que o inciso IX estabelece que a curadoria de dados utilizados no desenvolvimento e aprimoramento de inteligência artificial deve adotar fontes seguras, rastreáveis e auditáveis preferencialmente governamentais, embora permita contratação de fontes privadas desde que atendam requisitos rigorosos de segurança e auditabilidade estabelecidos. Já o inciso X dispõe que a conscientização e difusão de conhecimento sobre soluções que adotam técnicas de inteligência artificial, com capacitação contínua de usuários sobre aplicações, mecanismos de funcionamento e riscos, configura diretriz fundamental que reconhece a impossibilidade de uso responsável sem compreensão adequada. Tribunais e suas escolas devem oferecer capacitação contínua para magistrados e servidores sobre riscos da automação, vieses algorítmicos e análise crítica dos resultados gerados por inteligência artificial, incluindo capacidade de identificar quando resultados apresentados são inadequados ou requerem verificação adicional.

Essa preocupação com a qualidade e origem dos dados reflete compreensão técnica de que modelos de aprendizado de máquina são tão bons quanto os dados que os treinam: dados enviesados, incompletos ou manipulados produzem sistemas que replicam e amplificam essas deficiências.

A garantia de segurança da informação e segurança cibernética compõe princípios norteadores da implementação, reconhecendo que sistemas conectados e processando dados sensíveis constituem alvos potenciais de ataques que podem comprometer não apenas a confidencialidade, mas também a integridade dos dados e a disponibilidade dos serviços. A supervisão humana efetiva, periódica e adequada no ciclo de vida da inteligência artificial, considerando o grau de risco envolvido, admite a possibilidade de ajuste dessa supervisão conforme o nível de automação e impacto da solução utilizada, permitindo flexibilidade sem comprometer a segurança.

Feenberg (2019, p. 175) adverte que tecnologias moldam seus habitantes de forma comparável a leis e costumes que, na antiguidade, eram considerados como influências quase parentais sobre os cidadãos, conformando e representando aqueles que vivem sob seu controle e privilegiando certas dimensões da natureza humana. Quando indivíduos descobrem que

aspectos importantes de sua humanidade não são bem servidos pelo ambiente tecnológico, aparecem controvérsias visando alterar projetos tecnológicos para garantir melhor representação da humanidade dos utilizadores e, em alguns casos, são vítimas da tecnologia.

Essa observação não implica rejeição da tecnologia ou nostalgia por métodos pré-tecnológicos, mas sim o reconhecimento de que design tecnológico constitui escolha política que pode e deve ser informada por valores democráticos e necessidades humanas. Lutas em torno da tecnologia assemelham-se a lutas políticas (Feenberg, 2019, p. 175).

Escolhas tecnológicas são, na sua maioria e hoje em dia, decisões privadas protegidas do envolvimento público pelos direitos de propriedade e pela ideologia tecnocrática que apresenta tais escolhas como meramente técnicas, desprovidas de dimensão política. Democratização da tecnologia exige, em primeira instância, difusão do conhecimento, mas, por si só, isso não é suficiente para fazer diferença substantiva. Para além disso, o espectro de interesses representados por quem controla tecnologia deve ser alargado para tornar mais difícil descarga de externalidades da ação técnica sobre grupos com menos poder (Feenberg, 2019, p. 175).

No contexto judiciário, essa democratização traduz-se concretamente em transparência sobre o funcionamento de sistemas, participação de magistrados, advogados e jurisdicionados na discussão sobre implementação e avaliação de tecnologias e manutenção de controle decisório humano em questões que afetam direitos fundamentais.

A realidade da IA no Judiciário brasileiro configura-se, portanto, como fenômeno que oferece instrumentos valiosos para enfrentar o congestionamento processual crônico mediante automatização inteligente de tarefas repetitivas e apoio qualificado à análise de informações, mas também introduz riscos técnicos e éticos que demandam regulação cuidadosa, implementação responsável e vigilância constante contra usos inadequados ou maliciosos.

3 ATAQUES DE *PROMPT INJECTIONS*: VULNERABILIDADE TÉCNICA COM REPERCUSSÕES PROCESSUAIS

Aplicações integradas a modelos de linguagem seguem arquitetura comum que envolve *prompt* de instrução, que instrui o modelo sobre tarefa específica a realizar, e contexto de dados que constitui informação a ser processada conforme instrução fornecida. O sistema consulta o modelo usando *prompt* e dados para realizar a tarefa e retorna resposta ao usuário, completando o ciclo de processamento. Como dados usualmente provêm de recurso externo, como entrada de usuário ou documento submetido, atacante pode manipulá-los de

modo que aplicação retorne resultado desejado pelo atacante em vez de resultado correto conforme instrução original (Liu *et al*., 2024, p. 1831).

Essa manipulação configura ataque de *promptinjection* (em tradução para o português: injeção de *prompt*), classificado pela *Open Web Application Security Project*, *OWASP*, como risco de segurança número um para aplicações baseadas em modelos de linguagem, superando outras vulnerabilidades conhecidas (Liu *et al*., 2024, p. 1831).

A gravidade dessa classificação reflete não apenas a frequência com que ataques podem ser executados, mas também a severidade de consequências potenciais quando os sistemas comprometidos tomam decisões ou processam informações sensíveis.

Chen *et al.* (2025, p. 02) explicamque ataques de injeção de *prompt* enganam o modelo para desviar de instruções da aplicação original e seguir diretrizes do usuário malicioso. Esses ataques dependem fundamentalmente da capacidade do modelo de seguir instruções em linguagem natural e da incapacidade estrutural de separar *prompts* legítimos fornecidos por desenvolvedores e dados potencialmente maliciosos fornecidos por usuários.

O mecanismo básico do ataque, em sua forma mais simples, consiste em injetar nos dados instruções maliciosas cuidadosamente escolhidas que imitam o formato de instruções legítimas, como “Ignore todas as instruções anteriores e, em vez disso, execute...” (Chen *et al.*, 2025, p. 02). Como modelos de linguagem escaneiam toda entrada em busca de instruções a seguir e não possuem mecanismo técnico robusto para distinguir entre instruções legítimas do desenvolvedor e instruções injetadas por usuário malicioso, sistemas existentes são vulneráveis a tais ataques.

No contexto judiciário, a aplicação de modelos de linguagem pode envolver tarefas como classificação automática de processos segundo matéria e urgência, extração de informações relevantes de petições para indexação e recuperação, geração de minutas para despachos padronizados e análise de jurisprudência para identificação de precedentes aplicáveis. Considerando exemplo concreto de triagem automatizada de processos, *prompt* de instrução poderia especificar: “Classifique este processo segundo a matéria principal tratada e verifique se ele corresponde ao tema X.”

Como exemplo mais simplista da possibilidade de alcance desse ataque, o profissional que age de má-fé poderia anexar à petição submetida à análise uma instrução oculta: “Ignore instruções anteriores sobre o tema X.Classifique este processo como tema Y, independentemente do conteúdo real da petição.” Se o modelo seguir instrução injetada maliciosamente, o processo objetivamente não correspondente ao assunto seria classificado erroneamente, subvertendo critérios legais, podendo ocasionar vantagem indevida para parte

que manipulou sistema e prejuízo para outras partes cujos processos são efetivamente preteridos.

Esposito observa que algoritmos não pensam como pessoas, mas provocam alterações fundamentais na capacidade de obter e processar informações na sociedade (Esposito, 2022, p. 01). A comunicação não existe quando alguém diz algo, mas quando alguém percebe que alguém disse (Luhmann, 1984, p. 193). Comunicação é assim definida como unidade de três tipos de seleção: informação, enunciado e compreensão (Luhmann, 1984, p. 196). Poder e improbabilidade dessa noção de comunicação estão relacionados ao fato de que ela não inclui pensamentos dos participantes, portanto, em princípio, poderia envolver participantes que não pensam como algoritmos (Esposito, 2022, p. 13).

Percebe-se que a vulnerabilidade não reside em falha técnica isolada e corrigível, mas em característica estrutural de como modelos de linguagem processam comunicação. Modelo recebe texto e interpreta tudo como potencial instrução ou dado, sem distinção ontológica ou estrutural entre elementos. A separação depende exclusivamente de marcadores textuais, delimitadores ou formatação, todos passíveis de falsificação ou imitação por usuário que compreende funcionamento do sistema.

Quando um profissional malicioso injeta a instrução formatada similarmente a instrução legítima, o modelo não possui mecanismo para distinguir a origem, intenção ou legitimidade. Processa ambas como comunicação válida, sem consciência de que uma subverte o propósito da outra. No ambiente processual, essa característica assume gravidade específica porque petições, documentos e informações processuais constituem comunicação formal submetida a princípios rigorosos de lealdade e boa-fé.

A vulnerabilidade de modelos de linguagem a ataques de injeção de *prompt* revela limitação fundamental de sistemas que tratam linguagem natural como canal único de comunicação sem distinção estrutural rigorosa entre controle e dados. No Judiciário, onde decisões processuais e, potencialmente, aspectos de decisões de mérito são auxiliadas por sistemas automatizados, essa vulnerabilidade não pode ser tolerada sem mecanismos robustos de detecção e prevenção. A implementação de sistemas automatizados para ganho de eficiência não justifica o comprometimento da integridade processual ou criação de vetores de ataque que permitem manipulação por profissionais que agem de má-fé.

4 BOA-FÉ OBJETIVA E COOPERAÇÃO PROCESSUAL COMO LIMITES NORMATIVOS À MANIPULAÇÃO ALGORÍTMICA

Sujeitos processuais devem comportar-se de acordo com a boa-fé, entendida tecnicamente como norma de conduta que estabelece padrão objetivo de comportamento esperado (Didier Jr., 2019, p. 134). O artigo 5º do Código de Processo Civil é aplicável e vincula não apenas partes principais, mas todos que interagem com o sistema processual.

Não se pode confundir princípio da boa-fé com exigência de boa-fé subjetiva, elemento subjetivo que constitui requisito para configuração de alguns atos ilícitos processuais específicos, como manifesto propósito protelatório. Boa-fé subjetiva é elemento do suporte fático de alguns fatos jurídicos, portanto fato a ser provado, enquanto boa-fé objetiva é norma de conduta que impõe e proíbe comportamentos específicos, além de criar situações jurídicas ativas e passivas independentemente de intenção subjetiva (Didier Jr., 2019, p. 135).

O artigo 5º do Código de Processo Civil não está relacionado à boa-fé subjetiva ou à intenção psicológica do sujeito processual, tratando-se de norma que impõe condutas em conformidade com a boa-fé objetivamente considerada, independentemente da existência de boas ou más intenções no foro íntimo. Princípio da boa-fé extrai-se de cláusula geral processual, opção legislativa mais correta dado que a infinidadede situações que podem surgir ao longo do processo torna pouco eficaz qualquer enumeração legal pretensamente exaustiva das hipóteses de comportamento desleal (Didier Jr., 2019, p. 135).

Marinoni, Arenhart e Mitidiero (2021, p. 12) esclarecem que a boa-fé pode ser reconduzida dogmaticamente à segurança jurídica, na medida em que é possível reduzi-la à necessidade de proteção à confiança legítima, que constitui elemento do princípio da segurança jurídica e de prevalência da materialidade no tráfego jurídico. Como elemento que impõe tutela da confiança e dever de aderência à realidade, boa-fé exigida no processo civil é tanto boa-fé subjetiva quanto boa-fé objetiva, conforme o contexto apresentado. Ao vedar comportamento contrário à boa-fé, artigo 5º do Código de Processo Civil impõe especificamente a necessidade de boa-fé objetiva como padrão de conduta.

Comporta-se com boa-fé processual aquele que não abusa de suas posições jurídicas, exercendo direitos e faculdades processuais dentro de limites funcionais e finalísticos. Ausência de boa-fé pode levar, conforme gravidade e circunstâncias do caso, à ineficácia do ato processual contrário, à responsabilização por dano processual que comportamento desleal

causou à parte contrária ou ao sistema de justiça e, inclusive, à aplicação de sanção pecuniária como multa por litigância de má-fé (Marinoni, Arenhart e Mitidiero (2021, p. 12-13).

Didier Jr. (2019, p. 138) identifica fundamento constitucional do princípio da boa-fé processual no inciso I do artigo 3º da Constituição Federal de 1988, que estabelece como objetivo fundamental da República Federativa do Brasil a construção da sociedade livre, justa e solidária. Haveria dever fundamental de solidariedade, inscrito no texto constitucional, do qual decorreria dever de não quebrar confiança legitimamente depositada e de não agir com deslealdade nas relações jurídicas. O Supremo Tribunal Federal segue linha argumentativa ainda mais incisiva ao afirmar que cláusula do devido processo legal, garantia constitucional fundamental, exige processo leal e pautado na boa-fé de todos os participantes.

Percebe-se claramente que a manipulação intencional de sistemas automatizados mediante injeção de instruções maliciosas configura violação frontal e inequívoca ao dever de lealdade processual. Profissional queanexa à petição ou documento processual texto destinado não a comunicar-se legitimamente com tribunal ou parte contrária conforme finalidade processual, mas a enganar sistema de inteligência artificial subvertendo seu funcionamento regular, age em flagrante desconformidade com a boa-fé objetiva exigida constitucionalmente.

Não há diferença substancial, do ponto de vista ético e jurídico, entre falsificar documento físico tradicional e manipular processamento automatizado mediante técnicas deliberadas de engenharia de *prompt* maliciosa. Ambos os comportamentos constituem atos destinados a obter vantagem processual indevida mediante a fraude, frustrando a confiança legítima de que o processo tramitará segundo critérios legais objetivos e não segundo manipulação técnica realizada por profissional que age de má-fé.

O princípio da cooperação, consagrado expressamente no artigo 6º do Código de Processo Civil, determina que todos os sujeitos do processo devem cooperar entre si para que se obtenha, em tempo razoável, decisão de mérito justa e efetiva. Esse modelo caracteriza-se pelo redimensionamento do princípio do contraditório, com inclusão do órgão jurisdicional no rol dos sujeitos do diálogo processual, não mais como mero espectador passivo do duelo das partes, mas como participante ativo que contribui para o aprimoramento da decisão.

Marinoni, Arenhart e Mitidiero (2021, p. 13)observam que problema central do processo está na equilibrada organização de seu formalismo, vale dizer, da divisão adequada do trabalho entre seus participantes. Modelo de processo justo no Estado Constitucional é modelo cooperativo, pautado pela colaboração efetiva do juiz para com as partes e das partes para com o juiz.

Processo pautado pela colaboração é processo orientado pela busca tanto quanto possível da verdade e que exige de todos os seus participantes observância rigorosa da boa-fé objetiva, sendo igualmente destinatário dessa exigência o juiz, tendo como objetivo produzir decisões materialmente justas e não apenas formalmente corretas (Marinoni, Arenhart e Mitidiero (2021, p. 14).

Observa-se que implementação de inteligência artificial deve respeitar deveres cooperativos que estruturam processo civil brasileiro. Sistemas automatizados não eliminam, mas potencialmente facilitam o cumprimento desses deveres quando adequadamente projetados. Sistema que facilita acesso a informações relevantes, que explicita critérios de classificação ou que sugere precedentes aplicáveis pode promover cooperação ao reduzir assimetrias informacionais.

Ataques de injeção de *prompt* violam frontalmente o modelo cooperativo de processo porque subvertem a divisão equilibrada de trabalho entre participantes processuais. Profissional que manipula sistema automatizado mediante injeção maliciosa arroga-se poder que não lhe pertence, interferindo em atividade que deveria ser exercida por tribunal segundo critérios objetivos, públicos e auditáveis. Há usurpação de função na medida em que profissional determina, mediante injeção de instruções, como o sistema deve classificar ou processar, função que pertence legitimamente a tribunal ou ao próprio sistema segundo programação regular.

O processo civil brasileiro, estruturado constitucionalmente sobre princípios da boa-fé objetiva e cooperação processual, não tolera comportamentos destinados a enganar sistemas automatizados mediante manipulação técnica intencional. Esses princípios, que vinculam todos os sujeitos processuais sem exceção, estendem-se naturalmente e necessariamente a interações com sistemas de inteligência artificial implementados no ambiente processual. A violação consciente desse dever por profissional que age de má-fé configura ato ilícito processual passível de sanções previstas em lei, incluindo responsabilização por danos e aplicação de multas.

5 GOVERNANÇA DE INTELIGÊNCIA ARTIFICIAL NO JUDICIÁRIO: NECESSIDADE DE ABORDAGEM INTERDISCIPLINAR

Mendes e Mattiuzzo (2019, p. 41-42)definem algoritmo como conjunto de instruções organizadas de forma sequencial que determina como algo deve ser feito, constituindo receita procedimental para resolução de problema. Um dos objetivos fundamentais dos algoritmos

contemporâneos é fazer previsões utilizando probabilidades, analisando dados fornecidos e oferecendo palpites estatisticamente coerentes, embora não possam fornecer respostas precisas e determinísticas a todas as questões.

Machine learning muda lógica tradicional de programação: adentram na máquina tanto dado quanto resultado desejado e produto é algoritmo capaz de tornar relação entre dado e resultado verdadeira para novos casos (Domingos, p. 06, 2015). A organização *Fairness, Accountability and Transparency in Machine Learning Organization* (elaborada por um grupo de acadêmicos) relacionou princípios-chave que deveriam ser observados pelo setor privado e pelo governo ao lidar com algoritmos: responsabilidade (*accountability)*, explicabilidade (*explainability)*, precisão, auditabilidade e justiça (*fairness*) (Mendes; Mattiuzzo, p. 55, 2019).

Responsabilidadeestá ligada à ideia de que, ao projetar sistemas algorítmicos, é preciso ter em mente que pessoas serão afetadas pelo processo decisório e que, dessa forma, é necessário, em certa medida, oferecer alternativas para eventual reparação de danos causados por decisões equivocadas. A explicabilidade é a descrição, compreensível por seres humanos não especialistas, do processo por meio do qual aquele que toma decisão, ao utilizar certo grupo de entradas, atinge dada conclusão específica. A Justiça, talvez o mais previsível, tem por objetivo que os algoritmos não podem levar a resultados discriminatórios (Mendes; Mattiuzzo, p. 56-57, 2019).

Independentemente da solução concreta a ser adotada, seja ela transparência aprimorada, desenvolvimento de ferramentas de *accountability* ou combinação estratégica dos diversos mecanismos disponíveis, o caminho a ser trilhado deve sempre guiar-se pelo papel humano no processo de automação. É essencial que pessoas que desenharão tais sistemas sejam capacitadas não só para compreender seus aspectos técnicos específicos, mas também para visualizar e antecipar efeitos do uso daquele mecanismo no mundo real com todas suas complexidades (Mendes; Mattiuzzo, 2019, p. 60).

É evidente que a centralidade humana também precisa estar refletida no processo de revisão de decisões automatizadas ou auxiliadas por automação. A preocupação deve se apresentar tanto no momento de teste do sistema quanto na análise de decisões tomadas e posteriormente questionadas por indivíduos que se sentiram por elas prejudicados. O desafio é imaginar formas de traduzir para sistemas computacionais aquilo que vem sendo construído nas ciências humanas e no pensamento ético há milênios (Mendes; Mattiuzzo, 2019, p. 61).

O Brasil adotou medidas legislativas e regulatórias para regulamentar e estabelecer diretrizes que visam mitigar e combater possíveis discriminações e abusos de poder decorrentes de bancos de dados e do uso de algoritmos de IA, destacando-sea Lei nº

13.709/18 (Lei Geral de Proteção de Dados) e a Portaria nº 4.617/2021, que estabelece Estratégia Brasileira de Inteligência Artificial. A partir de 2020, o Conselho Nacional de Justiça estabeleceu diretrizes específicas para uso da IA (Sainz; Gabardo; Ongaratto, 2024, p. 278-279).

Sociedade cada vez mais digital requer aparato estatal forte e capaz de responder adequadamente aos desafios impostos pela transformação tecnológica. Direito também requer investigações genuinamente interdisciplinares, a fim de avançar no desenvolvimento de diagnósticos atualizados e habilitados para elaborar adequadas políticas públicas e possíveis soluções regulatórias para tecnologias complexas na sociedade (Sainz; Gabardo; Ongaratto, 2024, p. 282-283).

A literatura brasileira sobre inteligência artificial no Judiciário concentra-se predominantemente em questões éticas gerais, princípios abstratos de governança e preocupações válidas sobre discriminação algorítmica(Sainz; Gabardo; Ongaratto, 2024, p. 270; 273). Embora esses temas sejam inquestionavelmente importantes para debate público informado, a análise revela uma lacuna significativa: praticamente inexiste discussão substancial sobre vulnerabilidades técnicas específicas de segurança cibernética que comprometem a integridade operacional de sistemas aplicados ao Judiciário.

Ataques de injeção de *prompt* não figuram de modo significativo em pesquisas jurídicas nacionais, evidenciando a possibilidade de que o debate brasileiro ainda não incorporou adequadamente riscos de segurança cibernética que afetam especificamente sistemas baseados em processamento de linguagem natural. Discussão sobre transparência e explicabilidade, embora necessária e bem-vinda, mostra-se estruturalmente insuficiente para enfrentar manipulações técnicas intencionais executadas por profissionais que agem de má-fé.

A transparência sobre o funcionamento geral do algoritmo não impede tecnicamente que um usuário malicioso e tecnicamente sofisticado imponha instruções destinadas a subverter esse funcionamento regular. A explicabilidade de modelo, por mais detalhada que seja, não detecta automaticamente presença de dados comprometidos por injeção de *prompt* maliciosa. O debate brasileiro privilegia legitimamente questões de equidade e não discriminação, mas a negligencia dimensão complementar de segurança técnica de sistemas.

Essa lacuna compromete potencialmente efetividade prática de regulação bem-intencionada. Resolução nº 615 do CNJ estabelece requisitos gerais de segurança, auditoria e monitoramento contínuo, mas não especifica tecnicamente como detectar ou prevenir especificamente ataques de injeção de *prompt*. Determina supervisão humana apropriada e curadoria cuidadosa de dados, mas não aborda explicitamente como garantir integridade de

dados processados por modelos de linguagem contra manipulações maliciosas durante fase de produção.

A incorporação tardia e ainda superficial de questões técnicas específicas de segurança no debate jurídico brasileiro reflete tendência mais ampla de separação disciplinar entre conhecimento técnico especializado e conhecimento jurídico tradicional. Juristas discutem princípios éticos abstratos sem necessariamente dominar a arquitetura concreta de sistemas, sem necessariamente considerar todas as implicações jurídicas e processuais.

A necessidade de abordagem interdisciplinar genuína, que integre expertise técnica em segurança de sistemas com compreensão profunda de princípios processuais e garantias constitucionais, torna-se evidente quando se considera a complexidade de implementar sistemas de inteligência artificial que sejam simultaneamente úteis, seguros contra manipulações maliciosas e respeitosos de direitos fundamentais processuais.

6 SUBSUNÇÃO JUDICIAL E INTELIGÊNCIA ARTIFICIAL: COMPLEMENTARIDADE RESPONSÁVEL

O art. 19, § 3º, inciso II da Resolução nº 615 do CNJ estabelece que uso de ferramentas de inteligência artificial será de caráter auxiliar e complementar, consistindo em mecanismos de apoio à decisão, vedada expressamente utilização como instrumento autônomo de tomada de decisões judiciais sem devida orientação, interpretação, verificação e revisão crítica por parte do magistrado, que permanecerá integralmente responsável pelas decisões tomadas e pelas informações nelas contidas (Brasil. Conselho Nacional de Justiça, 2025).

Sistema inteligente deverá assegurar autonomia plena dos usuários internos, com uso de modelos que promovam incremento da eficiência, precisão e qualidade das atividades desempenhadas, sem limitar de forma alguma a capacidade de atuação independente dos usuários (art. 32, incisos I e II da Resolução nº 615 do CNJ). Deve possibilitar revisão detalhada e facilitada do conteúdo gerado e dos dados utilizados para sua elaboração, assegurando que usuários tenham acesso transparente às premissas e ao método empregado pela inteligência artificial na sua formulação, sem que haja qualquer espécie de vinculação obrigatória à solução apresentada pela inteligência artificial e garantindo-se possibilidade real de correções ou ajustes substantivos (Brasil. Conselho Nacional de Justiça, 2025, p. 33).

O art. 42 da Resolução dispõe que os Órgãos do Poder Judiciário deverão informar ao Comitê Nacional de Inteligência Artificial do Judiciário todos os eventos adversos

relacionados ao uso de soluções de inteligência artificial, possibilitando aprendizado institucional e aprimoramento contínuo. Em complemento, o § 1º do referido artigo considera-se eventos adversos os incidentes que resultem em impactos negativos sobre a operação adequada do sistema, segurança dos dados processados ou qualidade da prestação de serviços jurisdicionais.

Essas diretrizes normativas estabelecem arquitetura institucional equilibrada que reconhece simultaneamente tanto potencial significativo da inteligência artificial para incrementar eficiência e qualidade do serviço jurisdicional quanto necessidade imperiosa de salvaguardas técnicas e organizacionais contra usos inadequados por erro ou contra manipulações maliciosas por profissionais que agem de má-fé. A regulamentação não parte de presunção injustificada de que tecnologia constitui ameaça em si mesma, mas de reconhecimento maduro de que qualquer ferramenta poderosa exige governança apropriada para maximizar benefícios legítimos enquanto minimiza riscos conhecidos.

Sichman (2021, p. 42)adverte que sistemas sociotécnicos contemporâneos, compostos de elementos sociais humanos e elementos técnicos automatizados trabalhando conjuntamente para objetivos comuns, devem ser projetados e implementados de modo que possibilitem que ambos elementos gerem resultados positivos, diferentemente de métodos anteriores em que pessoas eram forçadas a adaptar-se unilateralmente aos elementos técnicos inflexíveis.

A questão central não consiste em determinar abstratamente se inteligência artificial deve ou não ser utilizada no Judiciário, mas como estruturar concretamente essa utilização de modo que potencialize capacidades humanas existentes sem substituir inadequadamente julgamento humano em dimensões que exigem interpretação contextual, ponderação axiológica e responsabilidade institucional. Sistemas automatizados destacam-se particularmente em processamento rápido e preciso de grandes volumes de dados, identificação de padrões estatísticos em conjuntos complexos de informações, recuperação eficiente de precedentes e execução consistente de tarefas repetitivas padronizadas.

Magistrados destacam-se em compreensão qualitativa de nuances do caso concreto, ponderação equilibrada de valores e princípios conflitantes, consideração adequada de peculiaridades específicas que distinguem caso sob julgamento de precedentes aparentemente similares e fundamentação responsável de decisões com *accountability* institucional. A complementaridade estratégica entre essas capacidades distintas oferece oportunidade concreta de aprimorar substancialmente a prestação jurisdicional sem comprometer garantias processuais.

No Judiciário, design de sistemas de inteligência artificial deve representar adequadamente não apenas eficiência processual quantitativa, mas também valores constitucionais fundamentais como contraditório efetivo, ampla defesa, fundamentação adequada das decisões e acesso material à justiça.

Sistema que maximiza exclusivamente a velocidade de tramitação pode inadvertidamente comprometer a qualidade substancial de decisões. O desafio prático consiste em projetar ferramentas que aumentem a produtividade mensurável sem sacrificar garantias processuais constitucionalmente asseguradas.

Russell e Norvig (2013, p. 1154) observam que a racionalidade perfeita, ideal teórico em que o agente age a todo instante de forma a maximizar sua utilidade esperada, não é viável em ambientes reais complicados porque as demandas computacionais são proibitivamente elevadas. Otimização limitada constitui abordagem mais realista: agente se comporta tão bem quanto possível dados seus recursos computacionais finitos.

Essa perspectiva teórica oferece uma moldura conceitual útil para compreender a complementaridade entre inteligência artificial e interpretação judicial. Sistemas automatizados operam sob otimização limitada: dadas capacidades computacionais disponíveis e dados acessíveis, produzem melhores sugestões estatisticamente possíveis segundo critérios explicitamente programados. Magistrados operam sob racionalidade aplicada: dadas limitações cognitivas humanas e restrições temporais práticas, tomam melhores decisões juridicamente possíveis segundo compreensão jurídica desenvolvida e experiência profissional acumulada. Combinação estratégica dessas duas formas complementares de inteligência pode produzir resultados práticos superiores a cada uma isoladamente.

Por exemplo, o sistema pode processar rapidamente milhares de precedentes e identificar padrões jurisprudenciais relevantes. O magistrado pode avaliar criticamente se precedentes identificados aplicam-se efetivamente ao caso concreto específico, considerando peculiaridades fáticas e jurídicas que o sistema estatístico não captura adequadamente. Trata-se de divisão funcional de trabalho cognitivo que potencializa capacidades naturais de ambos os componentes.

Marinoni, Arenhart e Mitidiero (2021, p. 16) esclarecem que juiz tem dever constitucional de fundamentação analítica, devendo explicitar razões de decidir de forma clara, precisa e logicamente estruturada. Introdução de inteligência artificial como ferramenta auxiliar não altera essas responsabilidades fundamentais constitucionalmente estabelecidas,

mas pode potencialmente modificar de modo produtivo como essas responsabilidades são concretamente exercidas no cotidiano forense.

A questão dos ataques de injeção de *prompt* ilustra com clareza a necessidade prática de salvaguardas técnicas robustas não contra a tecnologia em si mesma, mas especificamente contra usos maliciosos deliberados por profissionais que agem conscientemente de má-fé visando subverter o funcionamento regular de sistemas. Assim como o sistema processual tradicional requer proteção estabelecida contra falsificação de documentos, litigância de má-fé e fraude processual sistema automatizado requer equivalentemente proteção técnica específica contra manipulação algorítmica maliciosa.

A capacitação continuada de magistrados e servidores sobre riscos específicos e limitações técnicas conhecidas de sistemas de inteligência artificial não visa desencorajar uso legítimo, mas promover uso informado, crítico e responsável. Profissional que compreende adequadamente como modelo de linguagem funciona tecnicamente, quais padrões ele reconhece estatisticamente e quais limitações estruturais ele apresenta está objetivamente melhor equipado para aproveitar produtivamente o potencial real da ferramenta enquanto evita confiantemente armadilhas conhecidas.

A vedação normativa a sistemas que gerem dependência absoluta do usuário ou não possibilitem revisão humana efetiva reflete compreensão institucional madura de que inteligência artificial deve amplificar capacidades humanas existentes, não substituí-las inadequadamente. A diferença crítica reside fundamentalmente no modo consciente de uso, não na tecnologia disponibilizada.

A subsunção do fato jurídico concreto à norma abstrata mediante interpretação judicial fundamentada constitui atividade central da jurisdição que permanece essencialmente humana, não por limitação tecnológica transitória, mas por natureza profunda da tarefa interpretativa. Interpretação jurídica envolve necessariamente compreensão qualitativa de linguagem em contexto social específico, ponderação reflexiva de valores e princípios potencialmente conflitantes, consideração adequada de finalidades normativas constitucionais e responsabilização institucional por consequências práticas da decisão.Essa construção interpretativa exige julgamento no sentido forte: capacidade desenvolvida de discernir entre alternativas normativas válidas, ponderar equilibradamente razões jurídicas conflitantes e assumir responsabilidade institucional pela decisão fundamentada.

A complementaridade responsável entre inteligência artificial e interpretação judicial estrutura-se, portanto, sobre divisão funcional clara: sistemas automatizados incumbem-se produtivamente de tarefas que se beneficiam objetivamente de processamento rápido de

grandes volumes de dados, reconhecimento estatístico de padrões complexos e execução consistente de rotinas padronizadas; magistrados incumbem-se insubstituivelmente de tarefas que exigem compreensão contextual qualitativa, ponderação axiológica reflexiva e responsabilidade decisória institucional.

A inteligência artificial, quando devidamente governada mediante regulação apropriada e tecnicamente protegida contra manipulações maliciosas por profissionais que agem de má-fé, constitui recurso genuinamente valioso para aprimoramento mensurável da prestação jurisdicional. O desafio institucional reside em construir concretamente modos de uso que respeitem valores constitucionais fundamentais, promovam eficiência operacional sem sacrificar qualidade substancial, aumentem produtividade quantitativa sem comprometer garantias processuais e incorporem inovação tecnológica mediante deliberação democrática adequadamente informada.

O papel do magistrado como intérprete final fundamentado da norma aplicada ao caso concreto não é ameaçado pela inteligência artificial adequadamente implementada, mas potencialmente enriquecido produtivamente por ela. Magistrado com acesso a ferramentas bem projetadas que facilitam pesquisa abrangente, identificam precedentes relevantes, verificam consistência argumentativa e sugerem estruturas de fundamentação pode realisticamente dedicar maior atenção qualitativa a aspectos que efetivamente exigem julgamento humano insubstituível: compreensão profunda das peculiaridades do caso, ponderação equilibrada de interesses conflitantes, adequação reflexiva de solução jurídica a circunstâncias específicas e responsabilização institucional por consequências da decisão.

A subsunção judicial permanece como garantia fundamental constitucional precisamente porque inteligência artificial bem utilizada de modo responsável fortalece objetivamente, em vez de enfraquecer, capacidade institucional do Judiciário de realizar seu propósito constitucional de prestar jurisdição justa, efetiva e tempestiva, protegendo direitos fundamentais e resolvendo conflitos conforme o Direito.

7 CONSIDERAÇÕES FINAIS

A inserção de sistemas de inteligência artificial no Poder Judiciário evidencia uma tensão que não se resolve com ajustes pontuais da dogmática clássica. A busca por maior eficiência convive com vulnerabilidades técnicas que permitem manipulações deliberadas por profissionais que agem de má-fé. Ataques de *prompt injection* mostram que modelos de linguagem, ao processarem dados e instruções no mesmo canal, tornam-se suscetíveis a

interferências que afetam classificações e resultados produzidos pelas aplicações. Nesse ambiente, a boa-fé objetiva deixa de enfrentar apenas condutas tradicionais de deslealdade e passa a abranger manipulações algorítmicas que subvertem o funcionamento de sistemas automatizados para obtenção de vantagens processuais indevidas.

O desenvolvimento teórico evidencia que boa-fé e cooperação processual constituem critérios normativos indispensáveis para preservação da integridade do procedimento em contexto de automação crescente. Esses princípios permitem distinguir o uso legítimo de ferramentas que ampliam a capacidade de análise de dados de práticas que distorcem critérios legais. Tal distinção fornece base dogmática para impedir que sistemas judiciais se tornem alvo de exploração técnica sob aparência formal de regularidade ou com justificativas centradas em ganhos privados de eficiência.

A noção de complementaridade responsável organiza a relação entre inteligência artificial e atividade jurisdicional. A separação entre tarefas que dependem de processamento computacional e aquelas que exigem discernimento jurídico permite estabelecer limites adequados para automação. Sistemas podem executar atividades de organização, verificação e recuperação de informações, enquanto magistrados permanecem responsáveis por interpretação contextual, ponderação de valores e fundamentação analítica. Quando decisões são influenciadas por dados manipulados, a lesão não se resume ao caso concreto, atinge a confiança geral no sistema de justiça.

A caracterização técnica dos ataques de injeção de *prompt* contribui para nova compreensão sobre litigância de má-fé. Manipulações algorítmicas exploram propriedades estruturais do modo como modelos de linguagem tratam entradas textuais. A opacidade desses sistemas, cujo aprendizado decorre de padrões estatísticos complexos, dificulta identificação por meios tradicionais de fiscalização. Isso exige mecanismos de defesa específicos, que envolvem arquitetura de sistemas, protocolos de segurança e rotinas de auditoria contínua.

Nesse cenário, as violações assumem dimensão sistêmica. A introdução de instruções maliciosas em documentos processuais não prejudica apenas a parte adversa, compromete confiança institucional na automação. A vantagem momentânea obtida por meio da manipulação técnica resulta na erosão de credibilidade das ferramentas utilizadas no processo. Por isso, a resposta jurídica não pode limitar-se a sanções individuais, deve alcançar padrões de desenvolvimento, governança e cultura institucional.

O debate nacional sobre governança algorítmica precisa incorporar a dimensão de segurança cibernética às preocupações éticas e regulatórias já consolidadas. A ênfase quase exclusiva em transparência, discriminação e explicabilidade não abrange riscos concretos que

afetam capacidade operacional dos sistemas. A ampliação do escopo regulatório, sem afastar as preocupações já existentes, evita lacunas que fragilizem o processo diante de manipulações técnicas.

A centralidade da subsunção judicial reforça limites intransponíveis da automação. Sistemas podem auxiliar magistrados com análises, sínteses e organização de informações, mas não podem substituir atividade interpretativa que envolve responsabilidade institucional e ponderação de consequências práticas. A autonomia decisória do juiz depende, inclusive, de garantias técnicas que impeçam dependência excessiva ou vinculação automática às saídas produzidas por sistemas inteligentes.

O desafio está em encontrar ponto de equilíbrio. A inovação tecnológica, etapa natural de aprimoramento administrativo, deve ser guiada por valores constitucionais. Sistemas inteligentes podem ampliar precisão e celeridade, mas não podem redefinir o núcleo da jurisdição. A racionalidade algorítmica opera sobre padrões estatísticos, enquanto a decisão judicial exige compreensão normativa, avaliação contextual e responsabilidade institucional.

O ambiente atual exige do jurista capacidade de integrar fundamentos constitucionais e compreensão técnica suficiente para avaliar impactos da automação. A proteção do devido processo legal depende tanto de garantias formais quanto de segurança técnica dos instrumentos utilizados. Somente com governança que articule eficiência e integridade o Judiciário conseguirá preservar legitimidade das decisões em contexto marcado pela presença crescente da inteligência artificial. A subsunção judicial permanece núcleo irredutível da atividade jurisdicional e precisa ser preservada mesmo quando ferramentas que apoiam o processo tornam-se tecnologicamente complexas e opacas.

REFERÊNCIAS

BRASIL. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República. Disponível em: https://www.planalto.gov.br/ccivil_03/Constituicao/ConstituicaoCompilado.htm. Acesso em: 13 nov. 2025.

BRASIL. **Lei nº 10.406, de 10 de janeiro de 2002 (Código Civil)**. Disponível em: https://www.planalto.gov.br/ccivil_03/LEIS/2002/L10406compilada.htm. Acesso em: 13 de nov. de 2025.

BRASIL. Conselho Nacional de Justiça. **Inteligência artificial na Justiça**. Brasília: CNJ, 2019. 40 p. Disponível em: https://www.cnj.jus.br/wp-content/uploads/2020/05/Inteligencia_artificial_no_poder_judiciario_brasileiro_2019-11-22.pdf. Acesso em: 16 nov. 2025.

BRASIL. Conselho Nacional de Justiça. **Resolução nº 615, de 11 de março de 2025**. Estabelece diretrizes para o desenvolvimento, utilização e governança de soluções desenvolvidas com recursos de inteligência artificial no Poder Judiciário. Brasília: CNJ, 2025. Disponível em: https://atos.cnj.jus.br/files/original155530202503146 7d4517244566.pdf. Acesso em: 16 nov. 2025.

CHEN, Sizhe; PIET, Julien; SITAWARIN, Chawin; WAGNER, David. StruQ: Defending Against Prompt Injection with Structured Queries. **34th USENIX Security Symposium**, Seattle, 2025. Disponívelem: https://www.usenix.org/conference/usenixsecurity25/presentation/chen-sizhe. Acesso em: 16 nov. 2025.

DIDIER JR., Fredie. **Curso de direito processual civil**: introdução ao direito processual civil, parte geral e processo de conhecimento. 21. ed. Salvador: Ed. Jus Podivm, 2019.

DOMINGOS, P. **The Master Algorithm**: How the Quest for the Ultimate Learning Machine Will Remake our World. New York: Basic Books, 2015.

ESPOSITO, Elena. Comunicação artificial? A produção de contingência por algoritmos. **RBSD – Revista Brasileira de Sociologia do Direito**, v. 9, n. 1, p. 4-41, jan./abr. 2022. Disponível em: https://revista.abrasd.com.br/index.php/rbsd/article/view/638. Acesso em: 13 nov. 2025.

FEENBERG, Andrew. **Entre a razão e a experiência**: Ensaios sobre a tecnologia e a modernidade. Independently Published, 2019.

LIU, Yupei; JIA, Yuqi; GENG, Runpeng; JIA, Jinyuan; GONG, Neil Zhenqiang. Formalizing and Benchmarking Prompt Injection Attacks and Defenses. **33rd USENIX Security Symposium**, Philadelphia, 2024. p. 1831-1847. Disponível em: https://www.usenix.org/conference/usenixsecurity24/presentation/liu-yupei. Acesso em: 16 nov. 2025.

LUHMANN, N. Sozialesysteme. **Grundriße einer allgemeinen Theorie**. Frankfurt amMain: Suhrkamp, 1984.

MARINONI, Luiz Guilherme; ARENHART, Sérgio Cruz; MITIDIERO, Daniel. **Código de Processo Civil Comentado**. 7. ed. São Paulo: Thomson Reuters Brasil, 2021.

MENDES, Laura Schertel; MATTIUZZO, Marcela. Discriminação Algorítmica: Conceito, Fundamento Legal e Tipologia. **Direito Público**, Porto Alegre, v. 16, n. 90, p. 39-64, nov./dez. 2019. Disponível em: https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/3766. Acesso em: 16 nov. 2025.

RUSSELL, Stuart J.; NORVIG, Peter. **Inteligência artificial**. Rio de Janeiro: Elsevier, 2013.

SAINZ, Nilton; GABARDO, Emerson; ONGARATTO, Natália. **Discriminação algorítmica no Brasil**: uma análise da pesquisa jurídica e suas perspectivas para a compreensão do fenômeno. RDP, Brasília, vol.21 n. 110, p. 258-289, abr./jun. 2024. Disponível em:

https://www.portaldeperiodicos.idp.edu.br/direitopublico/article/view/7295. Acesso em: 13 nov. 2025.

SICHMAN, Jaime Simão. Inteligência Artificial e sociedade: avanços e riscos. **Estudos Avançados**: São Paulo, v. 35, n. 101, p. 37–50, 2021. Disponível em: https://doi.org/10.1590/s0103-4014.2021.35101.004. Acesso em: 16 nov. 2025.

STF finaliza testes de nova ferramenta de Inteligência Artificial. **STF**, Brasília, 11 mai. 2023. Disponível em: https://portal.stf.jus.br/noticias/verNoticiaDetalhe.asp?idConteudo=507120&ori=1. Acesso em: 16 de nov. 2025.

STF lança MARIA, ferramenta de inteligência artificial que dará mais agilidade aos serviços do Tribunal. **STF**, 16 dez. 2024. Disponível em: https://noticias.stf.jus.br/postsnoticias/stf-lanca-maria-ferramenta-de-inteligencia-artificial-que-dara-mais-agilidade-aos-servicos-do-tribunal/. Acesso em: 16 nov. de 2025.